

## 派生文法に基づく日本語動詞接尾辞の形態素解析

小川泰弘<sup>†</sup> 稲垣康善<sup>†</sup> ムフタル・マフスット<sup>‡</sup>

<sup>†</sup>名古屋大学工学部情報工学科 <sup>‡</sup>三重大学工学部情報工学科

日本語の膠着言語の性質と音韻論的性質に注目した清瀬の派生文法では、活用という考え方を  
用いなくて動詞接尾辞を考えることにより、動詞語形の形成を単純かつ体系的に取り扱うことを  
可能としている。本稿では、派生文法に基づく日本語形態素解析法を提案し、不規則動詞を含む  
各種の動詞語形の変化を簡単に解析できることを示す。また従来異形態の登録で対処されてきた  
音便形に対して、後方からの検索と子音の補完により余分な辞書登録をしなくて解析する手法を  
提案し、その有効性を示す。

### Morphological Analysis of Japanese Verbal Suffixes based on Derivational Grammar

OGAWA, Yasuhiro<sup>†</sup> INAGAKI, Yasuyoshi<sup>†</sup> Muhtar, Muhsut<sup>‡</sup>

<sup>†</sup>Dept. of Information Eng., Nagoya Univ. <sup>‡</sup>Dept. of Information Eng., Mie Univ.

The morphological analysis of Japanese language has been studied on the assumption that  
Japanese verbs and adjectives conjugate. But, the derivational grammar proposed by Kiyose  
who considers the agglutinative character of Japanese language is able to describe verbal  
clauses systematically without conjugation. In this paper, we propose the morphological  
analysis of Japanese language based on derivational grammar. Especially we propose back-  
ward search and supplementing a consonant enable to analyze the sandhi form effectively.

#### 1 はじめに

従来、日本語形態素解析は、用言の活用を前提に  
して行われてきた。しかし日本語の膠着言語とし  
ての性質に着目した清瀬の派生文法では、活用と  
いう考え方をいなくて、動詞接尾辞を考えること  
によって、動詞語形の形成を体系的に取り扱う  
ことを可能としている。本稿では、第2章で派生  
文法の概要を述べ、第3章で派生文法に基づく  
形態素解析システムを提案する。またこの中で  
従来音韻論的手法による形態素解析で問題にな  
っていた音便形に対して新たな処理法を提案す  
る。第4章においてこれまでの形態素解析法と  
の比較を行う。

#### 2 派生文法

学校文法では、動詞や助動詞などの用言が活用  
するとされ、その活用形は未然形、連用形、終  
止形、連体形、仮定形、命令形のように分類  
された。そのため日本語の形態素解析におい  
ては用言の活用をどのように処理するかが問  
題とされてきた。

日本語は言語の分類において膠着語であるとい  
われている。膠着語とは実質的観念を表す語幹  
に文法的機能を表す接辞が結合することによ  
って単語を形成するという性質をもつ言語の  
総称である。

アルタイ言語学者の清瀬は、膠着言語である  
日本語に活用が存在することを矛盾として捉え、  
活用を前提としない派生文法 [3][4] を提案  
した。派生文法の概要をその特徴が表れる  
動詞語形の形成の部分を中心に以下に紹介  
する。

#### 2.1 連結子音と連結母音

動詞の不変部分語幹と呼ぶ。いわゆる一段活  
用動詞「見ル」「食ベル」などの場合は、不  
変部分「見」「食ベ」が語幹であり、その際  
語幹は母音 i か e のいずれかで終る。いわ  
ゆる五段活用動詞の場合には、例えば「書カ  
」「書キ」「書ク」「書ケ」「書コ」のよ  
うに、末尾が変化するといわれる。これは  
音韻論的に考えれば、“kak-a”、“kak-  
i”、“kak-u”、“kak-e”、“kak-o”  
であり、“kak”を不変部分、すなわち語幹  
として取り出すことができる。いわゆるワ  
行五段活用をするといわれる「買ウ」  
“ka-u”のような動詞の場合には“kaw”  
を語幹とする。

