

## 日本語 HPSG：統語的複合動詞の統語・意味構造の処理

橋本 力

chashi@sils.shoin.ac.jp

神戸松蔭女子学院大学大学院

### 概要

近年、自然言語処理と理論言語学の乖離が著しい。しかし我々は、両者とも言語の計算体系に関する研究であり、2つのアプローチが接点を持ち、互いに有益な形で相互作用しうると考えている。我々はこの考えのもとで、計算機上で動作する日本語 HPSG 文法を、理論的な考察を踏まえた上で構築している。

今回我々は統語的複合動詞に関わる現象の分析と、その形式化、実装に取り組んだ。我々は日本語の統語的複合動詞を項構造の観点から3種類に分類した。この分類はθ付与の可能性の違いにより独立に支持される。その分類が、長距離受け身を中心とする統語的複合動詞に関わる様々な現象——動詞句照応、願望形、V<sub>1</sub>主語尊敬化——の説明に貢献することを示す。この分析に従って動詞階層や句階層を形式化し、文法開発環境 LKB を用いて計算機上に実装した。動詞階層は外項内項の有無による分類と、他の動詞を埋めこむか否かによる分類により構成される。実装に際して多少の再形式化が必要になったが、理論的な分析の知見をほぼ忠実に実装に反映することに成功した。

キーワード：統語的複合動詞、長距離受け身、HPSG、LKB

## Japanese HPSG : Treatment of Syntax and Semantics of Syntactically Complex Verbs

HASHIMOTO Chikara  
Kobe Shoin Graduate School

### Abstract

While theoretical linguistics and NLP are both studies of linguistic computation there has been a dearth of work recently addressing their interaction. This paper shows how by constructing a computer-executable Japanese grammar within the HPSG framework, while taking into account certain theoretical concepts, the two can benefit from each other.

We report our analysis, formalization, and implementation of a Japanese grammar involving syntactically complex verbs. We divide syntactically complex verbs into three types based on their argument structure. This classification is independently motivated by their theta role assignment possibilities, and helps us account for long distance passives as well as other phenomena — verbal anaphora, desideratives and V1 honorifics. We formalize a type hierarchy of verbs and phrases based on our analysis and implement the grammar using LKB software, a grammar development environment. The verb hierarchy classifies verbs along two dimensions: whether they have an external (or internal) argument, and whether they can embed a verb.

Key Words : Syntactically complex verbs, Long distance passive, HPSG, LKB

## 1 はじめに

かつては互いに影響しあっていた自然言語処理と理論言語学も、近年ではその乖離が著しい。しかし我々は、両者とも言語の計算体系に関する研究であり、この言語に対する2つのアプローチが接点を持ち、互いに有益な形で相互作用しうると考えている。現在我々はこの考え方のもとで、計算機上で動作する日本語 HPSG 文法を、理論的な考察を踏まえた上で構築している（橋本 2003a）。

日本語 HPSG 文法を計算機に実装する試みとして Mitsuishi, Torisawa, and Tsuji (1998), 松本 (2000) があるが、これらは文法としての精密さより処理効率や網羅性などの実用性を第一に考えている。一方我々は理論言語学の知見を限りなくそのままの形で実装することを目指す。

ヨーロッパ系の言語が接辞や屈折語尾を発達させているのに対し、膠着型言語である日本語は単語の複雑な複合構造を持つ。日本語の大きな特徴と言える複合構造は日本語文法を論じる上での主要なテーマの一つであり、今回我々はこの複合構造、とりわけ統語的複合動詞の処理の実装に取り組んだ。分析は橋本 (2003b)に基づき、実装には文法開発環境 LKB (Copestake 2002) を用いた。

以下、2節では、影山 (1993) による統語的複合動詞の観察と、それに関わる現象として長距離受け身を挙げる。3節では、長距離受け身やその他の統語的複合動詞に関わる現象の HPSG による分析を示し、それにに基づき、動詞と句の階層を形式化する。4節では、本研究で用いる文法開発環境について述べた後、実装用に再形式化された文法について具体例とともに述べる。5節は結論である。

## 2 統語的複合動詞

日本語の特徴である複合語には、レキシコンで生産される語彙的な複合と統語構造で生産される統語的複合がある。本稿では「書き終わる」「書きそびれる」「書き直す」のような「動詞 (V<sub>1</sub>) + 動詞 (V<sub>2</sub>)」で構成される統語的複合動詞を扱う。

### 2.1 3種の複合動詞と長距離受け身

影山 (1993) では統語構造で作られる複合動詞には3種類あることが論じられている。

- (1) a. 健が 日記を V<sub>1</sub> 書き V<sub>2</sub> 終わる
- b. 健が 日記を V<sub>1</sub> 書き V<sub>2</sub> そびれる
- c. 健が 日記を V<sub>1</sub> 書き V<sub>2</sub> 直す

- (2) 非対格タイプ (1a) : V<sub>2</sub> は主語にも目的語にも θ 付与しない。

他動詞型 VP 補語タイプ (1b) : V<sub>2</sub> は目的語には θ 付与せず、主語には θ 付与する。

他動詞型 V 補語タイプ (1c) : V<sub>2</sub> は主語にも目的語にも θ 付与する。つまり目的語は V<sub>1</sub>, V<sub>2</sub> の両方から θ 付与される。

