

# 形態素解析用辞書における動詞の形態・意味関係の記述

加藤 恒昭<sup>1</sup> 林 良彦<sup>2</sup> 伊藤 たかね<sup>1</sup>

東京大学 大学院総合文化研究科<sup>1</sup> 大阪大学 大学院言語文化研究科<sup>2</sup>

## 1 はじめに

形態素解析は言語処理、言語理解の入口で、その結果得られた形態素や語は、文の統語や意味に関する情報を提供する構造の構成要素となる。一方で、これらの形態素や語もその間に意味的な関係を持ち、その内部に構造を持っている。それらの情報はソーラスや語彙的オントロジとして整理されるものであるが、実際の言語処理の過程で得られるもの、つまり形態素解析の結果である形態素解析用辞書の見出し語をベースとした記述は、利用しやすさの点から言語処理に直接の利益をもたらすと思われる。加えて、形態素解析処理に用いられる辞書は工学的な考慮からか、その後の統語処理、意味処理にとって系統的でない部分もあるが、それに起因する問題もこのような記述によって緩和できると期待される。

本稿では、動詞に興味を絞り、chasen, mecab<sup>1</sup>等で利用される形態素解析用辞書 IPADIC (version 2.70) 中の動詞を対象として行った形態・意味関係の記述について報告する。まず、設計、記述の方針を説明し、その後、得られた記述の特徴を示す幾つかの統計量を示す。

## 2 動機

本稿で報告する関係記述は、形態素解析の結果を構成する語や形態素（見出し語）のうち、動詞に着目し、その間に存在する様々な形態・意味関係を明らかにした言語資源を構築することを目的とする。これを通じて、語義の記述や、含意等、語義間の意味関係の記述を体系的に行うための基盤を提供することを目指している。例えば、「こわれる」「壊れる」「壊す」「壊せる」「取り壊す」はそれぞれ独立した見出し語として辞書中に登録されているが、「こわれる」が「壊れる」の異表記であること、「壊れる」と「壊す」が対応する自動詞と他動詞であること、「壊せる」が「壊す」の可能形であること、「取り壊す」が複合動詞であり「壊す」がその構成要素であることが示されていれば、「壊れる」の語義に基づいて、これらの辞書項目の語義をそれと関係づけて体系的に表現できると期待される。もちろん、対応する自動詞と他動詞の間の意味関係が常に同一とは限らない [1] し、複合動詞においてはその語義とその構成要素の語義の間には（関係がない

場合を含めて）様々な関係があることが知られており [7]、その分析は今後の課題 [2] であるが、これらの関係記述はそのような分析を行うための資源としても必要となってくる。また、語義を陽に表現することが困難であるとしても、見出し語間の意味関係は、「オモチャを壊した」のであれば「オモチャが壊れた」というようなテキスト間の含意関係認識や、「オモチャを壊すことができる」を「オモチャを壊せる」とするような言い換え等に利用できるような、有益な情報を提供すると考えられる。

## 3 関係記述の方針

ここでは、IPADIC の見出し語を対象として、以下の3種類の記述を行った。

1. IPADIC の見出し語は、表記、読み、活用型で特徴づけられる。それに対し、語（動詞）を関係記述の基本要素として定義し、ある語とその語の表記となりうる表記を持つ見出し語とを関係づけた。見出し語どうしの異表記関係も、この関係を通じて得ることができる。
2. 動詞間を、可能関係、使役関係、能動受動関係、自他関係という形態・意味関係で関係づけた。
3. 動詞を語構成の観点から分類し、派生動詞、複合動詞である場合は、その構成要素と関係づけた。

以下、それぞれの関係について詳しく述べる。なお、これらの記述は図1に示す構造を持つ情報としてまとめられている。

### 3.1 語の定義と異表記関係

関係記述の要素となる語（動詞）を、同じ読み、活用型を持ち、送り仮名の違い等、漢字仮名交じり表記の揺れと判断される表記を持つ見出し語の集まりとして定義する。ここでの表記の揺れは漢字とひらがなの使用法の違いをいい、異なる漢字を含むものは異なる語と判断する。例えば、「取り壊す」と「取壊す」は同じ語の異なる表記であるが、「取り壊す」と「取り毀す」は異なる語の表記とされる。そして「とりこわす」はこのふたつの語に関係づけられる。このように定義された語に代表表記を与える。代表表記はその語と関係づけられた見出し語

<sup>1</sup><http://chasen-legacy.sourceforge.jp/>.  
<http://mecab.sourceforge.net/>.