いわゆる五段活用動詞の語幹はすべて子音  
で終るので子音幹動詞と呼び、いわゆる一  
段活用動詞の語幹はすべて母音で終るので  
母音幹動詞と呼ぶ。

派生文法においては、動詞の変形は動詞の語幹に

表 1: 動詞と接尾辞の接続例

未然形	kak-ana-i kak-are-ru kak-ase-ru kak-ou	tabe-na-i tabe-rare-ru tabe-sase-ru tabe-you	-(a)na-i -(r)are-(r)u -(s)ase-(r)u -(y)ou
連用形	kak-imas-u ka $\phi$ -ita	tabe-mas-u tabe-ta	-(i)mas-(r)u -(i)ta
終止形	kak-u	tabe-ru	-(r)u
連体形	kak-u	tabe-ru	-(r)u
仮定形	kak-eba	tabe-reba	-(r)eba
命令形	kak-e kak-una	tabe-ro tabe-runu	-e / -ro,-yo -(r)una

接尾辞が接続したものとして考える。そのため伝統的な文法でいう活用形の語尾や助詞、助動詞をすべて接尾辞として扱う。これらをいままでの活用の枠組に対応させると表1ようになる。なお表1における記号 $\phi$ は音便により対応する子音が消失したことを表している。

いわゆる終止形「書ク」「kak-u」「食ベル」「tabe-ru」の場合、接尾辞はそれぞれ“-u”、“-ru”である。派生文法ではこれを“(r)u”としてまとめて表記する。この括弧内の子音の有無は動詞語幹の末尾に依存して決まる。言い換えると、“kak-”に“-ru”が後接した場合、子音が連続するため接尾辞の先頭のrが欠落するのである。このような子音を連結子音と呼ぶ。

いわゆる否定の助動詞「ナイ」が接続する場合を考えてみる。これはいわゆる未然形に接続するといわれる接尾辞である。派生文法では「書カナイ」「食ベナイ」は“kaka-nai”、“tabe-nai”ではなく、表1に示すように“kak-ana-i”、“tabe-na-i”と解析する。ここで否定を表す接尾辞は“(a)na-”の形で表される。括弧内のaは母音が連続した場合に欠落し、このような母音を連結母音と呼ぶ。派生文法においては、動詞接続は以下の二つの規則を基本として行われる。

#### 接続規則 1:

子音幹動詞に連結子音を持つ接尾辞が接続する場合、連結子音を削除する

#### 接続規則 2:

母音幹動詞に連結母音を持つ接尾辞が接続する場合、連結母音を削除する

## 2.2 派生接尾辞

前節で否定の接尾辞を“(a)nai”ではなく“(a)na-i”と表記した。これは「書カナカッタ」「kak-ana-katta」のように更に他の接尾辞が接続することが可

表 2: 音便形

語幹末子音	音便形	語例	備考
-k -(i)ta	$\phi$ -ita	聞いた	イ音便
-g -(i)ta	$\phi$ -ida	泳いだ	イ音便+連濁
-r -(i)ta		切った	促音便
-t -(i)ta	-t - $\phi$ ta	立った	
-w -(i)ta		買った	
-b -(i)ta		飛んだ	撥音便+連濁
-n -(i)ta	-n' - $\phi$ da	死んだ	
-m -(i)ta		読んだ	
ik -(i)ta	it - $\phi$ ta	行った	例外
-s -(i)ta	-s -ita	貸した	音便変化なし

能だからである。これは動詞語幹に“(a)na-”が接続することにより新な語幹が派生したのである。こうした語幹を二次語幹と呼び、二次語幹を派生する接尾辞を派生接尾辞と呼ぶ。派生接尾辞には他にも“(r)are-”“(s)ase-”“(i)mas-”“(i)ta-”などがあり、それぞれ受身・可能・尊敬、使役、丁寧、希望の意義を表している。二次語幹に対する接尾辞の接続に関しても接続規則1、2は適用される。例えば表1では“kak-imas-fu”において、連結子音rが削除されている。派生接尾辞に対して、“(r)u”のように新たな語幹を派生しない接尾辞を統語接尾辞と呼ぶ。動詞に複数の接尾辞が接続する場合は、統語接尾辞が最後に接続する。