3者は、上の意味的な違いに加え、統語的にも異なる振舞いを示す。つまり V 補語タイプのみが長距離受け身<sup>1</sup> が可能である。

- (3) a. \*日記が 健に 書き 終わ られる
- b. \*日記が 健に 書き そびれ られる
- c. 日記が 健に 書き 直さ れる

## 3 HPSG による分析

### 3.1 繰り上げとコントロール

本稿では繰り上げ構文とコントロール構文を用いる。以下では HPSG でのそれらの分析を概観する (Sag & Wasow 1999)。繰り上げ動詞 *continue* の語彙項目は (4a) のようになる。<sup>2</sup>

- (4) a.  $\left[ \begin{array}{l} \text{PHON /continue/} \\ \text{ARG-ST} \left\langle \boxed{1}, \text{CP} \left[ \begin{array}{l} \text{SPR} \left\langle \boxed{1} \text{ NP } \right\rangle, \\ \text{SEM } \{ \text{INDEX } \boxed{2} \} \end{array} \right] \right\rangle \\ \text{SEM } \left[ \text{RESTR} \left\langle \left[ \begin{array}{l} \text{RELN continue} \\ \text{ARG } \boxed{2} \end{array} \right] \right\rangle \right] \end{array} \right]$

b. It  $\boxed{1}$  continues CP [  $\boxed{1}$  NP ] to rain.

ARG-ST の第1要素  $\boxed{1}$  は第2要素の CP の SPR の値と構造を共有している。*continue* は CP の指示対象  $\boxed{2}$  には θ 付与する (RELN *continue* の ARG) が、主語には θ 付与しない。(4b) の主語 *It* は指示対象を持たないが、*continue* から θ 付与されず、意味的な制限が課せられないので問題ない。コントロール動詞 *try* の語彙項目は (5a) のようになる。

- (5) a.  $\left[ \begin{array}{l} \text{PHON /try/} \\ \text{ARG-ST} \left\langle \text{NP}_i, \text{CP} \left[ \begin{array}{l} \text{SPR} \left\langle \text{NP}_i \right\rangle, \\ \text{SEM } \{ \text{INDEX } \boxed{2} \} \end{array} \right] \right\rangle \\ \text{SEM } \left[ \text{RESTR} \left\langle \left[ \begin{array}{l} \text{RELN try} \\ \text{TRIER } i \\ \text{ARG } \boxed{2} \end{array} \right] \right\rangle \right] \end{array} \right]$

b. Ken<sub>i</sub> tried CP [ (NP<sub>i</sub>) to sleep ].

*try* は CP の指示対象  $\boxed{2}$  にも (RELN *try* の ARG)、主語の指示対象  $i$  にも θ 付与する (TRIER)。つまり *try* の主語は TRIER として解釈し得るものでなくてはならない。よって \*It tried to rain. のように *try* が指示対象を持たない主語を取ることはできない。

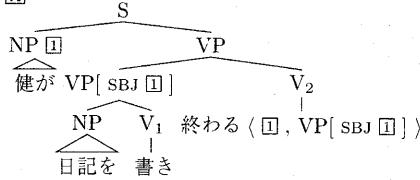
<sup>1</sup>V<sub>1</sub>-V<sub>2</sub> 全体が受動化される現象。

<sup>2</sup>PHON は音韻情報、ARG-ST は項構造、SPR は指定部、SEM は意味構造、RELN は意味的関係を表す。

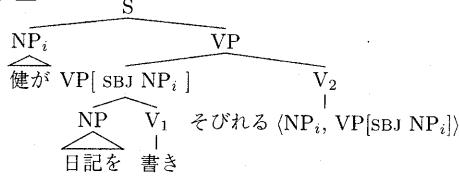
### 3.2 3種の複合動詞の分析

以下では影山(1993)の非対格タイプを[A型], VP補語タイプを[B型], V<sub>2</sub>補語タイプを[C型]と呼び, 各々を(6)-(8)のように分析する。<sup>3 4</sup>

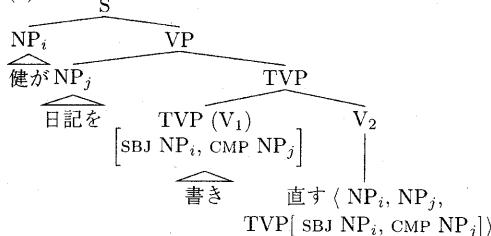
(6) [A]



(7) [B]



(8) [C]



(6) は、[A]では主語「健が」とVP「日記を書き」のSBJの値が構造を共有しており(繰り上げ), V<sub>2</sub>「終わる」が「健が」と「日記を書き」を下位範疇化することを示す。繰り上げ構文であるため, V<sub>2</sub>は主語「健が」にθ付与しない。(7) は、[B]では主語「健が」がVP「日記を書き」の主語をコントロールし, V<sub>2</sub>「そびれる」が「健が」と「日記を書き」を下位範疇化することを示す。コントロール構文であるため, V<sub>2</sub>は主語「健が」にθ付与する。(8) は、[C]では主語「健が」と目的語「日記を」がTVP(V<sub>1</sub>)「書き」の主語と目的語をコントロールし, V<sub>2</sub>「直す」が主語「健が」と目的語「日記を」, そしてTVP(V<sub>1</sub>)「書き」を下位範疇化することを示す。<sup>5</sup> V<sub>2</sub>は主語

<sup>3</sup> ノードラベルの定義は Gunji (1987) に従う。

S: 主語, 目的語が飽和した動詞句

VP: 主語が未飽和な動詞句

TVP: 主語と1つの目的語が未飽和な動詞句

<sup>4</sup> 以下ではV<sub>2</sub>のARG-STをその右もしくは下に示す。またSBJ, CMPは主語と目的語を表す。

<sup>5</sup> (8)の構造ではV<sub>2</sub>が目的語を下位範疇化しているが, 目的語の格を決定するのはV<sub>1</sub>の方である。

