

漢字仮名混じり文形態素解析における非サ変動詞の 分割単位設定について†

久 光 徹†† 新 田 義 彦††

漢字仮名混じり文の形態素解析については、従来数多くの研究があるが、分割単位の設定自体を自然言語処理の観点から考察した例はまれである。本論文では、「同一の語根から派生した自動詞/他動詞の扱い」と「活用形の扱い」の2点に焦点を絞り、非サ変動詞の活用形と助動詞活用形の複合部の、形態素解析における分割単位について考察する。従来第1の問題は考慮されておらず、第2の問題についても、自然言語処理の立場から、複数の手法を明示的に比較した例はない。第1の問題に関しては、上記のような動詞の語幹を、語根と自動詞化接辞、または、他動詞化接辞に分割する分割単位を提案する。新たな分割単位が、意味解析への入力として妥当であることを示すため、語根へ適切な意味表記を行えば、もとの動詞の必須補語の個数や、それらの意味役割が計算できることを述べる。語根と、上記の接辞を分割単位とすることは、自動詞/他動詞の派生対が数多く存在する日本語の処理において、きわめて有意義である。第2の問題に関しては、動詞語幹の末尾の子音を屈折接辞側へ付加することにより、構文解析に貫献しない分割を避け、しかも活用形処理にかかわる辞書見出しの数を最小にできることを示した。分割単位間の接続を表明するための接続コード体系もあわせて示す。

1. はじめに

日本語は通常分かち書きが行われないので、その機械処理のためには、入力文字列を辞書の見出しに記載された単位に分割する処理が必要である。このとき、分かち書きだけでなく、分割文字列に品詞などの構文情報、意味素性などの意味情報を付与するのが、形態素解析である。形態素解析は、構文解析、意味解析の前提となる日本語処理の基本技術である。

以下本論文では、上記の分割単位を、簡単のため「形態素」と呼ぶことにする。

形態素解析については、これまでも数多くの重要な研究が行われてきており、枚挙にいとまがないが、ここでは文献5), 13) のみを代表例として挙げておく。

ところで、従来の研究においては、主として

- 1) 分割処理効率の向上
- 2) 分割で生じる曖昧性の低減

に興味集中していたといえる。そして、それらの多くは、何を形態素とするかに本質的には依存しない(正確には、依存すべきでない)研究であった。ところが、これらに先立つべき

- 3) 形態素の設定

は、純粋に国語学的な問題とみなされてきたかのようであり、文献10)などの若干の例外を除いて、自然言

語処理独自の観点からの提案はみられなかった。

本論文では、従来あまり取り上げられなかった3)の問題を、自然言語処理の立場から取り扱う。なかでも、自然言語処理のかなめとなる、動詞活用形と助動詞活用形の複合部(便宜上、これを動詞句と呼ぶ)の解析について考察したい。本論文では特に、「同一の語根(本論文中で用いる言語学用語については付録を参照)から派生した自動詞/他動詞の扱い」と「活用形の扱い」の2点に焦点を絞り、非サ変動詞を含む動詞句(便宜上、これを非サ変動詞句と呼ぶ)を解析するための形態素の設定方法を考察する。

以下第2章では、上記の2点に関して、形態素設定における問題を明確にし、形態素設定に関する二つの要請を与える。第3章では、学校文法に準じる方法、形態音韻論に準じる方法を概括し、それらと対比して新たな形態素の設定方法を述べる。そして、これらのうちで新方法だけが、第2章で与えた要請を満たすことを示す。また、接続コードの体系についても述べる。第4章では、新方法で与えた形態素を意味処理に利用するために、新方法による形態素に意味表記する際の指針を示す。第5章では要約、問題点、課題について述べる。

言語学上の形態素と自然言語処理上の形態素は必ずしも一致しないため、必要な場合は後者を「言語学上の形態素」と呼び、区別する。また、本論文では形態音韻論による分析を引用することが多いが、「派生接辞」などに対応する適切な用語が国文法にないため、

† A Proper Segmentation of Verbal Complexes in Japanese Morphological Analysis by TORU HISAMITSU and YOSHIHIKO NITTA (Advanced Research Laboratory, Hitachi, Ltd.).

†† (株)日立製作所基礎研究所

文献 2), 8) に従って用語の統一をはかった。

2. 形態素設定への二つの要請

2.1 派生対をなす動詞の扱いについて

与えられた日本語の文を、論理式や依存図式など、個別言語に依存しない、なんらかの形の意味表現形式に写像することが、構文解析と意味解析の役割といえる⁶⁾。そのための前処理として形態素解析を位置づけ、従来自然言語処理の観点からは関心が払われてこなかった問題を指摘したい。

和語系の非サ変動詞には、{付く, 付ける}, {回る, 回す}, {落ちる, 落とす} のような、同一の語根から生じた派生対が数多く存在する。自然言語処理用の標準的な辞書として、福岡大学日本語処理用辞書(原辞書 KID-J82 より構成)を例にとると、このような対は 215 組にのぼり、490 個(三つ以上が 1 組となっているものもある)もの動詞を含む。