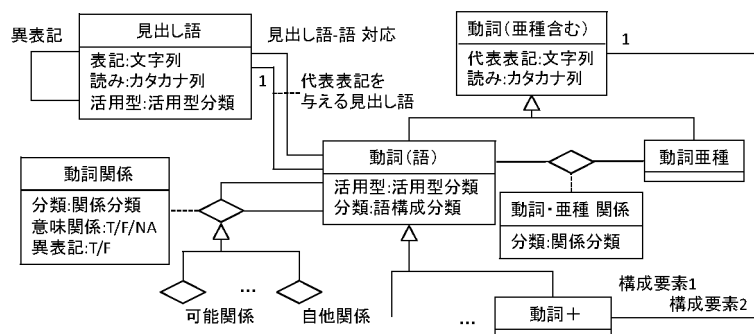


図 1: 動詞関係の情報構造

の表記のうち、漢字の使用が最多のものの中で文字数が最多のものを選択した。以下、表記もしくは見出し語は「」で囲むことで示し、語はその代表表記を [] で囲むことで表現する。

### 3.2 動詞間の形態・意味関係

寺村は態を「補語の格と相關関係にある述語の形態の  
関係」<sup>2</sup>とし、それを形態、統語、意味の3つの面から特  
徴づけることが必要だとしている。そして、格の移動（変  
化）と対応する動詞の形の中に予見可能的に出没する形  
態素があれば「文法的態」、予見不可能な対応であれば  
「語彙的態」であるとして、文法的態として受動態、可能  
態、自発態、使役態を、語彙的態として自動詞・他動詞  
の対立をあげ、それぞれの態と認定される動詞の形態的  
特徴を整理している [6].

本記述では、この考え方を参考にして、これらの態で用いられる動詞の形態的特徴、意味的特徴をその元となる動詞との関係として記述した。具体的には図2にその定義を示す6種類の関係を付与した。このうち、可能関係、使役関係1、使役関係2、能動受動関係1は、文法的態に関連するとして、最初に形態関係として定義し、その後、それに加えて意味関係が成り立っているかを判断して、その有無を記述している。例えば「結び付く」と「結び付ける」は形態的に可能関係にあるが、その意味関係は存在しないとされている。能動受動関係2と自他関係は語彙的態に関連するため、意味的な判断に基づいている。なお、自発態については、形態関係が可能関係と同一で、後述のように意味関係の判断は自他関係と同一であるとしたために独立した記述となっていない。

なお、可能形や受動形はひらがな表記しかもたない語も多く、それらと漢字仮名交じり表記を持つ語とは、異表記であることを表わす異表記フラグを立てて関係づけている。例えば、「つられる」という辞書項目はあるが、「釣られる」「吊られる」はないため、[つられる]は、[釣る]や[吊る]と異表記フラグが立った能動受動関係で関係づけられる。

自他関係はそれ以外の関係と意味的な重なりがあり、その付与に判断基準が必要になる。ここでは以下の基準を採用した

- 可能関係は自他関係と意味的な重なりを持たないので、問題とすべきはこれと同じ形態関係を持つ自発関係である。自発関係の条件を自他関係のそれと同じとし、自発関係にある対にはすべて自他関係を付与する。
- 使役1もしくは使役2の関係を持つ v1,v2 で、意味関係が成り立っている対については、v2 のヲ格となるものが、意志を持った主体に限られない場合に v2,v1 に自他関係を記述する。ただし、ヲ格が意志を持たないものであっても慣用的に定まった身体部位や精神、コンピュータ等に限定される場合は、自他関係を付与しない。[遊ぶ] - [遊ばせる] は、「土地を遊ばせる」から自他関係を付与し、[働く] - [働かす] は「頭を働かす」が慣用的であるとして、自他関係を付与しなかった。
- 能動受動関係1にある v1,v2 で、意味関係が成り立っている対については、自動詞 v2 を含んだ文において二格や「～によって」で示されるような行為者の存在を考える必要性がなく、自然もしくは自発的に事態が生じると考えられる場合に、自他関係を付与する。[産む] - [産まれる] と [生む] - [生まれる] の対に自他関係が付与されている。
- 能動受動関係2にある v1,v2 の間には自他関係は付与しない。