## 2.3 音便形

完了の統語接尾辞“(i)ta”が子音幹動詞に接続する場合は、接続規則1、2の例外となる。例えば「聞ク」「kik-u」の語幹“kik-”に接尾辞“(i)ta”が接続する場合、接続規則1に従えば「聞キタ」「kik-ita」となるものが、末尾子音kが欠落して「聞イタ」「ki-ita」となる。これは音便<sup>1</sup>と呼ばれる特別な語形変化であり、末尾子音がs以外の子音幹動詞において起こる。その変化は末尾子音に依存し、表2に示した変化を起こす。表中の $\phi$ は零記号で、接続によって $\phi$ に対応する音韻が消滅したことを示す。なおこの場合でも活用を考える必要はない。

動詞「行ク」「ik-u」の音便形は、この規則のさらに例外である。本来ならば末尾子音kが欠落して「行イタ」「i $\phi$ -ita」となるものが、末尾子音kがtに変化し「行ツタ」「it- $\phi$ ta」となる。

音便形をとる接尾辞には“(i)ta”、“(i)te”及び、それらから派生した“(i)tara”、“(i)temo”、“(i)teha”などがある。なお同じ形をしていても希望の派生接尾辞“(i)ta-”は音便形をとらない。

<sup>1</sup>派生文法では内的連声の一種とされる

## 2.4 形状動詞

伝統的文法でいう形容詞は、派生文法においては動詞の一種として扱われ形状動詞と呼ばれる<sup>2</sup>。派生文法では形状動詞においても活用を考える必要はない。「寒イ」“samui”は語幹“samu-”にいわゆる終止形の接尾辞“-i”が接続したものであり、この“-i”は形状動詞の語幹に接続する統語接尾辞である。このように形状動詞に接続する接尾辞を形状動詞接尾辞と呼ぶ。これらの接尾辞には連結子音や連結母音は存在しないので、形状動詞と接尾辞の接続は次の接続規則で表される。

### 接続規則 3:

形状動詞には形状動詞接尾辞が接続する

派生接尾辞にも形状動詞の語幹を派生するものがある。否定の“-(a)na-”、希望の“-(i)ta-”がそれぞれであり、「聞カナイ」“kik-ana-i”や「書キタッタ」“kak-ita-katta”のように、派生接尾辞の後に形状動詞接尾辞が接続する。すなわち“-(a)na-”や“-(i)ta-”は動詞の語幹に接続し新たに形状動詞の語幹を派生する派生接尾辞なのである。また“-gar-”は「寒ガル」“samu-gar-u”のように形状動詞の語幹に接続し新たに動詞の語幹を派生する派生接尾辞である。

## 2.5 不規則動詞

動詞の語幹に接尾辞が接続する様相は極めて規則的であるが、その例外となるのが不規則動詞である。伝統的文法でカ行変格活用をするといわれる「来ル」とサ行変格活用をするといわれる「スル」である。派生文法ではこの兩者についても活用は考えなくてよい。「来ル」は“ko-”、“k-”、“ku-”の三つ、「スル」は“se-”、“s-”、“si-”、“su-”の四つの語幹をもち、以下の規則で接尾辞と接続する。

### 接続規則 4.1:

“k-”、“s-”には、“-(i)ta”などの連結母音 i を持つ接尾辞や、“-(u)mai”などの連結母音 u を持つ接尾辞が接続する。

### 接続規則 4.2:

“ku-”、“su-”には“-(r)u”、“-(r)eba”など連結子音 r を持つ接尾辞が接続する。

### 接続規則 4.3:

“si-”には否定の派生接尾辞“-(a)na-”と命令形の派生接尾辞“-ro”が接続する

### 接続規則 4.4:

上記以外の接尾辞は“ko-”、“se-”に接続する。

<sup>2</sup>これに対していわゆる動詞は動作動詞と呼ばれる

この他にも日本語の動詞にはいくつか不規則な変化をする動詞がある。

話者が動作主に対する敬意を示す尊敬語の動詞「仰ル」“ossyar-u”に丁寧の接尾辞“-(i)mas-”が接続するとき、語幹末尾の r が欠落し「仰イマス」“ossyaφ-imas-u”という形を取る場合がある。このような動詞は派生文法では変則動詞と呼ばれる。

「アル」“ar-u”は語幹末尾が r である子音幹動詞であるが、否定の接尾辞“-(a)na-”が接続することがない、という点で特殊である。また前述の「行く」も音便変化においては不規則な変化をする動詞である。

