

日本語単一化文法による形態素解析と構文解析の融合*

吉村賢治**

Integration of Morphological Analysis and Syntactic Analysis using Japanese Unification Grammar

Kenji YOSHIMURA**

When we analyze Japanese sentences, we usually adopt morphological analysis and syntactic analysis sequentially. Although one of the main purpose of morphological analysis is disclosing a structure of word formation, the units used in word formation are not only morphemes but also phrases, which we recognize in syntactic analysis. So we can't divide the analyzing process of Japanese sentences into morphological analysis and syntactic analysis. In this paper, I introduce the unification grammar of Japanese, which enables us to accomplish morphological analysis and syntactic analysis in a single framework.

Key Words : Japanese unification grammar, word formation, morphological analysis, syntactic analysis, feature structure

1. はじめに

従来の日本語文の解析では、最初に入力された文字列を単語の列に変換する形態素解析を行い、その結果として出力される単語列に対して構文解析を行う段階的な手法が用いられている。しかし、形態素解析の処理対象である接辞による語形成の中には、統語的単位である句に対して適用されるものがある [1]。例えば、次の文 (1) と (2) はどちらも日本語の適格文である。

- (1) ウナギが食べたい
- (2) ウナギを食べたい

文 (1) と (2) は、それぞれ Fig. 1 に示す (a)、(b) の構造を持つ。(a) では、動詞 "tabe(食べ)" に接辞 "itai(たい)" が付いて形容詞 "tabetai(食べたい)" を形成し、形容詞 "tabetai" が後置詞句 "unagiga(ウナギが)" を補語としてとる。それに対して、(b) では動詞 "tabe(食べ)" と後置詞句 "unagiwo(ウナギを)" が結合して動詞句 "unagiwo tabe(ウナギを食べ)" になり、動詞句に接辞 "itai(たい)" が付いて形容詞を形成している。なお、ここでは学校文法における動詞などの活用を音韻変化で説明する文法 [2] を用いており、形容詞化接辞 "itai" の形成過程は省略している。

本稿では、このように形態論と統語論に跨った現象も一つの枠組みの中で処理できることを目的とした日本語の単一化文法とそこで用いる素性構造を提案する。本研究では Fig. 2 に示すようなシステムの構成を想定している。Fig. 2 において、レキシコンは見出しが形態素

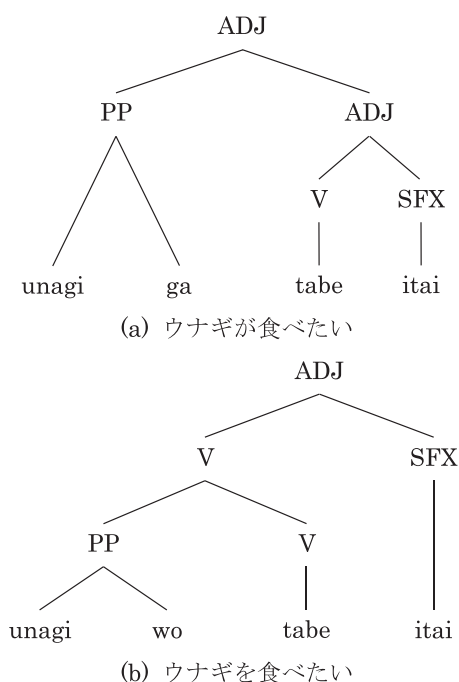


Fig. 1 Syntactic trees of sentence (1) and (2)

* 平成 28 年 11 月 28 日受付

** 電子情報工学科

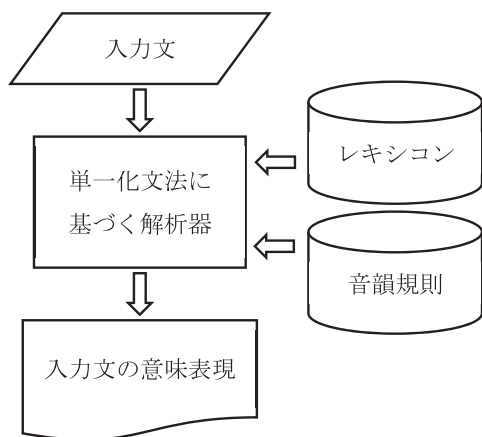


Fig. 2 System overview

でその素性構造を辞書内容とする単語辞書や生成辞書 (generative lexicon) [3] からの辞書群であり、音韻規則は形態素が接続したときに生じる音韻変化の規則から生成する有限状態トランスデューサ (FST) [4] である。