- 使役 1 もしくは使役 2 の関係を持つ v1, v2 で、意味関係が成り立っている対については、v2 のヲ格となるものが、意志を持った主体に限られない場合に v2, v1 に自他関係を記述する。ただし、ヲ格が意志を持たないものであっても慣用的に定まった身体部位や精神、コンピュータ等に限定される場合は、自他関係を付与しない。[遊ぶ] - [遊ばせる] は、「土地を遊ばせる」から自他関係を付与し、[働く] - [働かす] は「頭を働かす」が慣用的であるとして、自他関係を付与しなかった。

- 能動受動関係 1 にある v1,v2 で、意味関係が成り立っている対については、自動詞 v2 を含んだ文において二格や「～によって」で示されるような行為者の存在を考える必要性がなく、自然もしくは自発的に事態が生じると考えられる場合に、自他関係を付与する。[産む] - [産まれる] と [生む] - [生まれる] の対に自他関係が付与されている。

- 能動受動関係 2 にある  $v_1, v_2$  の間には自他関係は付与しない.

### 3.3 語構成の観点からの分類

語構成の観点からは、動詞は、活用語尾を除く意味的な単位である形態素、語基ひとつからなる単純動詞、複数の語基からなる複合動詞、語基に接辞が伴った派生動詞に分類される。ここで接辞とは他の語に付随してはじめて機能を発揮する、いわゆる接頭辞・接尾辞である [4]。この観点に基づき、動詞を図 3 に示す 14 種類に分類した。

<sup>2</sup>p.208. ここでの「述語」は本稿における動詞に対応する。

可能関係 動詞 v1 と v1 の語幹+eru の形態<sup>4</sup>をもつ動詞 v2 との関係<sup>5</sup>. v2 を v1 の可能形と呼ぶ. このうち, v2 が「v1 することができる」という意味を持つ対を意味関係ありとする.

使役関係 1 動詞 v1 と v1 語幹+(s)asu の形態をもつ動詞 v2 との関係. v2 を v1 の使役形と呼ぶ. このうち, 「A が B に v2 する」と「B が v1 する」という含意を持つ対を意味関係ありとする. B は意志を持った主体 (人間や動物等) に限定しないので, 自他対応の関係にある v1, v2 も使役関係 1 が意味的に成り立つものがある.

使役関係 2 動詞 v1 と v1 の語幹+(s)aseru の形態をもつ動詞 v2 との関係. v2 を v1 の使役形と呼ぶ. 意味関係の有無の判断は使役関係 1 と同様.

能動受動関係 1 動詞 v1 と v1 の語幹+(r)areru の形態をもつ動詞 v2 との関係. v2 を v1 の受動形と呼ぶ. 「A が B を v1 する」と「B が A に v2 する」もしくは「A が B に v1 する」と「B が A に v2 する」([取り付く] 等) という含意を持つものを意味関係ありとする.

能動受動関係 2 形態的に能動受動関係 1 にない v1, v2 で, 「A が B を v1 する」と「B が A に v2 する」もしくは「A が B に v1 する」と「B が A から v2 する」という含意を持つ関係. v2 を v1 の受動形と呼ぶ. [教える] - [教わる], [捕まえる] - [捕まる] 等が含まれる.

自他関係 「A が B を v1 する」と「B が v2 する」という含意を持つ他動詞 v1 と自動詞 v2 との関係.

図 2: 形態・意味関係の定義

単純 [歩く] [壊す] 等, ひとつの語基からなる単純動詞.

動詞+ [打ち壊す (打つ, 壊す)] [売れ残る (売れる, 残る)] 等, 動詞連用形+動詞の複合動詞.

名詞 [大人びる (大人)] [大人ぶる (大人)] [春めく (春)] 等, 名詞からの派生動詞. [愛す (愛)] や, [映ずる (映)] のように漢字 1 文字の音読みに「す」「する」が付属した一字漢語動詞を含めた.

名詞+ [色褪せる (色, 褪せる)] [役立つ (役, 立つ)] 等, 名詞+動詞の複合動詞.

形容 [暖める (暖かい)] [楽しむ (楽しい)] 等, 形容詞からの派生動詞.

形容+ [近付く (近い, 付く)] [若返る (若い, 返る)] 等, 形容詞+動詞の複合動詞.

形動 [静まる (静か)] [馬鹿げる (馬鹿)] 等, 形容動詞からの派生動詞. 形容動詞は「～な」で名詞を修飾できることで, 形容詞や名詞と区別する.

形動+ [主立つ (主, 立つ)] [真面目くさる (真面目, くさる)] 等, 形容動詞+動詞の複合動詞.

接辞 [ざわめく (ざわ, めく)] [べとつく (べと, つく)] 等, 単独で語にならない接辞や擬態語の一部に「めく」「つく」等が付属した派生動詞. 「つく」を接辞としている.