## 3 形態素解析システム

前章で紹介した派生文法に基づく形態素解析システムを作成した。これまでに派生文法を利用した形態素解析の研究としては [8] がある。これは本来音韻論的解析である派生文法を、日本語の通常表記である漢字仮名混じり文を扱えるように変更し、JUMAN[10] 上に適用したものである。

これに対して本システムでは派生文法がもつ簡明さを活かすために、入力文を音韻単位で解析した。音韻単位で解析するためにはローマ字表記を用いる必要がある。そこで本システムでは通常の漢字仮名混じりの入力文が与えられたとき、平仮名の部分を自動的に日本式ローマ字表記に変換し解析する。

派生文法を形態素解析に適用する際の問題点と、それに対する解決法を以下に示す。

### 3.1 連結子音、連結母音の欠落

連結子音・連結母音の欠落は規則的ではあるが、そのまま適用すると欠落音の復元が必要となりシステムが複雑になる。そこでこの規則をそのまま適用するよりも、[5] と同様に欠落した接尾辞を同時に辞書に登録する手法をとった。これにより接続規則は以下のように変更される。

#### 接続規則 1':

子音幹動詞には母音で始まる接尾辞が接続する。

#### 接続規則 2':

母音幹動詞には子音で始まる接尾辞が接続する。

統語接尾辞については、母音で始まるものを子音幹接尾辞統語接尾辞(略して子統語接尾辞)、子音で始まるものを母音幹接尾辞統語接尾辞(略して母統語接尾辞)として分類する。例えば“-(r)u”に対しては、“-ru”を母統語接尾辞、“-u”を子統語接尾辞として辞書に登録する。

派生接尾辞には様々な種類がある。これらは接続する語幹と新たに派生する語幹から次のように分類する。例えば使役を表す“-sase-”は母音幹動詞に接

表 3: 接続行列

	母音幹動詞	子音幹動詞	母統語接尾辞	子統語接尾辞	母母派生接尾辞	母子派生接尾辞	子母派生接尾辞	子子派生接尾辞
母音幹動詞	○	○	○		○	○		
子音幹動詞				○			○	○
母統語接尾辞	○	○						
子統語接尾辞	○	○						
母母派生接尾辞	○	○	○		○	○		
母子派生接尾辞				○			○	○
子母派生接尾辞	○	○	○		○	○		
子子派生接尾辞				○			○	○

続いて新たな母音幹を派生するので母音幹接続母音幹派生接尾辞(略して母母派生接尾辞)と分類する。同様に“-imas-”は子子派生接尾辞、“-ana-”は子母派生接尾辞である。

品詞間の接続規則は、表3に示されるような接続行列で表される。これは接続行列のうち動詞と動詞接尾辞に関するものを抜き出したものである。例えば第2行第4列の記号‘○’は、子音幹動詞の直後に子統語接尾辞が接続可能であることを示している。

### 3.2 音便処理

本システムでは動詞と接尾辞の接続は活用形を調べることなく、接続行列を用いるだけで処理できる。しかし動詞の音便形は動詞の語幹および接尾辞が変化するため、そのままでは処理できないという問題が生じる。

従来の手法 [5][7] では、子音幹動詞の異形態を辞書に登録することによりこの問題を解決していた。例えば、「書ク」「書k-」に対しては「書-」を、「買ウ」「買w-」に対しては「買t-」を異形態として登録していた。この手法は単純ではあるが、語幹末尾がs以外のすべての子音幹動詞に対して異形態の登録が必要となり、その数は膨大なものになる。[7]では子音幹動詞2805個に対して約1600個の異形態の登録が必要であったし、EDR日本語単語辞書においては動詞26091個(うち子音幹動詞14353個)の中で異形態の登録が必要な動詞は9247個にもものぼる。また新たな子音幹動詞を辞書に追加する場合には、異形態も同時に登録する必要がある。