例として、{落ちる, 落とす} を形態音韻論に従って分析すれば、それぞれの語幹「oti」と「otos」は、語根「ot」に、それぞれ自動詞化接辞「i」と他動詞化接辞「os」が付加して派生したものである。文献 11) では、これらの派生接辞による派生の型を形態的に大きく 9 種類に分類している。これにもとづき、それぞれの種類から 1 組ずつ動詞を採り、例文を示したのが表 1 である。文頭の*は、非文を示す。「e₁」「e₂」の添え字「₁」「₂」は、機能が異なる二つの「e」を区別するためである。

{ar, e₂, re, i, ...} は自動詞化接辞, {e₁, s, as, os, ...} は他動詞化接辞である。以下では、表 1 に従って、「ar-φ 型の派生対」のような呼び方をする。

なお 4 類について、文献 11) では「r」と「s」の対

表 1 日本語の非サ変動詞の派生対
Table 1 Derivational pairs of Japanese verbs.

類 型	例 文	
	a)	b)
1 ar-φ	*彼女の手が覆まった	彼女が彼女の手を覆んだ
2 ar-e ₁	彼女が彼に英語を教わった	彼女が彼女に英語を教えた
3 φ-e ₁	蓋が開(あ)いた	彼が蓋を開(あ)けた
4 φ-s	ハンドルが回った	彼がハンドルを回した
5 e ₁ -as	予算が増えた	政府が予算を増やした
6 e ₁ -s	椅子が倒れた	彼が椅子を倒した
7 i-os	石が落ちた	彼が石を落とした
8 i-as	会議が延びた	議長が会議を延ばした
9 φ-as	石が動いた	彼が石を動かした

立をもつ型としているが、ここでは「r」は、屈折接辞の一部とみなし、接辞としてはφとしておく。

2 類～9 類に属する派生対について、さまざまな用例を相互に比較すると、

A) 1 類の派生対の一部を除き、動詞の各必須補語に与えられる意味役割や、各必須補語と格助詞との対応が、対をなす動詞の間で規則的に変化する(必須補語と意味役割に関する説明は第 4 章で行う)。

B) 各対で、a の意味は、b の意味に含まれる。という事実が得られる。換言すれば、派生対をなす 2 類～9 類の動詞の間に、構文的、意味的に明確な対応が存在するのである。この対応は、使役文と平常文の対応に酷似しており、自然言語処理の立場からは看過できない。これらの動詞を含む文は、使役文などと統一的に処理することが望ましい。そのためには、形態素解析で動詞語幹を語根と派生接辞に分離して構文解析に引き渡す必要がある。

言語学においても、これらの派生対を別動詞として扱わず、使役、受身などと統一して扱おうとする井上の研究⁹⁾がある。井上は変形文法の立場から、自動詞文と他動詞文の比較を行い、他動詞文を、埋め込み自動詞文の変形により説明している。井上が対をなす自動詞と他動詞を、異なる語彙として扱うことに反対している理由を概括すれば次のようになる：

- 自動詞文と他動詞文では、表層の主語と目的語の対応など、基本的な文法関係の対応がみられる。
- 他動詞化接辞「as」と使役接辞「ase」のように、構文上の働きが非常に似たものがあり、一方を語彙の問題、一方を構文の問題とすることは有効な一般化を妨げる。(注：使役接辞は、使役助動詞語幹に対応する生成文法の用語である)
- 主語の選択に関する構文上の必要から自動詞文または他動詞文が成立すると考えられる場合があるが、自動詞と他動詞を別個の語彙項目とすると、これらの関連が見失われる。

我々は変形文法の立場に立つわけではないが、これらの論点は、我々の提案に一つの根拠を与える。

以上から、派生対をなす動詞の形態素解析における形態素の設定に関して、まず次の要請を立てる：

I. 動詞の語根と派生接辞を、形態素として考慮すること

2.2 活用形の扱いについて

次に活用形の扱いについて考える。さまざまな要請

が考えられるが、なかでも重要と思われる二つをあげたい。

第1は、構文解析に用いられない無意味な分割の発生を避けることである。具体例で述べれば、「行かない」を「行/か/な/い」と細分しないということである。この場合の「か」は接続の仲立ちにすぎない。このような分割を行わないことは、効率の面からも自然な要請である。

第2は、活用形の解析のために必要な形態素の総数と、意味表記すべき対象の実数との差を、できるだけ小さくするということである。ここで、この要請を正確に述べるため、用語を二つ導入する。言語学用語の異形態の定義を流用し、「教わっ」と「教わら」のように、特定の形態素が出現環境により形態を変えた形態素どうしを互いに異形態と呼び、そのなかで標準形とみなす一つを基本形と呼ぶことにする。これらの用語を用いると、第2の要請は、活用形の解析のために必要な異形態の数と、基本形の数との差を、できるだけ小さくするということである。