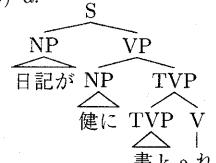
i. 健が 日記を 書き 直す

ii. 健が 先輩に 挨拶し 直す

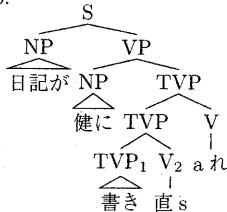
だけでなく目的語も下位範疇化しており, 目的語にもθ付与する。これらの繰り上げとコントロールの違いにより(2)の意味的な違いを説明できる。

本研究ではGunji (1987)に従い, 直接受け身の「られ」はTVPを下位範疇化すると考える。<sup>6</sup>

(9) a.



b.



3つの統語的複合の内[C]だけが長距離受け身に必要な(V<sub>2</sub>を含む)TVPノードを持つ。よって[C]だけが長距離受動化可能であることが(6)-(8)の構造から導ける。

### 3.3 他の現象の分析

#### 3.3.1 動詞句照応

「そうする」がVPだとする。<sup>7</sup> すると[A], [B]がVPを下位範疇化するのに対し[C]がTVPを下位範疇化することから, [A], [B]においてV<sub>1</sub>が「そうする」で置き換えられること(10)(11), [C]では不可能なこと(12)が説明される。<sup>8</sup>

- (10) a. 健が [VP 日記を 書き] 終わった  
b. 奈緒美も [VP そうし] 終わった

- (11) a. 健が [VP 日記を 書き] そびれた  
b. 奈緒美も [VP そうし] そびれた

- (12) a. 健が [VP 日記を [TVP [TVP 書き] 直した]]  
b. \*奈緒美も [VP そうし] 直した  
c. 奈緒美も [VP そうした]

#### 3.3.2 願望形

[A]に「-たい」が付与された場合目的語は対格しか許さない。一方[C]だと主格でも許容される。

- (13) a. 僕は 日記\*が / を 書き 終わりたかった  
b. 僕は 日記 が / を 書き 終えたかった

<sup>6</sup> 以後, 本構造においてTVP(V<sub>1</sub>)をTVP<sub>1</sub>と書く。

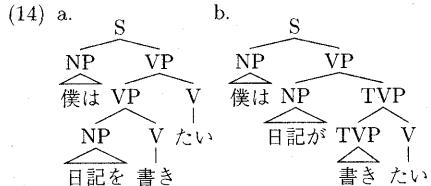
<sup>7</sup> 「そうする」がVPであるかどうかの議論についてはHinds (1973), 井上(1976), Hasegawa (1980), Gunji (1987)を参照。

<sup>8</sup> ii. のV<sub>2</sub>は[B]型のV<sub>2</sub>であるが非文となる。これらは、「そうする」の先行詞が意図的な行為を表す動詞でなければならぬと規定することで排除する。

i. 雨が ふり 終わった  
ii. \*雪も そうし 終わった

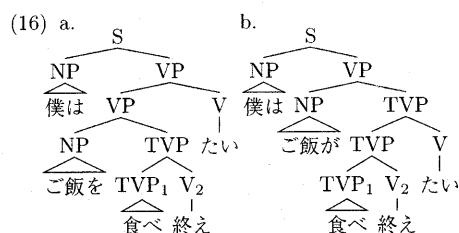
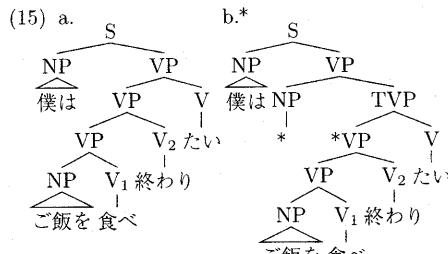
「-たい」は目的語に主格を付与することもあるが、動詞語幹が目的語に与える格をそのまま付与することもある。そこで2つの「-たい」を考える。

1. VP-たい: (主語, VP) (14a)
2. TVP-たい: (主語, 主格目的語, TVP) (14b)



(14a)では「VP-たい」が使われておらず、目的語「日記を」の対格は「書き」によって与えられる。(14b)では「TVP-たい」が使われておらず、目的語「日記が」の主格は「たい」によって付与される。

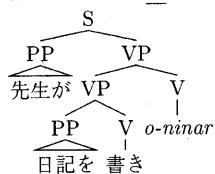
[A], [B]にはV<sub>2</sub>を含むTVPノードが無いので「VP-たい」のみが(15), [C]にはV<sub>2</sub>を含むTVPノードがあるので「VP-たい」と「TVP-たい」の両方が使われる(16)。(15b)では「TVP-たい」がVPを下位範疇化し、また主格目的語に相当するものが無いので認められない。



### 3.3.3 V<sub>1</sub>主語尊敬化

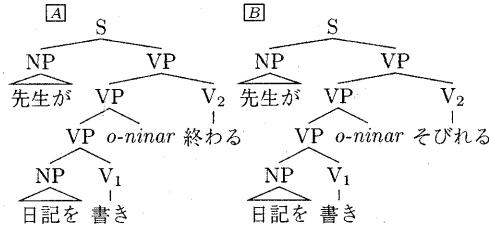
主語尊敬動詞 *o-ninar* は統語的には VP を下位範疇化し VP を形成し、音韻形態的には *o* が隣接する動詞群の接頭辞になる (Gunji 1999)。