本システムでは、多数の異形態の登録が必要という欠点を克服するため、後方からの探索および子音の補完という手法で音便形を解析した。

異形態の登録が必要となるのは、音便変化をした動詞語幹が辞書に登録された形と異なり、単純な辞書引きでは検索ができなくなるからである。しかし

図 1: 音便処理

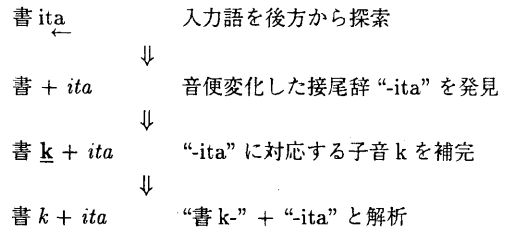


表 4: 動詞幹への補完

接尾辞	補う子音	用例
-ita	k	聞 ita → 聞 k-ita
-ida	g	泳 ida → 泳 g-ida
-tta	r	切 tta → 切 r-tta
	t	立 tta → 立 t-tta
-n'da	w	買 tta → 買 w-tta
	b	飛 n'da → 飛 b-n'da
-n'da	n	死 n'da → 死 n-n'da
	m	読 n'da → 読 m-n'da
-ita	φ	貸 sita → 貸 s-ita

文を後方から探索した場合、子音幹動詞の異形態より先に統語接尾辞の異形態が見つかる。例えば“(i)ta”の音便形による異形態に対しては表2の音便形に示してある“-ita”、“-ida”、“-tta”、“-n'da”の4種類の異形態のいずれかが見つかる。そこで、この4種類の接尾辞を辞書に登録しておき、これらを見つけたときに例外処理を行う。具体的には失われたと思われる子音を表4に従って補い、後方からの探索を続けるのである。この処理は図1のようになる。“-tta”、“-n'da”に対しては補う子音の候補が複数あるが、その場合にはそれぞれを補って適当なものを選択する必要がある。現在のシステムでは例えば“-ida”をg音便接尾辞として辞書に登録しておき、g音便接尾辞を見つけた場合にgを補うという例外処理を発生させるようにしている。

この手法においては完了の接尾辞“(i)ta”に対して“-ita”、“-ida”、“-tta”、“-n'da”という異形態を登録しなくてはならない。しかしこうした登録が必要なのは完了の“(i)ta”、“(i)te”及びそれらから派生した“(i)tara”、“(i)temo”など限られており、子音幹動詞の異形態を登録する従来の手法に比べてはるかに少数ですむ。現在のシステムでは音便処理のために48個の異形態が登録されているにすぎない。またこれは今後辞書に新たな動詞が追加されても増えることはない。

なお「行ク」「行k-u”は音便形が特殊であるが、これは異形態“行t-”を辞書に登録することで対処した。この種の動詞はEDR日本語辞書においては

14個しかなく、異形態の登録は少数ですむと判断したからである。

### 3.3 子音 w の補完処理と変則動詞

動詞の語形変化を音韻論的に解析する場合において、語幹末尾が w である子音幹動詞については独特の問題を生じる。「買う」は派生文法で考えれば語幹は“kaw-”であり、「買ワナイ」“kawanai”は“kaw-ana-i”と解析される。しかし「買う」を解析しようとした場合、そのローマ字表記は“kau”であり、動詞の語幹末尾の子音 w が表記されない。このため“kau”を解析しようとしても、辞書に登録された“kaw-”と形が異なるため解析できないという問題が生じる。

これに対する単純な解決方法は“kaw-”に対して異形態“ka-”を辞書に登録することであるが、語幹末尾が w である動詞と同数の異形態の登録が必要となる。この点が音便形の解析と類似していることに着目し、音便処理の応用で解決した。音便処理と同様に入力文を後方から解析し、子音幹接尾辞を発見した場合に子音 w を補完すればよい。しかし常に w を補完すると、子音幹動詞に子音幹接尾辞が接続した場合を解析できなくなるので、直前に母音があった場合のみ w を補完をする。例えば先程の“ka-u”の例では、まず子音幹接尾辞“-u”を発見し、更にその直前が母音 a であるから、子音 w を補完して“kaw-u”と解析できる。また現在のシステムでは漢字ローマ字混じり文を扱っているため、直前の語が漢字である場合(例えば直前の語が“買”である場合)も w を補完する必要がある。

変則動詞「仰ル」“仰 r-u”は、i で始まる接尾辞“-imas-”が接続する場合に、末尾の r が欠落し「仰イマス」“仰 φ-imas-u”という形になるという点で不規則な動詞である。このため語幹末尾で補完する子音として w の他に r も考慮しなくてはならない。しかしながら、「仰ル」などの変則動詞は EDR 日本語単語辞書において 18 個と少数である。そこで子音補完の例外処理を行うよりも、異形態を辞書に登録することで対処した。