第1と第2の要請は一見両立が困難に思えるが、共に重要である。これらをまとめて次の要請を立てる：

- II. 無意味な分割を避けつつ、活用形処理のために必要な異形態の数と、基本形の数との差をできるだけ小さくする

3. 三つの方法

以下で、形態素解析の対象は標準的に漢字仮名混じり表記されていると仮定する。考察の対象としたのは、第2章で述べた辞書にある非サ変動詞2,807個である。異形態の数などの定量的評価も、これらを用いて行った。その内訳は表2のようである。

以下本章では、学校文法にもとづく形態素の設定、形態音韻論にもとづく形態素の設定を概括し、これらと対比して新たな形態素の設定方法を示す。そして、

表2 考察対象とした動詞の内訳
Table 2 A classification of sample verbs.

型	個数
イ音便型I (カ行五段)	232 個
イ音便型II (ガ行五段)	51 個
促音便型 (バ行, マ行, ナ行五段)	250 個
撥音便型 (ラ行, タ行五段)	664 個
ワ行五段	197 個
サ行五段 (音便なし)	443 個
一段動詞	970 個
計	2,807 個

新方法だけが、第2章であげた要請I, IIを満たすことを示す。

3.1 学校文法にもとづく形態素の設定方法

まず、学校文法では、形態素として派生接辞を切り離して扱うことはなく、第2章の要請Iは満たされない。

次に要請IIについて考える。「教わる」を例にとると、学校文法に単純に従えば、「教わる」の活用形を含むさまざまな動詞句は、

教わらない → 教わら/な/い
 教わろう → 教わろ/う
 教わりながら → 教わり/ながら
 教わった → 教わっ/た
 教わって → 教わっ/て
 教わる → 教わる
 教われれば → 教われ₁/ば
 教われ → 教われ₂

のように分割される。しかし、{教わら, 教わろ, 教わっ, 教われ₁, 教わり}と異なり、{教わる, 教われ₂}には、テンスやモーダルの情報が、「る」と「れ」によって付加されている。したがって、形態素解析に一貫性をもたせるため、より正確には、

教わる → 教わ/る
 教われ₂ → 教わ/れ

とする必要がある。したがって、この立場で形態素解析を行う場合、活用形処理のために必要な異形態の数は、{教わら, 教わろ, 教わっ, 教われ₁, 教わり, 教わ}の6個である。これらを用いれば、無意味な分割は生じない。

一方、対象とする全動詞における、活用形処理のために必要な異形態の数と基本形の数との差は、

{イ音便型の動詞数+促音便型の動詞数+撥音便型の動詞数+五段ワ行の動詞数}×5+{五段サ行の動詞数}×4+{一段活用の動詞数}×3=11,652 個

となり、異形態数は基本形の数の数倍である。したがって、要請IIの後半を満たしているとはいえない。

3.2 形態音韻論にもとづく形態素の設定方法

日本語の動詞の活用を記述する試みは、学校文法以外にも幾つか存在する(文献12)、第4章)が、それらに最も大きな影響を与えたのは、米国の構造主義言語学者 Bernard Bloch の分析¹⁾であろう。彼は、音素単位で活用形の再分析を行った。以下、表3の説明までは、de Chene²⁾による、形態音韻論的分析を概括したものであり、Bloch 流の分析を形式化したもの

表3 動詞接尾辞の語彙項目エントリー (東京方言)
(文献2), p. 165, 表 5.5 より)
Table 3 The lexical entries for verbal affixes of
Tokyo dialect.

音韻表示	環 境	範 ちゅう
∅	V] vs_] v	連用形
i] vs_] v	連用形
u] vs_] v	終止形
eba] vs_] v	仮定形
o	V] vs_] v	命令形
e] vs_] v	命令形
<y>ou	<V>] vs_] v	意志形
na	V] vs_] AS	否定
ana] vs_] AS	否定
are] vs_] vs	受身
<s>ase	<V>] vs_] vs	使役

である (ただし、用語、例、ローマ字表記などは原文とは若干異なる)。

形態音韻論では、語彙項目となる最小基本単位は、語幹や、屈折接辞であり、語根や派生接辞は単独で語彙項目とはならない。たとえば「教わる」は、「osowar」と「u」に分割して登録する。動詞の屈折形は、語幹と屈折接辞の連鎖に、音韻ルールを適用して生成される。音韻ルールは、記号 V (verb), VS (verb stem), AS (adjective stem) を用いて表3のようにまとめられる (ただし、音便は無視している)。

大文字のVは語幹の末尾が母音であることを、「]」は言語学上の形態素の境界を示す。特に「] vs」と「] AS」は、語幹と屈折接辞の境界を、「] v」は屈折接辞の境界をそれぞれ示す。「_」は、子音「r」が挿入される位置を示す。「<」と「>」を含むルールは、「<」と「>」で囲まれた要素の有無で二つのルールに展開できる。たとえば「osowaraseru (教わらせる)」は、語幹「osowar」と、屈折接辞「ase」、「u」の接続に対して、表3の11行目の括弧を含まない方のルールと、3行目のルールが適用され、はじめの挿入位置だけに「r」が挿入されて導出される：