- (17) 先生が日記を お書き になる

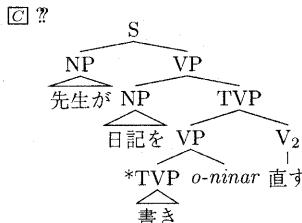


[A], [B]はV<sub>1</sub>主語尊敬化が可能だが、[C]は不可能である。

- (18) [A]: 先生が日記を お書き になり 終わる
- [B]: 先生が日記を お書き なり そびれる



- (19) [C]: ?先生が日記を お書き なり 直す



(17)の形式化と、[A], [B]は VP を下位範疇化するが [C]は TVP を下位範疇化するという本稿の分析により上の事実が予測される。つまり、(19)において、本来 VP を下位範疇化すべき *o-ninar* が TVP を下位範疇化していることが容認度低下の原因となる。

## 3.4 HPSGによる形式化

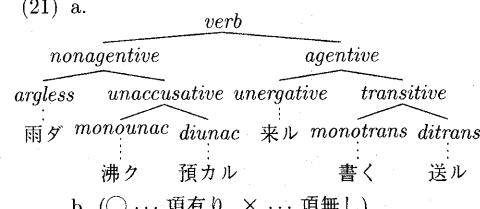
### 3.4.1 動詞の階層

我々の文法では語の ARG-ST は、郡司 (2000)に基づき、外項 (EXT) と 2 つの内項 (INT) により構造化されている。

- (20) [EXT ..., INT1 ..., INT2 ...]

動詞階層も郡司 (2000)に基づき外項内項の有無により (21) のように考える。

- (21) a.



- b. (○ ... 項有り, × ... 項無し)

		EXT	INT1	INT2
nonagent	<i>argless</i>	×	×	×
	<i>unaccu</i>	×	○	×
agentive	<i>unergative</i>	○	×	×
	<i>transi</i>	○	○	×

例えば *monounac* は外項を持たないが内項を 1 つ (INT1) 持つ. *ditrans* は外項、2 つの内項全てを持つ.

本稿では **A**, **B**, **C** に (22) の ARG-ST を想定する. (22) から **A** が *diunac*, **B** が *monotrans*, **C** が *ditrans* に属することが分かる.

- (22) **A**:  $\left[ \text{EXT} \langle \rangle, \text{INT1 } \boxed{1}, \text{INT2 VP} [ \text{SBJ } \boxed{1} ] \right]$   
**B**:  $\left[ \text{EXT NP}_i, \text{INT1 VP} [ \text{SBJ NP}_i ], \text{INT2 } \langle \rangle \right]$   
**C**:  $\left[ \text{EXT NP}_i, \text{INT1 NP}_j, \text{INT2 TVP} [ \text{SBJ NP}_i, \text{CMP NP}_j ] \right]$

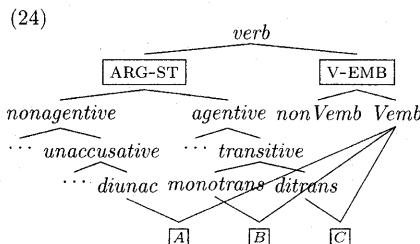
つまり **A** では、INT2 に VP を取り、INT1 にはその VP の主語の値と構造を共有している(繰り上げられている)NP を取る. **B** では、INT1 に VP を取り、EXT にはその VP の主語をコントロールする NP を取る. **C** では、INT2 に TVP を取り、EXT にはその TVP の主語をコントロールする NP を、INT1 にはその TVP の目的語をコントロールする NP を取る.

(22) の ARG-ST からの統語構造へのリンクは、以下の項具現化原理により行われる.

- (23) a.  $\left[ \text{SBJ } \boxed{1}, \text{CMP1 } \boxed{2}, \text{CMP2 } \langle \rangle, \text{ARG-ST } \left[ \text{EXT } \langle \rangle, \text{INT1 } \boxed{1}, \text{INT2 } \boxed{2} \right] \right]$   
b.  $\left[ \text{SBJ } \boxed{1}, \text{CMP1 } \boxed{2}, \text{CMP2 } \boxed{3}, \text{ARG-ST } \left[ \text{EXT } \boxed{1}, \text{INT1 } \boxed{2}, \text{INT2 } \boxed{3} \right] \right]$

EXT が無い場合は INT1 が主語に、INT2 がもしあれば目的語にリンクされ(a), EXT がある場合はそれが主語、残りの内項が目的語にリンクされる(b).

動詞階層に (21) の項構造による分類 ARG-ST に加え、動詞の埋め込みの有無による分類 V-EMB を導入する. **A**, **B**, **C** の階層における位置は (24) のようになる.



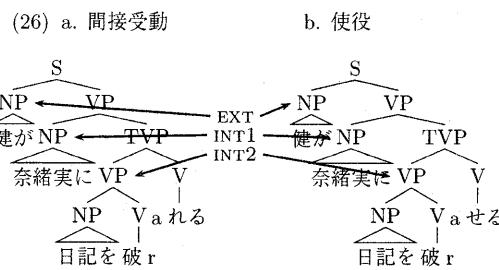
Vemb では、動詞 ( $V_1$ ) を埋めこむ動詞 ( $V_2$ ) の内項の 1 つが動詞句 ((25) の  $\boxed{1}$ ) であること、 $V_1$  と  $V_2$  が隣接すること (ADJacency) が制約として導入される (25).<sup>9</sup>

- (25)  $\left[ \text{SYN} | \text{HEAD} | \text{ADJ } \boxed{1} \left[ \text{phrase} \left[ \text{SYN} | \text{HEAD } v \right] \right], \text{ARG-ST } | \text{INT1(or 2) } \boxed{1} \right]$

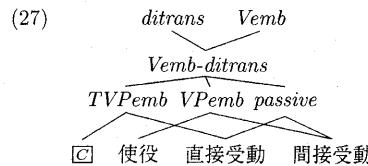
<sup>9</sup>HEAD 素性は head daughter が mother と共有する情報を表す.