### 3.4 不規則動詞

不規則動詞は 2.5 節で述べた規則で接続する。これらは動詞の種類が限られているため、接尾辞による例外処理で扱うよりも、動詞の異形態を辞書に登録する手法を採った。派生文法では「来ル」の語幹は母音幹動詞“ko-”、“ku-”、子音幹動詞“k-”の 3 種類である。しかし“k-”に関しては音便処理で補完する k と混同される場合があるので、“k-”ではなく“ki-”を語幹として登録した。このため本シ

ステムでは“kita”は“k-ita”ではなく、“ki-ta”と解析される。また「スル」については補完による問題は生じないため、“se-”、“su-”、“si-”を母音幹動詞、“s-”を子音幹動詞として登録している。しかし“si-”と“s-”が同時に登録されるため、例えば“sita”を“s-ita”と“si-ta”の二通りに解析するという問題が生じる。そのため“s-”を登録する必要があるかどうかは現在検討中である。

日本語動詞にはこの他にも変則動詞や「行く」が不規則は変化をするが、これらについては異形態の登録で対処することを既に示した。なお「アル」も不規則な動詞であるが、これは特に区別せず末尾が r の子音幹動詞として登録した。「アル」は否定の接尾辞“-ana-”が接続することがないという点で不規則であるが、これは“ar-ana-i”という形を過剰受理するだけであり、このような形は入力文に表れないと考えたからである。

## 4 従来の手法との比較

本システムによる形態素解析を従来の伝統的文法に基づく手法と比較し、本システムの長所を述べる。

「書かせられたがった」という語を例にとってみる。従来の伝統的文法論では、

書か	動詞「書く」の「未然形」
せ	未然形接続の助動詞「セル」の「未然形」
られ	未然形接続の助動詞「ラレル」の「連用形」
たがっ	連用形接続の助動詞「タガル」の「連用形」
た	連用形接続の助動詞「タ」の「終止形」

と解析されてきた。一方本システムでは、

書 k-	子音幹動詞「書 k-」
-ase-	子母派生接尾辞 “-(s)ase-”
-rare-	母母派生接尾辞 “-(r)are-”
-ta-	母形派生接尾辞 “-(i)ta-”
-gar-	形子派生接尾辞 “-gar-”
-tta	r 音便接尾辞 “-(i)ta”

と解析される。

伝統的な文法に基づく場合、動詞を辞書引きしたあとその活用形を調べる。その次に例えば「セ」は五段活用動詞の未然形に接続可能といった接続可能性を調べる。しかし派生文法に基づく本システムではこうした活用形を調べる必要はなく、ただ品詞間の接続を調べるだけで解析できる。

#### 長所 1: 動詞語形の変形の処理が単純である

活用形を調べる手間を省くために、活用形の語尾を分離して品詞として用いる方法がある。この手法は [1] で用いられているが、「カ」はカ行五段活用

動詞の語幹に接続して未然形となる、といった非常に細かな品詞の分類が必要になり、[1]では437種類に分類されていた。

伝統的文法ではカ行五段活用動詞といった細かな品詞の分類が必要であり、例えば[9]では動詞を17種、活用語尾を含む附属語を71種の品詞に分類していた。しかし、本システムでは動詞の区別は子音幹動詞と母音幹動詞の2種類しかなく、また活用語尾も存在しないため、附属語も35種類と少なくすむ。これにより接続行列も比較的小さくなる。

#### 長所 2: 形態素の種類が少数である

派生文法を用いるもう一つの利点は、派生接尾辞を用いることにより辞書に登録する動詞の数を少なくすることが可能な点である。「散歩スル」などのいわゆるサ変動詞は名詞「散歩」に母派生接尾辞“-su-”が接続した形として解析される。「寒ガル」は形状動詞の語幹「寒-」に形子派生接尾辞“-gar-”が接続したもとして解析できる。こうした解析は従来JUMAN[10]などでも行われ辞書に登録する動詞の数を減らすのに役立っている。