[[[osowar]vs_ase]vs_u]v →
osowar ∅ ase r u → osowaraseru

(以上概括終わり)。

以上から、形態音韻論に従う場合も、第2章の要請Iは満たされない。

次に要請IIについて考える。上記の「r」挿入ルールなどは本来生成方向のルールであり、音便形の処理に関するルールを付け加えれば、形態素生成には応用できる。しかし、形態素解析の際に、子音挿入や音便変形を生成ルールを用いて試行錯誤的に処理するのは

効率上好ましくないで、子音挿入や音便形は、一群の異形態を辞書に記載して処理する必要がある。

形態音韻論的な分析を実際の解析に応用する場合、漢字ローマ字混じりで語幹を記述することになる。 「教わる」を例にとると、語幹の側では「教war」と「教wat」(あるいは「教wa」)の二つの形態素を用意しなければならない。「教wat」は連用形「教わって」などを処理するために必要である。

はじめに入力文字列の平仮名部分をローマ字化した後、次のような分割を行うことになる：

教わらない →教waranai →教war/ana/i
教わらせる →教waraseru →教war/ase/ru
教わった →教watta →教wat/ta
教わりながら→教warinagara →教war/inagara
教わる →教waru →教war/u

(以下省略)

この場合も、無意味な分割は生じない。対象とする全動詞における、活用形処理のために必要な異形態の数と、基本形の数との差は、

{五段サ行,ワ行以外の五段動詞数} + {五段ワ行の動詞数} × 2 + {連用形の音便処理に必要な異形態の数から基本形の数を引きいた数} = 1,598 個

となり、学校文法に比べて激減している。したがって、要請IIは、ほぼ満たされている。ただ、平仮名部分をローマ字化し、形態素を音素単位で表記するため、形態素がさまざまな子音で終わることを許すので、辞書引き回数が増大するおそれがある。この点については、3.3節で述べる。

3.3 提案する形態素の設定方法

我々は、第2章の要請Iを満たすため、語根や派生接辞を、屈折接辞とともに形態素と考える。これらの形態素の基本形として、基本的には形態音韻論にもとづく言語学上の形態素の基本形を用いるが、我々の解析対象は漢字仮名混じり文であるため、以下では漢字ローマ字混じり表記を用いる。たとえば、「死ぬ」の語幹は、「死n」と表記する。また、「ん」は、ナ行との混同を避けるため「x」とローマ字表記する。

3.3.1 動詞句解析に必要な形態素の設定

i) 「死n」、「歩k」、「殺s」など、派生の対をもたない動詞の扱い

すべての屈折形語幹の間で共有される、先頭側の連続部分文字列を形態素とする。たとえば「死ぬ」については、{死nana, 死nase, …}のすべてにおいて共有される「死」を形態素とする。

例) 死nana } 殺sana }
 死nare } 殺sare }
 死xda } 殺sita } → 殺s
 ⋮ } ⋮ }
 死ne } 殺se }

ii) ar-φ 型の派生対の扱い

これらは、2.1 節で指摘した対応関係 A, B が成り立たない対 (「擱mar」と「擱m」) と、成り立つ対 (「助kar」と「助ke」) に分かれる。

a) 対応関係が成り立たない対

個別の動詞について、i)と同様にして形態素を定める。

例) 擱mana } 擱marana }
 擱mare } 擱marare }
 擱xda } 擱matta } → 擱ma
 ⋮ } ⋮ }
 擱me } 擱mare }

b) 対応関係が成り立つ対

第一段階)

対になる動詞の語幹を漢字ローマ字混じり表記し、先頭側の最大一致部分文字列を切り出す。

例) 助kar } → 助k
 助ke }

第二段階)

第一段階で切り出した文字列の末尾が子音なら、それを末尾から除いた文字列を形態素とする。

例) 助k → 助

iii) ar-φ 型以外派生対をなす動詞の扱い

ii-b)と同じ手続きを行う。

例) 回r } → 回 動k }
 回s } → 回 動kas }

i)~iii)の手続きで、派生対をなす動詞から得た形態素を以下では単に語根と呼ぶ。

iv) 接辞の扱い

a) 派生接辞

派生接辞を語幹に含む動詞の場合、iii)までの手続きで定めた語根を各動詞の語幹から取り除くと、派生接辞そのもの、もしくは派生接辞に特定の子音が付加された文字列が残る (たとえば、「教e」の「e」、「教war」の「war」)。これらには屈折接辞が後続するので、i)と同様に屈折形先頭の共通部分を切り出し、その結果を派生接辞に対応する形態素とする。