2. 単一化文法

単一化文法は、文脈自由文法における非終端記号を単なる記号ではなく、その範疇がもつ文法的な属性や意味的な属性などを素性と素性値の対を要素とする集合で表現する [5]。本研究では、代表的な日本語の単一化文法である日本語句構造文法 (JPSG : Japanese Phrase Structure Grammar) [6] に基づき、その原理と素性構造を拡張して、日本語文の解析を形態素解析と構文解析の2段階に区別せずに行うことができる単一化文法の構築を目指している。

JPSG は (3) に示す一つの書き換え規則と素性構造の単一化に関する幾つかの原理で構成される。

(3) $M \rightarrow CH$

ここで、 M 、 C 、 H はそれぞれ素性構造で、 M を親、 C と H を子とよぶ。特に、 H は日本語における右側主要部の規則 (Righthand Head Rule) から主辞 (head) とよび、左側の子 C と区別する。

素性には Table 1 に示すようなものがある。 pas は受け身化の可否をで示す二値素性である。これに対して、 pos 、 $pform$ 、 gr などは複数個の値の中の一つを値とする多値素性である。 pos 、 $pform$ 、 gr は主辞素性であり、head 素性の値である素性構造を構成する要素となる。 sem は述語論理式や素性構造など、何らかの形式で表された意味表現を保持する素性である。 $subcat$ と $adjacent$ は主辞が必要とする補語の集合を表現している下位範疇化素性とよばれる素性である。 $adjunct$ の値はその範疇が修飾する隣接する主辞の素性構造である。以上の素性が局所素性 (local feature) とよばれるのに対して、 $slash$ 、 $refl$ は非局所素性 (nonlocal feature) とよばれる。局所素性の値は主辞とその親との間で部分的に共有されるのに

Table 1 Features and their values

素性名	値	備考
pas	+,-	受け身化の可否
pos	v, n, p, ...	品詞
$pform$	ga, wo, ni, ...	格助詞
gr	sbj, obj	述語との関係
sem	任意の形式的表現	意味表現
$subcat$	素性構造の集合	下位範疇化
$adjacent$	素性構造	隣接を要求
$adjunct$	素性構造	修飾先の主辞
$slash$	素性構造の集合	空所を表す
$refl$	pp[sbj]	再帰代名詞の先行詞

対して、非局所素性の値は補語や付加語と親の間でも共有される。非局所素性は、文の中の離れた句との関係を記述するために用いられ、 $slash$ は主題文における空所を表すために、 $refl$ は再帰代名詞の先行詞を表すために用いられる。

JPSG の代表的な原理に主辞素性の原理と下位範疇化素性の原理がある。

主辞素性の原理 (head 素性)

親 M の主辞素性の値は主辞 H の主辞素性の値と単一化する。

下位範疇化素性の原理 (subcat 素性, adjacent 素性)

- ① 補語構造の場合、親 M の下位範疇化素性の値は、主辞 H の下位範疇化素性の値から子 C (補語) と単一化可能な素性構造を取り除いたものと単一化する。
- ② 付加構造の場合、親 M の下位範疇化素性の値は、主辞 H の下位範疇化素性の値と単一化する。
- ③ 等位構造の場合、親 M の下位範疇化素性の値は、子 C と主辞 H の下位範疇化素性の値と単一化する。

基本的な日本語文の記述に必要な原理としては、その他に以下のようなものがある。

意味素性の原理 (sem 素性)

- ① 補語構造の場合、親 M の意味素性の値は、主辞 H の意味素性の値と単一化する。
- ② 付加構造の場合、親 M の意味素性の値は、付加語 C の意味素性の値と単一化する。

非局所素性の原理 (slash 素性, refl 素性)

親 M の非局所素性の値は子 C と H の非局所素性の値の和集合からその句で束縛される素性構造を取り除いたものと単一化する。

付加素性の原理 (adjunct 素性)