ADJ 素性は理論的には形態音韻的制約 (Gunji 1999) により参照されるのだが、この形態音韻的制約は、本稿で実装の際に用いた文法開発環境が語順と統語構造を切り離して扱うことができないため、実装はできない.

直接受動動詞「られ」は、**C** と同じ項構造で、(9) にあるように、EXT に主語、INT1 に目的語、INT2 に TVP (受動化される動詞語幹) を取る。よって直接受動動詞の「られ」は **C** と階層の同じ位置に存在することになる。しかし間接受動動詞の「られ」や使役動詞の「させ」は INT2 に TVP ではなく VP を取ることから (26)、動詞階層において、**C** と直接受動の位置と、間接受動と使役の位置を区別する必要がある。

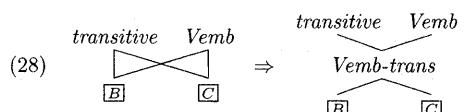


そこで、動詞階層の *ditrans* & *Vemb* 以下を (27) のように拡張する。



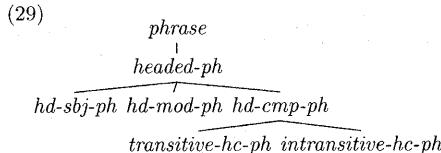
**C** と直接受動は *Vemb-ditrans* と *TVPemb* から制約を引き継ぎ、間接受動と使役は *Vemb-ditrans* と *VPemb* から制約を引き継ぐ。なお、直接間接両受動は *passive* からも制約を引き継ぐ。

HPSG 文法を実装するにはそのタイプ階層の形式化を厳密なものとすることが要求される。タイプ階層は数学的にはグラフとして捉えることができるが、well-formed なものにするには、greatest lower bound (glb) を必ず一意に定めなくてはならない。そこで、いくつかの中間タイプを設定し glb が一意に定まるようにした。例えば (24) では *transitive* と *Vemb* の glb が一意に定まらないので、中間タイプ *Vemb-trans* を導入する。



### 3.4.2 句の階層

句階層は基本的に Sag and Wasow (1999) に従う。しかし本文法は TVP のような「(主語と)目的語が 1 つ未飽和な句」という概念を用いるので、主語と目的語が 1 つ未飽和な句を “Transitive Phrase”，主語が未飽和で目的語が全て飽和した句を “Intransitive Phrase” と名付け、句階層を (29) のように定義する。



*phrase* で意味的構成性原理が導入される。意味的構成性原理は、mother の意味は全ての daughter の意味を足し合わせたものである、ということを表す原理である。<sup>10</sup>

$$(30) \left[ \begin{array}{l} \text{SEM} \mid \text{RESTR } \boxed{1} \oplus \cdots \oplus \boxed{n} \\ \text{DTRS} \left\langle \left[ \text{RESTR } \boxed{1} \right], \dots, \left[ \text{RESTR } \boxed{n} \right] \right\rangle \end{array} \right]$$

*headed-ph* で主辞素性原理と意味的継承原理 (31) が導入される。主辞素性原理は、mother と head daughter の間で HEAD 値が共有されることを、意味的継承原理は、mother と head daughter の間で INDEX 値が共有されることを表す。<sup>11</sup>

$$(31) \left[ \begin{array}{l} \text{SYN} \mid \text{HEAD } \boxed{1}, \text{ SEM} \mid \text{INDEX } \boxed{2}, \\ \text{HD-DTR} \left[ \text{SYN} \mid \text{HEAD } \boxed{1}, \text{ SEM} \mid \text{INDEX } \boxed{2} \right] \end{array} \right]$$

## 4 統語的複合動詞の処理の実装

### 4.1 文法開発環境 LKB

理論的考察を踏まえて開発した日本語 HPSG 文法を計算機上に実装するために、文法開発環境 LKB を用いる。LKB はスタンフォード大学 CSLI で LinGO プロジェクトの一環として継続的にメンテナンスされている文法開発用ソフトウェアである。LKB はタイプ付き素性構造の形式に従った文法が入力されると、その文法に従って構文解析、文生成を行う。

### 4.2 実装のための再形式化

実装の際には LKB による制限のため文法の形式化をいくつか変更した (橋本 2003a)。

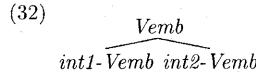
#### 4.2.1 動詞の階層

*Vemb*において、その下位タイプの動詞は内項に動詞を持つことが制約として導入されたが、内項は 2 つあり、どちらに動詞を持つのかを明示しなくては

<sup>10</sup>DTRS は句の構成素を、RESTR は意味的制限を表す。

<sup>11</sup>HD-DTR は head daughter を、INDEX は mother または daughter の意味的な指示対象を表す。

ならない。そこで、*Vemb* の下位タイプとして *int1-Vemb* と *int2-Vemb* を導入した (図は前者、団は後者)。



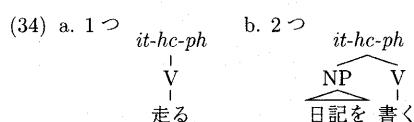
項具現化原理は、(23) の形式化をそのまま実装することはできないので、外項内項の有無に関する 6 つのリンクの可能性を別々に分けて実装した。