例) ena } warana }
 erare } warare }
 ete } watte } → e } wa
 ⋮ } ⋮ }
 ero } ware }

我々の立場では、これらはそれぞれ「e」、「ar」の異形態である。

b) 屈折接辞

以上の処理で、多くの場合、語幹の基本形の末尾子音が切り出されずに残った。これらは屈折接辞側へ付加し、屈折接辞基本形の異形態をつくる。たとえば、「死n」からは「死」を語根として切り出した結果、語幹「死n」の末尾「n」が残っている。これは、後続する屈折接辞に付加し、たとえば基本形「ana」から異形態「nana」をつくる。また、連用形の音便も屈折接辞の異形態として扱う。たとえば「死xda」の「xda」は、完了時制詞「ita」の異形態とみなす。

以上で定めた形態素を用いると、3.1, 3.2 節で用いた例は、次のように解析される：

教わらない → 教 waranai → 教/wa/rana/i
 教わらせる → 教 waraseru → 教/wa/rase/ru
 教わった → 教 watta → 教/wa/ta
 教わりながら → 教 warinagara → 教/wa/rinagara
 教わる → 教 waru → 教/wa/ru

(以下省略)

この場合も、無意味な分割は生じない。対象とする全動詞における、活用形処理のために必要な異形態の数と、基本形の数の差は、

{屈折接辞に由来する異形態数と、屈折接辞基本形数との差} + {派生接辞に由来する異形態と、派生接辞基本形数との差} = 125 個

である。したがって、第2章の要請IIは、よく満たされている。

ここで、我々の方法は、形態音韻論による方法と同様、ローマ字化にともない文字数が増加するため、形態素解析の手続きにおいて、辞書引き回数が増加が心配される。実際、形態音韻論の方法では、{b, d, g, k, m, n, r, s, t, w} で終わる形態素が現れるため (「d」は、助動詞「だ」を、「d/a」と分解するため生じる)、これらの子音の直後を辞書引き開始位置から排除できない。また、これらの子音で終わる部分文字列も辞書引きの対象から排除できない。

しかし、我々の形態素は「s」と「d」以外の子音では終わらないので、「s」、「d」以外の子音の直後位置は辞書引き開始位置から排除でき、「d」1文字、または、「s」以外の子音で終わる文字列も辞書引き対象から排除できる。したがって、平仮名部分をローマ字化しても、辞書引き回数にあまり影響はない。

3.3.2 接続コードについて

上で切り出した形態素の間の接続可否を表現するための接続コード体系について述べる。我々の手法では語幹の末尾子音を後続の形態素に付加するため、その子音を示す情報を接続コードに含ませる必要がある。そのため、各形態素に与える接続コードとして単なるシンボルを用いた場合、付加子音によってコードの細分が発生し、コード数が増大する。たとえば、使役を表す形態素の基本形「ase」に、「cau」というコードを与えたとすると、これに「n」や「k」を付加した異形態「nase」、「kase」には、「n-cau」、「k-cau」のようなコードを与えて、付加子音に関する情報を表現しなければならないからである。このままではコード間の接続テーブルの大きさが増大し、テーブルの作成、保守の面から好ましくない。

この問題を解決するため、各形態素に、第1成分を「受けコード」、第2成分を「係りコード」と呼ぶコードの順序対 (c₁, c₂) を与え、これを用いて形態素の間の接続を表現する、すなわち、形態素 A に (a₁, a₂)、形態素 B に (b₁, b₂) を与えたとき、A に B が接続可能であるか否かを、係りコード a₂ と受けコード b₁ の接続可否で表現する。これは、それぞれ 1, 0 で接続テーブルに記載する。

具体的に「歩 kaseru」に現れる形態素 {歩, kase, ru} のコードを示すと、たとえば次のようになる (コードの説明は表4参照) :

歩 (vs, v1k)

kase (k-infl, cau)

ru (r-infl, non-pft)

「歩 kaseru」は、v1k と k-infl, cau と r-infl の接

表4 コードの例
Table 4 Examples of codes.

コード	コードの説明
vs	動詞語幹を示す。
v1r	「r」を付加した屈折接辞のみが接続できることを示す。
k-infl	屈折接辞の基本形 {ase, are, ...} に、子音「k」を付加した異形態の集合を示す。
r-infl	屈折接辞の基本形 {ase, are, ...} に、子音「r」を付加した異形態の集合を示す。
cau	使役接辞を示す。
non-pft	非完了時制詞を示す。
v3w	「e _i 」と「ar」(の異形態)のみが接続できる語根を示す。v3w の「w」は、後続する派生接辞への付加子音が「w」であることを示す。ただし、「we」は「e」と同一視する。
w-af1	af1 は、「wa」の基本形が「ar」であることを、「w-」は付加された子音が「w」であることを示す。

表5 形態素の係りコードの一部

Table 5 A part of Kakari-codes of morphemes.