親 M の非局所素性の値は子 C の付加素性の値から主

辞Hと単一化可能な素性構造を取り除いた値と単一化する。

例として、名詞「犬」と格助詞「が」、動詞「吠e(吠える)」、接辞「ita(た)」が、それぞれ辞書項目として次のような素性構造をもつとする。

$$(4) \text{ 犬} : \left[\begin{array}{l} \text{head: } [\text{pos: } n] \\ \text{sem: } \left[\begin{array}{l} \text{denote: } X \\ \text{restriction: } \{\text{dog}(X)\} \end{array} \right] \end{array} \right]$$

$$(5) \text{ が} : \left[\begin{array}{l} \text{head: } \left[\begin{array}{l} \text{pos: } p \\ \text{pform: } ga \\ \text{gr: } sbj \end{array} \right] \\ \text{sem: } X \\ \text{adjacent: } \left\{ \left[\begin{array}{l} \text{head: } [\text{pos: } n] \\ \text{sem: } X \end{array} \right] \right\} \end{array} \right]$$

$$(6) \text{ 吠e} : \left[\begin{array}{l} \text{head: } [\text{pos: } v] \\ \text{sem: } \left[\begin{array}{l} \text{denote: } \left[\begin{array}{l} \text{relation: } bark \\ \text{agent: } X \end{array} \right] \\ \text{restriction: } XR \end{array} \right] \\ \text{subcat: } \left\{ \left[\begin{array}{l} \text{head: } \left[\begin{array}{l} \text{pos: } p \\ \text{pform: } ga \\ \text{gr: } sbj \end{array} \right] \\ \text{sem: } \left[\begin{array}{l} \text{denote: } X \\ \text{restriction: } XR \end{array} \right] \\ \text{adjacent: } \{ \} \end{array} \right] \right\} \end{array} \right]$$

$$(7) \text{ ita} : \left[\begin{array}{l} \text{head: } [\text{pos: } v] \\ \text{sem: } \left[\begin{array}{l} \text{denote: } \left[\begin{array}{l} \text{tens: } past \\ \text{process: } X \end{array} \right] \\ \text{restriction: } XR \end{array} \right] \\ \text{adjacent: } \left\{ \left[\begin{array}{l} \text{head: } [\text{pos: } v] \\ \text{sem: } \left[\begin{array}{l} \text{denote: } X \\ \text{restriction: } XR \end{array} \right] \\ \text{subcat: } \{ \} \end{array} \right] \right\} \end{array} \right]$$

ここで、素性の名称等は概ね JPSG の記法に従っているが、sem 素性の内容などでは本研究独自の素性名を使用しているところもある。また、sem 素性の値を記述している restriction 素性の値は述語論理式の集合で記述しているが、述語論理式は素性構造の略記法として使用している。例えば $\text{dog}(X)$ は素性構造で表現すると

$$(8) \left[\begin{array}{l} \text{predicate: } dog \\ \text{arg: } X \end{array} \right]$$

となる。素性構造の記述において、大文字で始まる X や XR は変数を表している。変数の有効範囲 (scope) はその変数を含む一番外側の素性構造の内部だけである。

これらの素性構造に基づいて文 (9) の解析過程を説明する。

(9) 犬が吠えた

文 (9) は次の形態素列から構成される。

(10) 犬 + ga + 吠e + ita

まず、(5) の adjacent 素性の値と (4) の素性構造を単一化し、下位範疇化素性の原理と主辞素性の原理、意味素性の原理に従って (11) が後置詞句「犬が」の素性構造として得られる。

$$(11) \text{ 犬が} : \left[\begin{array}{l} \text{head: } \left[\begin{array}{l} \text{pos: } p \\ \text{pform: } ga \\ \text{gr: } sbj \end{array} \right] \\ \text{sem: } \left[\begin{array}{l} \text{denote: } X \\ \text{restriction: } \{\text{dog}(X)\} \end{array} \right] \\ \text{adjacent: } \{ \} \end{array} \right]$$

後置詞句 (11) の素性構造は動詞 (6) の subcat 素性の値である集合の要素と単一化し、下位範疇化素性の原理と主辞素性の原理、意味素性の原理に従って (12) が動詞句「犬が吠e」の素性構造として得られる。

$$(12) \text{ 犬が吠e} : \left[\begin{array}{l} \text{head: } [\text{pos: } v] \\ \text{sem: } \left[\begin{array}{l} \text{denote: } \left[\begin{array}{l} \text{relation: } bark \\ \text{agent: } X \end{array} \right] \\ \text{restriction: } \{\text{dog}(X)\} \end{array} \right] \\ \text{subcat: } \{ \} \end{array} \right]$$