- (33) a.  $\boxed{\text{aryless-arp-lexeme}}$   
SYN [SBJ ⟨ ⟶, CMP1 ⟨ ⟶, CMP2 ⟨ ⟶ ]  
ARG-ST [EXT ⟨ ⟶, INT1 ⟨ ⟶, INT2 ⟨ ⟶ ]
- b.  $\boxed{\text{monounac-arp-lexeme}}$   
SYN [SBJ  $\boxed{1}$ , CMP1 ⟨ ⟶, CMP2 ⟨ ⟶ ]  
ARG-ST [EXT ⟨ ⟶, INT1  $\boxed{1}$ , INT2 ⟨ ⟶ ]
- c.  $\boxed{\text{diunac-arp-lexeme}}$   
SYN [SBJ  $\boxed{1}$ , CMP1  $\boxed{2}$ , CMP2 ⟨ ⟶ ]  
ARG-ST [EXT ⟨ ⟶, INT1  $\boxed{1}$ , INT2  $\boxed{2}$  ]
- d.  $\boxed{\text{unergative-arp-lexeme}}$   
SYN [SBJ  $\boxed{1}$ , CMP1 ⟨ ⟶, CMP2 ⟨ ⟶ ]  
ARG-ST [EXT  $\boxed{1}$ , INT1 ⟨ ⟶, INT2 ⟨ ⟶ ]
- e.  $\boxed{\text{monotrans-arp-lexeme}}$   
SYN [SBJ  $\boxed{1}$ , CMP1  $\boxed{2}$ , CMP2 ⟨ ⟶ ]  
ARG-ST [EXT  $\boxed{1}$ , INT1  $\boxed{2}$ , INT2 ⟨ ⟶ ]
- f.  $\boxed{\text{ditrans-arp-lexeme}}$   
SYN [SBJ  $\boxed{1}$ , CMP1  $\boxed{2}$ , CMP2  $\boxed{3}$  ]  
ARG-ST [EXT  $\boxed{1}$ , INT1  $\boxed{2}$ , INT2  $\boxed{3}$  ]

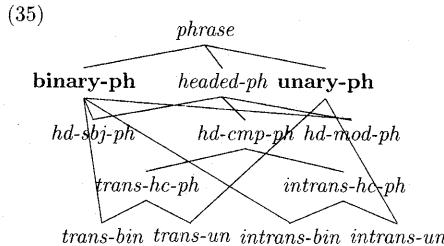
(25) で隣接制約を表す ADJ 素性について触れたが、郡司 (2000) のようにこの素性を SBJ, CMP 同様結合価素性の 1 つとして扱うと、(33)において、内項 (INT1 or INT2) が結合価 ADJ にマッピングされる場合を別に考慮しなくてはならず、項具現化原理がさらに複雑になる。本稿では ADJ を主辞素性として扱うことによりこれを避けている。

#### 4.2.2 句の階層

実装に際して、句はその構成素の数が明示されなくてはならない。本文法では、例えば *intransitive-hc-ph* の構成素が 1 つの場合と 2 つの場合を別に規定しなくてはならない。

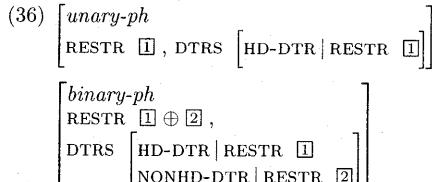


そこで句階層に *unary-phrase* と *binary-phrase* を導入し、また *transitive-hc-ph* と *intransitive-hc-ph* を各々 *unary* と *binary* に分ける。

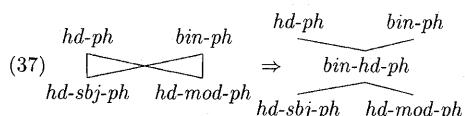


*transitive-bin-hc-ph* は 3 項動詞で使われるが、3 項動詞は先に間接目的語と結びつくようにした。<sup>(25)</sup> で形態音韻的制約が実装できなかったことについて触れたが、間接目的語が先に来るようにならため、<sup>(26)</sup> 型において  $V_1$  と  $V_2$  が隣接することを保障している。<sup>12</sup>

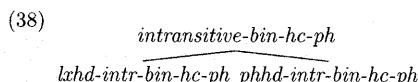
また、意味的構成性原理も *unary* と *binary* の場合で分けなくてはならない。



動詞階層と同様に、句階層においても glb は一意に定めなくてはならない。よって句階層でもいくつかの中間タイプを導入する。例えば、(35) のままでは *headed-phrase* と *binary-phrase* の glb が一意に定まらないので、中間タイプ *bin-hd-phrase* を導入する。



また、*intransitive-bin-hc-ph* は主辞が V (word) の場合 (*lexical head*) と TVP (*transitive-bin-hc-ph*) の場合 (*phrasal head*) <sup>13</sup> の 2 通りがある。



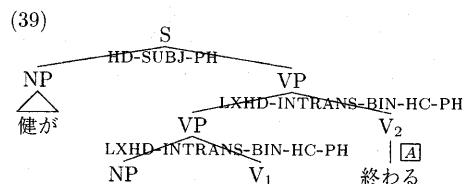
#### 4.3 解析例

以下に統語、意味構造の解析の具体例を示す。意味構造は Minimal Recursion Semantics (MRS) に基づく (Copestake, Flickinger, Sag, & Pollard 1999)。

<sup>12</sup> **A**, **B** 型では CMP1 に埋め込み動詞が来るが、CMP2 が空で、かつ、CMP1 が SBJ より先に動詞と結びつくので問題無い。