型	基本係りコード	例と係りコードの枝番
φ	v1	歩, v1k
ar-φ	v2	塞, v2g
ar-e _i	v3	教, v3w
φ-e _i	v4	開, v4k
e _i -φ	v5	消, v5
φ-s	v6	回, v6r
e _i -s	v7	隠, v7r
e _i -as	v8	増, v8y
i-os	v9	起, v9k
i-as	v10	延, v10b
φ-as	v11	動, v11k

表6 3方法の比較

Table 6 A comparison among the three methods.

	学校文法にもとづく方法	形態音韻論にもとづく方法	提案する方法
要請 I	×	×	○
要請 II	×	○	◎

続可能性により受理される。

語根と派生接辞異形態の間の接続は、語根の係りコードと、派生接辞の異形態の受けコードの接続可否によって表現する。表5に示したのは、各語根の係りコードである。たとえば、「教 waru」は、語根「教」と、派生接辞「ar」の異形態「wa」、非完了時制詞「u」の異形態「ru」に分解されるが、はじめの二つには、それぞれコード対 (vs, v3w), (w-af1, v1r) が与えられ、v3w と w-af1 の接続可能性により「教」に「wa」が接続可能なことが示される。

コードの対による接続表現は、受けコードと係りコードを比較的少数用意すれば、組合せにより多数のきめ細かい接続関係が表現でき、接続テーブルの大きさを抑えることができる。

3.4 三つの方法の比較

以上から、3.1~3.3節で述べた3種の方法について、第2章の要請 I, II を満たすかどうかをまとめると表6のようになる。「×」は、その要請が満たされなことを示す。要請 II における「○」と「◎」の違いは、異形態の数を反映させたものである。3.3節で示した方法だけが、二つの要請を満たしている。

4. 語根への意味表記の指針

本章では、第3章で示した形態素を、意味解析への入力として利用できるようにするための、語根へ意味表記する際の指針を述べる。意味表記の方法は、文献

7), 9) のようにさまざまなものが考えられるが, ここでは, 具体的な意味表記の形式に関しては議論しない。それは本論文の趣旨から外れるだけでなく, 以下の議論は具体的な意味表記の形式からは独立しているからである。

以下では, 動詞の必須補語の数と, その意味役割が問題となる。動詞の補語とは, 動詞を用いて描写される事象に登場する人, 物, 概念などであり, 必須補語は, 単に事象を詳細に叙述するための補語ではなく, それが欠けると事象の叙述が不完全になるような補語である(文献4)。以下の, 意味役割に関する分析は, 文献3), 4)を参考にした。

4.1 意味役割

意味役割は, 動詞の必須補語に付与され, その補語がその動詞に対してどのような意味的関係を結ぶかを示す。文献4)では, 中核的な意味役割として, 動作主(動作の主体), 経験者(感情, 知覚などの精神活動, 状態を経験する存在), 着点(人や物が到達する場所), 起点(人や物が離れてゆく場所), 相手(事象の主体に対する相手役, または属性叙述の対象が関係する相手), 対象(他の意味役割が表す意味的関係以外を広く表す)と, これらが複合した二重意味役割をあげている。ここでは, これらに文献3)で述べられている使役動作主(他の人や物が動作主または経験となる動作, 感情, 知覚などを生じさせる主体)を加える。意味役割は, 各補語に格助詞を結びつける手がかりとなる。意味役割を示すために以下で用いる記号は次のとおり:

{CA: 使役動作主, A: 動作主, E: 経験者,
S: 起点, SA: 起点動作主, G: 着点,
GA: 着点動作主, O: 対象}

4.2 必須補語がもつ意味役割と動詞対

以下では, 動詞の必須補語の個数と, それらの意味役割が, 派生対をなす動詞の間でどのように変化するかをみる。まず, 第2章であげた動詞について, 必須補語に付加される意味役割を比較すると, 表7のようになる:

a)とb)の対応を整理して述べるため, 動詞を, 授与動詞と呼ばれるものとそうでないものに分ける。

4.2.1 授与動詞

授与動詞とは, {教える, 教わる, 授ける, 授かる, やる, 貰う, 借りる, 貸す, …} など, S(SA), G(GA), Oを意味役割とする3個の必須補語をもち, Sに対応するものからGに対応するものへ, Oに対応するもの

表7 派生接辞と意味役割

Table 7 A relation between derivational affixes and semantic roles.