動詞句 (12) の素性構造は過去の接辞 (7) の adjacent 素性の値である集合の要素と単一化し、下位範疇化素性の原理と主辞素性の原理、意味素性の原理に従って (13) が動詞句「犬が吠えた」の素性構造として得られる。

$$(13) \text{ 犬が吠えた} : \left[\begin{array}{l} \text{head: } [\text{pos: } v] \\ \text{sem: } \left[\begin{array}{l} \text{denote: } \left[\begin{array}{l} \text{tens: } past \\ \text{process: } X \end{array} \right] \\ \text{restriction: } \{\text{dog}(X)\} \end{array} \right] \\ \text{adjacent: } \{ \} \end{array} \right]$$

文 (9) の意味表現を完成するためには、この後、断定を示す零の様相辞の adjacent 素性値と単一化をおこなって素性構造 (14) を求め、

$$(14) \left[\begin{array}{l} \text{head: } [\text{pos: } v] \\ \text{sem: } \left[\begin{array}{l} \text{denote: } \left[\begin{array}{l} \text{modal: } declarative \\ \text{event: } \left[\begin{array}{l} \text{tens: } past \\ \text{process: } X \end{array} \right] \end{array} \right] \\ \text{restriction: } \{\text{dog}(X)\} \end{array} \right] \\ \text{adjacent: } \{ \} \end{array} \right]$$

文脈から主題 (ここでは主題を変数 Y で表し、 Y が満たすべき制約を変数 YR で表す) を補完して素性構造像 (15) を求める。

$$(15) \left[\begin{array}{l} \text{head: } [\text{pos: } v] \\ \text{sem: } \left[\begin{array}{l} \text{denote: } \left[\begin{array}{l} \text{topic: } Y \\ \text{comment: } \left[\begin{array}{l} \text{modal: } declarative \\ \text{event: } \left[\begin{array}{l} \text{tens: } past \\ \text{process: } X \end{array} \right] \end{array} \right] \end{array} \right] \\ \text{restriction: } \{\text{dog}(X)\} \cup YR \end{array} \right] \\ \text{adjacent: } \{ \} \end{array} \right]$$

最後に、文 (9) の場合は情報の提供 (inform) を表す零の陳述辞が添加されて、head 素性として素性構造 (16)、sem 素性の denote 素性として素性構造 (17) をもつ素性構造が得られる [6]。

(16) [pos: s]

$$(17) \left[\begin{array}{l} \text{mood: } inform \\ \text{judgement: } \left[\begin{array}{l} \text{topic: } Y \\ \text{comment: } \left[\begin{array}{l} \text{modal: } declarative \\ \text{event: } \left[\begin{array}{l} \text{tens: } past \\ \text{process: } X \end{array} \right] \end{array} \right] \end{array} \right] \end{array} \right]$$

3. 形態論のための素性

3.1 音韻規則

本研究で用いている音韻論的な分析に基づく形態論では、学校文法における活用形による活用語の語尾変化を音韻変化で説明する。基本的な音韻変化の規則は、子音脱落と母音脱落である。

子音脱落

形態素の接続時に子音が連続した場合、後接の子音が脱落する。

母音脱落

形態素の接続時に母音が連続した場合、後接の母音が脱落する。

例えば、動詞「kak (書く)」に非完了体を表す文法接辞「ru (る)」が接続した場合、後接の子音 r が脱落して

(17) kak + ru → kak 0 0u

となる。ここで、「+」は接続を表す記号で、「0」は空の文字を表す記号である。また、動詞「mi (見る)」に完了体を表す文法接辞「ita (た)」が接続した場合、後接の母音 i が脱落して

(18) mi + ita → mi 0 0ta

となる。

日本語の代表的な音韻変化は音便形に関連している。

イ音便

イ音便は k と g で終わる子音動詞（「行く」は例外的に促音便する）に完了体を表す文法接辞「ita (た)」などが接続した場合に発生し、

(19) kak + ita → ka0 0 ita (書いた)

(20) oyog + ita → oyo0 0 ida (泳いだ)

のように変化する。

撥音便

撥音便は n と m で終わる子音動詞に完了体を表す文法接辞「ita (た)」などが接続した場合に発生し、

(21) sin + ita → sin 0 0da (死んだ)