<sup>13</sup> **C** 型の  $V_2$  と、受動、使役の場合。

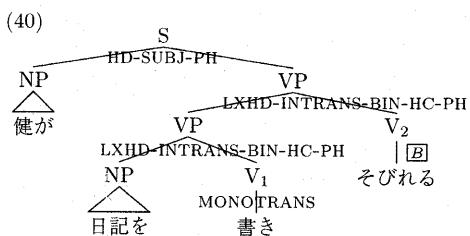
健が日記を書き終わる (6)



INDEX  $e_1$   
RSTR  $\left\langle \begin{bmatrix} ken\_rel \\ ARG0 x_2 \end{bmatrix} \begin{bmatrix} nikki\_rel \\ ARG0 x_3 \end{bmatrix} \begin{bmatrix} kak\_rel \\ ARG0 e_4 \\ ARG1 x_2 \\ ARG2 x_3 \end{bmatrix} \begin{bmatrix} owar\_rel \\ ARG0 e_1 \\ ARG1 e_4 \end{bmatrix} \right\rangle$

「 $V_1$ -終わる」は **A** タイプの複合であり、 $V_1$  「書く」の主語が  $V_2$  「終わる」の主語と構造を共有する繰り上げ構造である。よって「終わる」は「書く」の指示する事象  $e_4$  には  $\theta$  付与するが、主語の「健」にはしない。つまり **A** は意味的には 1 項動詞となる。

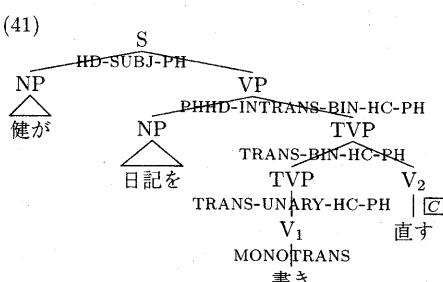
健が日記を書きそびれる (7)



INDEX  $e_1$   
RSTR  $\left\langle \begin{bmatrix} ken\_rel \\ ARG0 x_2 \end{bmatrix} \begin{bmatrix} nikki\_rel \\ ARG0 x_3 \end{bmatrix} \begin{bmatrix} kak\_rel \\ ARG0 e_4 \\ ARG1 x_2 \\ ARG2 x_3 \end{bmatrix} \begin{bmatrix} sobirer\_rel \\ ARG0 e_1 \\ ARG1 x_2 \\ ARG2 e_4 \end{bmatrix} \right\rangle$

「 $V_1$ -そびれる」は **B** タイプの複合であり、 $V_1$  「書く」の主語を  $V_2$  「そびれる」の主語がコントロールする構造である。「そびれる」は「書く」事象  $e_4$  と主語「健」  $x_2$  に  $\theta$  付与し、統語的にも意味的にも 2 項動詞である。

健が日記を書き直す (8)



INDEX e <sub>1</sub>	$\left[ \begin{array}{c} \text{ken\_rel} \\ \text{ARG0 } x_2 \end{array} \right] \left[ \begin{array}{c} \text{nikki\_rel} \\ \text{ARG0 } x_3 \end{array} \right] \left[ \begin{array}{c} \text{kak\_rel} \\ \text{ARG0 } e_4 \\ \text{ARG1 } x_2 \\ \text{ARG2 } x_3 \end{array} \right] \left[ \begin{array}{c} \text{naos\_rel} \\ \text{ARG0 } e_1 \\ \text{ARG1 } x_2 \\ \text{ARG2 } x_3 \\ \text{ARG3 } e_4 \end{array} \right]$
RSTR	$\left\langle \left[ \begin{array}{c} \text{ken\_rel} \\ \text{ARG0 } x_2 \end{array} \right] \left[ \begin{array}{c} \text{nikki\_rel} \\ \text{ARG0 } x_3 \end{array} \right] \left[ \begin{array}{c} \text{kak\_rel} \\ \text{ARG0 } e_4 \\ \text{ARG1 } x_2 \\ \text{ARG2 } x_3 \end{array} \right] \left[ \begin{array}{c} \text{naos\_rel} \\ \text{ARG0 } e_1 \\ \text{ARG1 } x_2 \\ \text{ARG2 } x_3 \\ \text{ARG3 } e_4 \end{array} \right] \right\rangle$

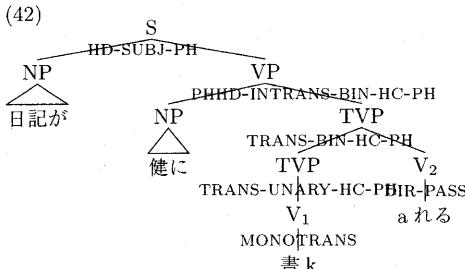
(43) は長距離受け身の構造であり、「られ」は統語的には3項動詞だが、意味的には複合動詞で表される事象 e<sub>5</sub> のみをとる1項動詞である。

## 5 結論

本稿では長距離受け身などの統語的 V<sub>1</sub>-V<sub>2</sub> 複合動詞に関わる現象を分析し、それに基づき日本語 HPSG 文法を形式化、実装した。実装に際して多少の再形式化が必要になったが、理論言語学の知見をほぼ忠実に実装に反映することに成功した。

「V<sub>1</sub>-直す」は□タイプの複合で、V<sub>1</sub>「書く」の主語と目的語をV<sub>2</sub>「直す」の主語と目的語がコントロールする。「直す」は「健」x<sub>2</sub>, 「日記」x<sub>3</sub>, 「書く」事象 e<sub>4</sub>にθ付与する3項動詞である。