類 型	例			
	a)		b)	
1 ar-φ	助kar	<O>	助ke	<A, O>
2 ar-e _i	教war	<S, GA, O>	教e	<SA, G, O>
3 φ-e _i	開k	<O>	開ke	<A, O>
4 φ-s	回r	<O>	回s	<A, O>
5 e _i -as	増e	<O>	増yas	<A, O>
6 e _i -s	倒re	<O>	倒s	<A, O>
7 i-os	落ti	<O>	落tos	<A, O>
8 i-as	延bi	<O>	延bas	<A, O>
9 φ-as	動k	<O>	動kas	<A, O>

が, 物理的ないしは抽象的な意味で移動するという共通した内容をもつ。特徴は, SまたはGをもつ必須補語に焦点が当てられ, その必須補語が, 二重意味役割SAやGAをもつことである。授与動詞が派生対をなす場合, 次の特徴がある:

- 派生対の間で必須補語の数が一致する。
- 派生対の間で, 他動詞化接辞をもつ方はSが, 自動詞化接辞をもつ方はGが, それぞれ動作主との二重意味役割となる。

例) [授ke <SA, G, O>
授kar <S, GA, O>

4.2.2 非授与動詞

非授与動詞においては, 表1に示した2類~9類のすべての派生対で, b)の動詞の必須補語の数は, a)の動詞の必須補語の数より1多く, あらたに加わる意味役割は, 動作主, または, 使役動作主である。

4.3 接辞付加と意味役割

4.2節における観察にもとづき, 我々が第3章で設定した形態素を, 構文解析, 意味解析で用いるための意味表記上の指針を述べる。以下では, 議論をみやすくするため, 動詞語幹などは, 漢字ローマ字混じり表記した言語学上の形態素を用いて表す。

4.3.1 派生対をもたない動詞

「死」, 「歩」, 「殺s」などはそれぞれ, 「死n」, 「歩k」, 「殺s」に対する意味表記を考えればよい。「擱mar」と「擱m」など, 第2章の関係A, Bが成立しない一部の動詞もそれぞれ別動詞として扱い, 「擱ma」と「擱」に対して個別に意味表記すればよい。

4.3.2 派生対をもつ動詞

i) 授与動詞

これらの対では, 三つの必須補語に与えられる意味役割は, 基本的には {S, G, O} である。「SからGへ

のOの移動」の記述を、派生対の間で共通に用意し、語根の意味表記とする。授与動詞の語根はマークし、構文解析、意味解析の段階で、他動詞化接辞が付加された場合SをSAに、自動詞化接辞が付加された場合GをGAにする。

例) 「教」〈S, G, O〉

+e → 教e 〈SA, G, O〉

+war → 教war 〈S, GA, O〉

ii) 非授与動詞

● 派生関係にある動詞の、片方の派生接辞がφの場合

● 他動詞化接辞がφの場合

派生接辞をもたない方の動詞の意味表記を、語根の意味表記とする。自動詞化接辞が付加されると、動作主の意味役割を付与されていた補語は必須補語ではなくなる。

例) 「折」〈A, O〉

+φ → 折r 〈A, O〉

+re → 折re 〈O〉

● 自動詞化接辞がφの場合

派生接辞をもたない方の動詞の意味表記を、語根の意味表記とする。他動詞化接辞が付加される場合、その事態を生じさせる意志的主体として、動作主の意味役割をもつ必須補語を加える。すでに意味役割として動作主があるときは、使役動作主の意味役割をもつ必須補語を加える。

例) 「回」〈O〉

+s → 回s 〈A, O〉

+φ → 回r 〈O〉

「着」〈A, O〉

+φ → 着r 〈A, O〉

+se → 着se 〈CA, A, O〉

● 派生関係にある動詞の、派生接辞が両方ともφでない場合

派生対の間で、必須補語の数が少ない方の動詞の意味表記を、語根の意味表記とする。他動詞化接辞が付加される場合、あらたに動作主の意味役割をもつ必須補語を加える。

例) 「増」〈O〉

+yas → 増yas 〈A, O〉

+e → 増e 〈O〉

以上では簡単にしか述べなかったが、適切な意味表記のもとで、我々の形態素を用いて動詞の必須補語と意味役割の組を計算することができる。意味役割を用いて、各補語に対して表層の格助詞を指定する方法は

省略する(文献4)、第2部第2章を参照)。

5. おわりに

本論文で我々は、日本語漢字仮名混じり文において、非サ変動詞句を解析するための形態素の設定方法を考察した。このとき、二つの要請

I. 動詞の語根と派生接辞を、形態素として考慮する

II. 無意味な分割を避けつつ、活用形処理のために必要な異形態の数と、基本形の数の差をできるだけ小さくする

を立てた。Iは、同一語根から派生した自動詞/他動詞の対が、多数存在する日本語の特性を考慮し、意味解析との連携を高めるための要請であり、IIは、活用形処理の合理化を徹底するための要請である。

我々は、要請Iを満たすため、語根と派生接辞に対応する形態素を設定する手続きを示した。動詞の活用は、語幹末尾子音を屈折接辞側へ付加し、屈折接辞の異形態に吸収することにより捉えることにより、要請IIも同時に満たすことができた。

さらに、新しく設定した形態素の間の接続関係を表現するための接続コードの体系を示し、これらの形態素を用いて、もとの動詞の必須補語と意味役割の組が計算できることも示した。

新しく設定した形態素により、従来異なる語彙として個別に取り扱われてきた派生対を統一的に扱うことが可能となり、有意味最小単位という本来の定義に照らし、非サ変動詞句の解析にとって、より合理的な形態素が設定できたと考える。