本システムではさらに可能動詞の登録も省くことができる。伝統的文法では五段活用動詞が可能の意味を表すとき、一段活用する別の動詞になる考えられてきた。例えば「書ク」に対しては「書ケル」が別の動詞として辞書に登録された。しかし本システムは“-e-”という可能の子母派生接尾辞を導入することにより“書keru”は“書k-e-ru”と解析され、可能動詞を辞書に登録することなく解析できる。

#### 長所 3: 辞書登録する動詞を少なくできる

また[8]でも指摘されている通り、口語的な表現への対処が容易である。いわゆるら抜き言葉である「食バレル」は可能の母派生接尾辞“-re-”を登録することで対処できる。また「書カセル」「食バサセル」に対する「書カス」「食バサス」といった表現も、形子派生接尾辞“-as-”と母子派生接尾辞“-sas-”を登録することで対処できる。

#### 長所 4: 口語的表現への対処が容易である

最後に、従来の音韻論に基づく手法の欠点であった子音幹動詞の異形態の登録[5][7]も、後方からの探索と末尾子音の補完を用いることにより、本システムでは必要ない。また子音幹末尾wの補完も同様に行うことができる。

#### 長所 5: 従来の音韻論に基づく手法の欠点であった音便処理のための動詞の異形態の登録が必要ない

しかしながら、本システムは音韻論的解析であり、入力文をローマ字表記に変換しなくてはならない。その結果、解析する文字列が長くなり辞書引き

の効率が落ちるといいう短所がある。

## 5 実験結果

現在のシステムはまだ実験段階であり、複数の解析候補がある場合には単純な後方最長一致法を用いて解を一意に決定している。本システムでEDR日本語コーパスから150文を取り出し動詞の出現421個に対して形態素解析を行った。その結果396個(94.1%)については正しく解析できた。失敗した例としては、「モデルとなり」の「なり」を動詞“nar-i”ではなく助詞“nari”と解析する最長一致法が原因のものや、可能的接尾辞“e”と命令の接尾辞“e”の区別で失敗したものなどがあつた。これらは今後最小コスト接続法[6]などを導入して改善できると予測される。

本システムでは後方からの検索という人間の直観に反する解析を行っているが、それが原因となる失敗例はなかった。また本システムの特徴である子音の補完による失敗例もなかった。以上から派生文法に基づく少数の品詞分類でも、十分な精度の形態素解析が行えるものと考えられる。

#### 参考文献

- [1] Tanaka Hozumi, Tokunaga Takenobu and Aizawa Michio. *Integration of Morphological and Syntactic Analysis based on LR Parsing Algorithm* 自然言語処理 第2巻第2号, 1995, pp.59-74
- [2] 小川泰弘, ムフタル・マフスット, 稲垣康善. “派生文法による日本語動詞接尾辞の形態素解析” 人工知能学会全国大会(第10回)論文集, 1996, 人工知能学会
- [3] 清瀬義三郎則府. “日本語学とアルタイ語学” 1991, 明治書院
- [4] 清瀬義三郎則府. “日本語文法新論 - 派生文法序説 -” 1989, 桜楓社
- [5] 西野博二, 鷲北賢, 石井直子. “派生文法による日本語構文解析” 自然言語処理研究会 87-6, 1992, 情報処理学会
- [6] 久光徹, 新田義彦. “接続コスト最小法による形態素解析の提案と計算量の評価について” 信学技報 Vol. 90, No. 116, pp. 17-24, 1994, 情報処理学会
- [7] 久光徹, 新田義彦. “日本語形態素解析における効率的な動詞活用処理” 自然言語処理研究会 103-1, 1994, 情報処理学会
- [8] 淵武志, 米澤明憲. “日本語形態素解析システムのための形態素文法” 自然言語処理 Vol. 2, No. 4, pp. 37-65, 1995
- [9] 丸山宏, 萩野紫穂. “正規文法に基づく日本語形態素解析” 情報処理学会論文誌 Vol. 35, No. 7, 1994
- [10] 松本裕治, 黒橋禎夫, 宇津呂武仁, 妙木裕, 長尾真. “日本語形態素システム JUMAN 使用説明書 version 2.0” 1994年, 奈良先端技術大学院大学技術論文
- [11] 日本電子化辞書研究所. “EDR 電子化辞書 1.5 版仕様説明書” 1996