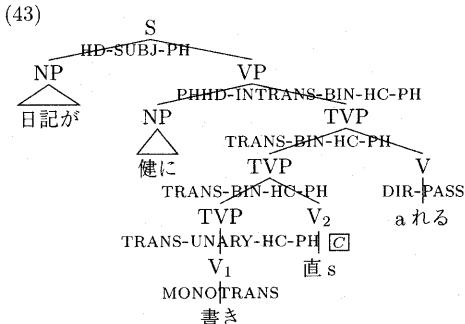
日記が健に書かれる (9a)



INDEX e <sub>1</sub>	$\left[ \begin{array}{c} \text{nikki\_rel} \\ \text{ARG0 } x_2 \end{array} \right] \left[ \begin{array}{c} \text{ken\_rel} \\ \text{ARG0 } x_3 \end{array} \right] \left[ \begin{array}{c} \text{kak\_rel} \\ \text{ARG0 } e_4 \\ \text{ARG1 } x_3 \\ \text{ARG2 } x_2 \end{array} \right] \left[ \begin{array}{c} \text{rare\_d\_rel} \\ \text{ARG0 } e_1 \\ \text{ARG1 } e_4 \end{array} \right]$
RSTR	$\left\langle \left[ \begin{array}{c} \text{nikki\_rel} \\ \text{ARG0 } x_2 \end{array} \right] \left[ \begin{array}{c} \text{ken\_rel} \\ \text{ARG0 } x_3 \end{array} \right] \left[ \begin{array}{c} \text{kak\_rel} \\ \text{ARG0 } e_4 \\ \text{ARG1 } x_3 \\ \text{ARG2 } x_2 \end{array} \right] \left[ \begin{array}{c} \text{rare\_d\_rel} \\ \text{ARG0 } e_1 \\ \text{ARG1 } e_4 \end{array} \right] \right\rangle$

直接受動動詞「られ」は「られ」の主語と目的語が動詞語幹の目的語と主語をそれぞれコントロールする。統語的には主語と目的語、動詞語幹を取る3項動詞だが、意味的には動詞語幹が表す事象 e<sub>4</sub>のみを取る1項動詞である。

日記が健に書き直される (9b)



INDEX e <sub>1</sub>	$\left[ \begin{array}{c} \text{nikki\_rel} \\ \text{ARG0 } x_2 \end{array} \right] \left[ \begin{array}{c} \text{ken\_rel} \\ \text{ARG0 } x_3 \end{array} \right]$
RSTR	$\left\langle \left[ \begin{array}{c} \text{kak\_rel} \\ \text{ARG0 } e_4 \\ \text{ARG1 } x_3 \\ \text{ARG2 } x_2 \end{array} \right] \left[ \begin{array}{c} \text{naos\_rel} \\ \text{ARG0 } e_5 \\ \text{ARG1 } x_3 \\ \text{ARG2 } x_2 \end{array} \right] \left[ \begin{array}{c} \text{rare\_d\_rel} \\ \text{ARG0 } e_1 \\ \text{ARG1 } e_5 \end{array} \right] \right\rangle$

## 参考文献

- Copestake, A. (2002). *Implementing Typed Feature Structure Grammars*. CSLI Publications.
- Copestake, A., Flickinger, D., Sag, I. A., & Pollard, C. (1999). Minimal Recursion Semantics: An introduction. draft.
- Gunji, T. (1987). *Japanese Phrase Structure Grammar*. D. REIDEL PUBLISHING COMPANY.
- Gunji, T. (1999). On Lexicalist Treatments of Japanese Causatives. In Levine & Green, G. (Eds.), *Studies in Contemporary Phrase Structure Grammar*. Cambridge University Press.
- Hasegawa, N. (1980). The VP constituent in Japanese. In *Linguistic Analysis*.
- Hinds, J. (1973). On the status of the VP node in Japanese. In *Language Research (Seoul)*.
- Mitsuishi, Y., Torisawa, K., & Tsuji, J. (1998). HPSG-Style Underspecified Japanese Grammar with Wide Coverage. In *Proceedings of COLING-ACL '98*.
- Sag, I. A. & Wasow, T. (1999). *Syntactic Theory: A Formal Introduction*. CSLI Publications.
- 影山太郎 (1993). 『文法と語形成』. ひつじ書房.
- 松本裕治 (2000). 「HPSGにもとづく日本語文法について—実装に向けての精緻化・拡張—」. 『自然言語処理』, Vol.7, No.5, pp.19-49】.
- 郡司隆男 (2000). 「日本語の動詞の形態素とそれに基づくタグ付け」. 郡司隆男 (編), 『制約に基づく文法の連続量の概念を取り入れた拡張の研究』. 平成9~11年度科学研 究費補助金.
- 橋本力 (2003a). 「計算機上で動作する日本語 HPSG 文法の構築」. 『言語処理学会第9回年次大会発表論文集』.
- 橋本力 (2003b). 「長距離受け身の HPSG に基づく分析」. 『日本言語学会第126回研究大会予稿集』.
- 井上和子 (1976). 『変形文法と日本語』. 大修館書店.

## 補遺: V<sub>2</sub> の例

- Ⓐ: 得る, 終わる, かける, きる, 過ぎる, だす, 続ける, 始める, まくる
- Ⓑ: 飽きる, かねる, 遅れる, そこなう, そこねる, そびれる, つける
- Ⓒ: 合う, 終える, かえす, 尽くす, 続ける, 直す, 憇れる, 抜く, 残す, 始める, 間違う, 間違える, 忘れる