本提案の形態素で扱えないのは、派生対をなす動詞の一方に、他方に対応しない語彙依存的な意味が存在する例外的な場合である。そのような場合は、必要に応じて個別に考慮しなければならない。また、{狭まる, 狭める}のように、語根が形容詞と共有できる場合の、動詞と形容詞を統合した扱いについては言及しなかった。別の機会に報告したい。

謝辞 日本語処理用辞書の資料としての使用を認めてくださり、かつ貴重なコメントを頂いた、福岡大学・首藤公昭教授に感謝の意を表します。

参考文献

- 1) Bloch, B.: Studies in Colloquial Japanese, Part I, Inflection, *Journals of the American Oriental Society*, Vol. 66, pp. 97-109 (1946).
- 2) de Chene, B.: 形態音韻論における心理的実在

- 性(ことばからみた心, 第4章), 東京大学出版会, 東京(1987).
- 3) 井上和子: 変形文法と日本語(下), 大修館書店, 東京(1976).
 - 4) 益岡隆志: 命題の文法—日本語文法序説一, くろしお出版, 東京(1987).
 - 5) 長尾 真ほか: 国語辞書の記憶と日本語文の自動分割, 情報処理, Vol. 19, No. 6, pp. 514-521 (1978).
 - 6) Nitta, Y. et al.: A Proper Treatment of Syntax and Semantics in Machine Translation, *Proc. of COLING '84*, pp. 159-166 (1984).
 - 7) Okada, N.: Towards a Unified Understanding of Natural Language and Picture Patterns, *Proc. of International Symposium on Language and Artificial Intelligence*, Kyoto, pp. 10-1-10-23 (1986).
 - 8) 大塚高信ほか(監修): 新英語学辞典, 研究社, 東京(1982).
 - 9) Schank, R. C.: *Conceptual Information Processing*, North-Holland, Amsterdam and American Elsevier, New York (1975).
 - 10) 首藤公昭ほか: 日本語文の機械処理のための文節構造モデル, 電子通信学会論文誌, Vol. J 62-D, No. 12, pp. 872-879 (1979).
 - 11) 寺村秀夫: 日本語のシンタクスと意味I, くろしお出版, 東京(1982).
 - 12) 寺村秀夫: 日本語のシンタクスと意味II, くろしお出版, 東京(1984).
 - 13) 吉村賢治ほか: 文節数最小法を用いたべた書き日本語文の形態素解析, 情報処理学会論文誌, Vol. 24, No. 1, pp. 40-46 (1983).

付録 用語の定義

文献8), 11)を参考にして, 本論文中で用いた言語学用語を説明する。

- 1) 形態素: 特定の意味を有す最小の単位であり, 発話において独立して使用可能な自由形態素と, そうでない拘束形態素よりなる。
- 2) 異形態: 形態素が, 出現環境によって音韻変化を起こしたもの。
- 3) 語幹: 拘束形態素の一つ。原則として品詞をもたず派生接辞を受けて語幹を生じる。たとえば「ugoku」の「ugo-」。

4) 語幹: 品詞をもち, 屈折接辞を付け加えられる。たとえば, 「ugoku」の「ugok-」。

5) 接辞: 拘束形態素の一つで, 語幹または語根に接続する形でしか現れない。接頭辞と接尾辞からなり, 接尾辞はさらに屈折接辞と派生接辞に分かれる。

屈折接辞は, ある一定の品詞をもつ語幹に例外なく付加でき, 語幹に付け加える意味は規則的である。たとえば, 「ugoita」の「-ita」は屈折接辞である。学校文法で, いわゆる活用形に接続する助詞や助動詞に対応する。

派生接辞は, 特定の語根にだけ付加でき, 語幹を派生する。自動詞化/他動詞化接辞を含む。「ugokasita」の「-as-」は他動詞化接辞。

(平成元年 5月24日受付)

(平成2年 11月13日採録)



久光 徹 (正会員)

昭和35年生。昭和61年東京大学大学院数学科修士課程卒業。同年(株)日立製作所入社。現在同社基礎研究所勤務。自然言語処理の研究に従事。ACL 会員。



新田 義彦 (正会員)

昭和21年生。昭和44年東京大学理学部数学科卒業。同年より(株)日立製作所中央研究所に勤務。昭和49年より新設の同システム開発研究所に勤務。昭和60年より新設の同基礎研究所に勤務。現在, 同所主任研究員。この間昭和51年~52年スタンフォード大学工学部 OR 学科(M.S.)。形式言語, 情報検索, 機械翻訳, 人工知能, 自然言語理解の研究に従事。最近はロジックと自然言語のギャップに関心が集中している。「OAのソフトウェア」(第5章自然言語処理ソフトウェアを執筆, オーム社)。日本ソフトウェア科学会, 電子情報通信学会, ACM, ACL 各会員。