(22) yom + ita → yon 0 0da (読んだ)

のように変化する。

促音便

促音便は t と r, w で終わる子音動詞と子音動詞 ik (行く) に完了体を表す文法接辞「ita (た)」などが接続した場合に発生し、

(23) tat + ita → ta0 0 tta

(24) kir + ita → ki0 0 tta

(25) kaw + ita → ka0 0 tta

(26) ik + ita → i0 0 tta

のように変化する。

2階層モデルを実装した PC-KIMMO [7] の書式では、子音 k で終わる子音動詞に対するイ音便の規則は次のように書ける。

(27) k:0 ⇒ _ +:0 i:i

ここで、書式 L:S は語彙レベルの文字 L と表層レベル

の文字 S の対を表している。+ は形態素の境界を表す記号、0 は空記号である。この規則は、語彙レベル +i の前にある k は脱落することがあることを表している。同様に子音脱落、母音脱落の規則は、それぞれ次のように書ける。

(28) R:0 ⇔ C:@ +:0 _

(29) A:0 ⇔ V:@ +:0 _

ここで、R={r, s, y}、C={k, g, s, t, n, b, m, r, w}、A={a, i}、V={i, e} であり、@ はすべての文字と一致するワイルドカードである。⇔ は左辺の変化が起きるのは右辺の環境の中であり、そのときだけであることを表している。

このように2階層モデルの書式では音韻変化とそれが発生する前後の文字列で規則を表現するが、音便変化で説明したように子音脱落、母音脱落以外の日本語の音韻変化の発生は特定の形態素が接続したときだけに限られる。そこで本研究では音韻変化の規則に形態素のカテゴリに対する制約を付加して、

(30) 音韻変化 <前カテゴリ + 後カテゴリ>

の形式で記述する。この表記を使うとイ音便の規則は次のように書くことができる。

(31) k:0 +:0 i:i <cv_k+sfx_it>

(32) g:0 +:0 i:i t:d <cv_g+sfx_it>

ここで、cv_k、cv_g はそれぞれ子音 k、子音 g で終わる子音動詞の形態素カテゴリコードで、sfx_it は文法接辞 ita や ite に対して付与される形態素カテゴリコードである。同様にして撥音便の規則は、子音 n、m で終わる子音動詞の形態素カテゴリコードをそれぞれ cv_n、cv_m とすると、

(33) n:n +:0 i:0 t:d <cv_n+sfx_it>

(34) m:n +:0 i:0 t:d <cv_m+sfx_it>

となり、促音便の規則は、子音 t、r、w で終わる子音動詞の形態素カテゴリコードをそれぞれ cv_t、cv_r、cv_w、動詞 ik (行く) の形態素カテゴリコードを cv_ik とすると、

(35) t:0 +:0 i:t <cv_t+sfx_it>

(36) r:0 +:0 i:t <cv_r+sfx_it>

(37) w:0 +:0 i:t <cv_w+sfx_it>

(38) k:0 +:0 i:t <cv_ik+sfx_it>

となる。子音脱落、母音脱落の規則は形態素カテゴリコードに対するワイルドカード # を用いて、

(39) C:@ +:0 R:0 <##>

(40) V:@ +:0 A:0 <##>

と書く。

3.2 形態素カテゴリ素性

形態論における素性は形態素の素性構造の中に morph 素性として記述する。3.1 で示した形態素カテゴリコードは morph 素性の素性値として

(41) [mcat: cv_k]

のように mcat 素性として与えることが可能であるが、「書いておく」における複合表現 ite+ok を一つの形態素

iteok として扱うことを可能にするためには文頭側の形態素カテゴリコードと文末側の形態素カテゴリコードを指定しておく必要がある。ここでは、文頭側の形態素カテゴリコードと文末側の形態素カテゴリコードをそれぞれ *lmc* 素性と *rmc* 素性として記述する。例えば、複合的な形態素 *iteok* の形態素カテゴリ素性は

$$(42) \begin{bmatrix} \textit{lmc}:\textit{sfx_it} \\ \textit{rmc}:\textit{cv_k} \end{bmatrix}$$

となる。これらの形態素カテゴリ素性は次の原理に従って子から親に伝搬する。

形態素カテゴリ素性の原理 (*lmc* 素性、*rmc* 素性)

- ① 親 M の *rmc* 素性の値は、主辞 H の *rmc* 素性の値と単一化する。
- ② 親 M の *lmc* 素性の値は、子 C の *lmc* 素性の値と単一化する。

3.3 語のレベルに関する素性

語形成を行う最小の単位は形態素であるが、形態素が結合して語形成を行う対象には、形態素によって制限が付く場合がある。例えば、形容詞を名詞化する接尾辞に「さ」と「み」がある。「さ」は「甘さ、深さ」のように単純な形容詞にも付くし、「甘酸っぱさ、奥深さ」のように複合形の形容詞にも付くが、「甘み、深み」における「み」は「甘酸っぱい」や「奥深い」には付けることができない。この現象を説明するために、影山は日本語の語形成に語根、語幹、語という3つのレベルに区別することを提案している [8]。3つのレベルにおける最小の単位は語根であり、語根と語根が結合して語幹になる。語幹は単独で語になることもできるが、語幹と結合したのも語を形成する。語と語が結合したものは語である。これを句構造文法における生成規則の形式で記述すると次のようになる。

- (43) a. 語幹 → 語根・語根
- b. 語 → 語幹
- c. 語 → 語幹・語幹
- d. 語 → 語・語

語のレベルを設定することにより、接尾辞「み」は形容詞の語根のみと結合し、「さ」は語幹以下のレベルの形容詞と結合するという制約を与えることができる。本研究でも、*morph* 素性の値として、語根であることを示す *root*、語幹であることを示す *stem*、語であることを示す *word* のいずれかを値とする *level* 素性を設定する。ここで、影山の規則 (43) は等価な (44) の規則に書き換えることができる。

- (44) a. 語幹 → 語根・語根
- b. 語 → 語根・語根
- c. 語 → 語幹・語幹
- d. 語 → 語・語

(44)a と (44)b から語根と語根が接続した場合には、語幹が形成されるときと語が形成されるときがあることが分かる。この区別を各語根で指定できるようにするために、語根の *morph* 素性に親の素性構造の *level* 素性の値を指定する *mlevel* 素性を追加する。語幹と語の *mlevel* 素性の値は *word* になる。以上のことから、語のレベル素性 (*level* 素性) の原理は次のようになる。

語のレベル素性の原理 (*level* 素性)

- ① C と H の *morph:level* 素性値は単一化可能でなければならない。
- ② M の *morph:level* 素性の値は、補語構造の場合には H の *morph:mlevel* 素性の値と単一化し、付加構造の場合には C の *morph:mlevel* 素性の値と単一化する。
- ③ C と H の *morph:level* 素性値が *root* 以外するとき、M の *morph:mlevel* 素性値は *word* と単一化する。

以上の形態論に関する素性を加えた形容詞「甘い」の語根 *ama* の素性構造を (45) に示す。

$$(45) \textit{ama}:\begin{bmatrix} \textit{head}:[\textit{pos}:\textit{adj_root}] \\ \textit{morph}:\begin{bmatrix} \textit{lmc}:\textit{adj_root} \\ \textit{rmc}:\textit{adj_root} \\ \textit{level}:\textit{root} \\ \textit{mlevel}:\textit{stem} \end{bmatrix} \\ \textit{sem}:\begin{bmatrix} \textit{denote}:\begin{bmatrix} \textit{property}:\textit{sweet} \\ \textit{arg}:\textit{X} \end{bmatrix} \\ \textit{darg}:\textit{taste} \\ \textit{restriction}:\textit{XR} \end{bmatrix} \\ \textit{p_subcat}:\left\{\begin{array}{l} \textit{head}:\begin{bmatrix} \textit{pos}:\textit{p} \\ \textit{pform}:\textit{ga} \\ \textit{gr}:\textit{sbj} \end{bmatrix} \\ \textit{sem}:\begin{bmatrix} \textit{denote}:\textit{X} \\ \textit{restriction}:\textit{XR} \end{bmatrix} \\ \textit{adjacent}:\{\} \end{array}\right\} \end{bmatrix}$$

ここで *p_subcat* 素性は本研究で導入した素性であり、潜在的な *subcat* 素性であることを表している。形容詞の語根の状態では補語をとることができない。ここでは西山による形容詞の研究 [9] を参考にして、*p_subcat* 素性は次に示す形容詞語根を述語化する繫辞 (*predicative copula*) *k* によって *subcat* 素性に変えられると考える。

$$(46) \textit{k}:\begin{bmatrix} \textit{head}:[\textit{pos}:\textit{adj_stem}] \\ \textit{morph}:\begin{bmatrix} \textit{lmc}:\textit{adjpc} \\ \textit{rmc}:\textit{adjpc} \\ \textit{level}:\textit{root} \\ \textit{mlevel}:\textit{stem} \end{bmatrix} \\ \textit{sem}:\textit{X} \\ \textit{subcat}:\textit{Y} \\ \textit{adjacent}:\begin{bmatrix} \textit{head}:[\textit{pos}:\textit{adj_root}] \\ \textit{sem}:\textit{X} \\ \textit{p_subcat}:\textit{Y} \end{bmatrix} \end{bmatrix}$$

形容詞の取り扱いに関しては稿を改めて詳しく報告する予定である。ここでは、形容詞に付いて派生名詞を形成する接尾辞「sa (さ)」と「mi (み)」の素性構造におけ

$$(47) \text{ sa: } \left[\begin{array}{l} \text{head: } [\text{pos: } n] \\ \text{morph: } \left[\begin{array}{l} \text{level: } \text{root} \\ \text{mlevel: } \text{word} \\ \text{lmcat: } \text{sfx} \\ \text{rmcat: } \text{sfx} \end{array} \right] \\ \text{sem: } \left[\begin{array}{l} \text{denote: } X \\ \text{restriction: } \cup \left(PR, \left\{ \left[\begin{array}{l} \text{property: } P \\ \text{arg: } Y \\ \text{darg: } D \\ \text{degree: } X \end{array} \right], \text{scalable}(P) \right\} \right) \end{array} \right] \\ \text{adjacent: } \left\{ \left[\begin{array}{l} \text{head: } [\text{pos: } \text{adj_root}] \\ \text{morph: } [\text{level: } \text{stem}\sqrt{\text{root}}] \\ \text{sem: } \left[\begin{array}{l} \text{denote: } \left[\begin{array}{l} \text{property: } P \\ \text{arg: } Y \\ \text{darg: } D \end{array} \right] \\ \text{restriction: } PR \end{array} \right] \end{array} \right\} \end{array} \right]$$

$$(48) \text{ mi: } \left[\begin{array}{l} \text{head: } [\text{pos: } n] \\ \text{morph: } \left[\begin{array}{l} \text{level: } \text{root} \\ \text{mlevel: } \text{word} \\ \text{lmcat: } \text{sfx} \\ \text{rmcat: } \text{sfx} \end{array} \right] \\ \text{sem: } \left[\begin{array}{l} \text{denote: } X \\ \text{restriction: } \cup \left(PR, \left\{ \left[\begin{array}{l} \text{property: } P \\ \text{arg: } X \\ \text{darg: } D \end{array} \right], \text{decomposable}(Y), \text{part_of}(X, Y) \right\}, YR \right) \end{array} \right] \\ \text{subcat: } \left\{ \left[\begin{array}{l} \text{head: } [\text{pos: } n] \\ \text{gr: } \text{no} \\ \text{sem: } \left[\begin{array}{l} \text{denote: } Y \\ \text{restriction: } YR \end{array} \right] \end{array} \right\} \\ \text{adjacent: } \left\{ \left[\begin{array}{l} \text{head: } [\text{pos: } \text{adj_root}] \\ \text{morph: } [\text{level: } \text{root}] \\ \text{sem: } \left[\begin{array}{l} \text{denote: } \left[\begin{array}{l} \text{property: } P \\ \text{arg: } X \\ \text{darg: } D \end{array} \right] \\ \text{restriction: } PR \end{array} \right] \end{array} \right\} \end{array} \right]$$

る語のレベルに関する制約について説明する。形容詞を名詞化する接尾辞 sa は pos 素性値が adj_root である形容詞の語根を補語にとって名詞化するが、morph:level 素性値は root または stem であればよい。これに対して、同じく形容詞を名詞化する接尾辞 mi も pos 素性値が adj_root である形容詞の語根を補語にとって名詞化するが、morph:level 素性値は root でなければならない。接尾辞 sa と接尾辞 mi の素性構造をそれぞれ (47) と (48) に示す。ここで、表記 $\cup(A, B)$ は集合 A と B の和集合を表している。形容詞語根に接尾辞 sa が付いて形成される派生名詞は形容詞が表している属性の「程度」(degree 素性の値)を表しており、その属性は計量可能 (scalable) である必要がある。一方、接尾辞 mi が付いて形成される派生名詞は形容詞が表している属性の「程度」を表す場合とある物 (素性構造 (48) では「の格」で与えられる Y) の属性 P を持つ部分 (X) を表す場合があるが [10]、素性構造 (48) では接尾辞 mi が付いて形成される派生名詞の代表的な意味である後者の場合を記述している。その場合、ある物 Y は部分に分けることができる (decomposable) が必要である。

素性構造 (47) において、adjacent 素性の記述で補語にとる語のレベル (level) の値はアトム stem または root と単一化できることを表している。

素性構造 (48) には subcat 素性と adjacent 素性が含まれている。同様な構造は主題化の助詞「は」の素性構造でも見ることができる [11]。主題化の助詞「は」の素性構造は (49) に示すように adjacent 素性と adjunct 素性を持つ。

$$(49) \text{ ha: } \left[\begin{array}{l} \text{head: } [\text{pos: } p] \\ \text{sem: } \left[\begin{array}{l} \text{topic: } X \\ \text{event: } Y \end{array} \right] \\ \text{adjunct: } \left\{ \left[\begin{array}{l} \text{head: } [\text{pos: } s] \\ \text{sem: } Y \end{array} \right] \right\} \\ \text{adjacent: } \left\{ \left[\begin{array}{l} \text{head: } [\text{pos: } n] \\ \text{sem: } X \end{array} \right] \right\} \end{array} \right]$$

(48) や (49) のように一つの素性構造が、素性値が空集合でない adjacent 素性と共に subcat 素性や adjunct 素性を持つ場合には、adjacent 素性の適用を優先しなければならない。そのために次に示す adjacent 素性優先の原理を設定する。

adjacent 素性優先の原理

一つの素性構造が、素性値が空集合でない adjacent 素性と共に subcat 素性や adjunct 素性を持つ場合には、adjacent 素性の適用を優先する。

4. あとがき

本稿では単一化文法の枠組みを利用して、日本語文の形態素解析と構文解析を融合するために必要となる素性構造について報告したが、ここで示した素性構造はまだ完全なものではなく、今後、接頭辞による語形成や名詞が連続した複合語の形成などに関する考察を行い、実用的な使用に耐える素性構造を完成する必要がある。また、日本語単一化文法で仮定している唯一の書き換え規則はチョムスキーの標準形であり、構文解析のアルゴリズムとしては CYK 法の利用が考えられるが、有限状態トランスデューサを用いた音韻変化の解析アルゴリズムと構文解析を行う CYK 法のアルゴリズムの効率的な融合についても検討していく必要がある。

参考文献

- [1] 伊藤たかね・杉岡洋子, 英語学モノグラフシリーズ 16・語の仕組みと語形成, 東京: 研究社, 2002.
- [2] 三浦睦美・吉村賢治・首藤公昭, “日本語形態素解析のための派生文法,” 福岡大学工学集報第 54 号, 1995.
- [3] J. Pustejovsky, *The Generative Lexicon*, Massachusetts: The MIT Press, 1995.
- [4] L. K. Kenneth R. Beesley, *Finite State Morphology*, Stanford: CSLI Publications, 2003.
- [5] S. M. Shieber, *An Introduction to Unification-Based Approaches to Grammar*, Stanford: CSLI Publications, 1993.
- [6] 郡司隆男, 自然言語, 東京: 日本評論社, 1994.
- [7] E. L. Antworth, *PC-KIMMO: A Two-level Processor for Morphological Analysis*, Dallas, TX: International Academic Bookstore, 1990.
- [8] 影山太郎, 文法と語形成, 東京: ひつじ書房, 1993.
- [9] K. Nishiyama, “Adjectives and the Copulas in Japanese,” *Journal of East Asian Linguistics* 8, 1999.
- [10] 長谷部郁子, “日本語の形容詞派生名詞の形成について,” *MLF2009*, 2009.
- [11] 吉村賢治, 自然言語処理の基礎 [改訂版], 東京: サイエンス社, 2012.