接辞+ [すつ飛ぶ (すつ, 飛ぶ)] [蹴落とす (蹴, 落とす)] 等, 単独で語にならないような接辞や擬態語の一部と動詞からなる, 動詞由来の派生動詞.

サ動+ サ変動詞の連用形+動詞の複合動詞. [理解し合う] [和解し合う] とその可能形のみ.

句 [悦に入る] 等の助詞を含む慣用句, [打って出る (打って, 出る)] 等のテ形を含んだもの等, 語と考えづらいもの.

古語 現代仮名遣いでないもの.

不明 国語辞典にも掲載がなく分類が不明なもの.

図 3: 語構成に基づく動詞分類の定義

複合動詞や派生動詞の構成要素が動詞である場合は, 本記述の動詞と関連づけた. ただし, この対応づけは, 以下のような場合があり必ずしも単純ではない.

- 構成要素となる語が, 単独では極めて稀にしか用いられないために辞書 (本記述の構成要素である語の集まり) 中に含まれない. これには, (1) [言い古す] の [古す] [いきり立つ] の [いきる] (「熱る」) のように同じ読みで漢字異表記を持った語が存在しない場合と, (2) [付き随う] の [随う] [極め付ける] の [極める] のように同じ読みで漢字異表記と判断される表記の語 ([従う] [決める]) が含まれる場合がある.
- 例えばある語の可能形であるような複合動詞において, その構成要素は可能形が辞書中に含まれていない. 例えば [生き返れる] の [返れる] は辞書に含まれておらず, [返る] だけが存在する.
- 構成要素部分がひらがな表記のみで, 関係づけのために意味的な判断が必要になる. 例えば [さきはじめる] の [さく] [はじめる] は語として存在せず,

「さく」「はじめる」が [咲く] [始める] の表記であることの判断が必要となる. これは動詞間の形態・意味関係で異表記フラグが立つ場合とほぼ同じ状況である.

- 形態的には動詞のようであるが, 実は動詞ではなく, その意味で複合動詞という判断が不適当なものがある. [言い付かる] は, 一見 [付かる] という動詞を構成要素とするようであるが, [付かる] という動詞は存在せず, [言い付かる] 全体として [言い付ける] から派生したと考えられる.

これらについて, 辞書中の動詞と何らかの形で関係づけられる場合, つまり第一の分類の (1) の場合以外, 辞書中に存在しない動詞を動詞亜種として導入し構成要素とするという記述をし, その動詞亜種と辞書中の動詞の関係を別に記述することで, できるだけ多くの構成要素

<sup>4</sup>益岡らの日本語文法 [3] に従い, 五段活用動詞を子音までを語幹とする子音動詞と考える.

<sup>5</sup>動詞 v1 と v1 語幹+(ra)reru も可能関係となるが, このような対は辞書中に含まれなかった. このため, 可能関係と自発関係は形態的に重なることになる.

表 1: 付与された形態・意味関係数

	関係数	意味関係	異表記	自他関係
可能	1,384	1,165	37	47
使役 1	165	132	7	70
使役 2	108	73	17	28
能動受動 1	65	61	47	2
能動受動 2	9	-	-	-
自他	584	-	-	-

表 2: 語構成からの動詞の分類

分類	動詞数	分類	動詞数
単純	3,112	形動	33
動詞+	2,904	形動+	3
名詞	523	接辞	51
名詞+	341	接辞+	215
形容	90	その他	68
形容+	23		

と動詞を関係づけるようにしている。例えば、[生き返れる]を扱うために、[返れる]という辞書に含まれない語を動詞亜種として導入し、[生き返れる]の構成要素はこの[返れる]とし、[返れる]と[返る]を可能関係で関係づけている。

## 4 記述の概要

得られた記述は以下のような統計量を持つ。

対象とした IPADIC には、動詞-自立の品詞分類を持つ見出し語が 14,819 項目あり、そこから 7,363 語の動詞が定義された。ある動詞が持つ表記数（ある動詞と対応づけられる見出し語数）の平均は 2.2 で、734 動詞がひとつの表記、4,888 動詞がふたつの表記を持つ。最も多くの表記を持つ動詞は[跳ね上がる]で、7 種類の表記を持つ。逆にある見出し語が対応する動詞の数の平均は 1.1 で、13,765 の見出し語がただひとつの動詞と対応する。最も多くの動詞と対応する見出し語は「かける」「たつ」「つく」「ひく」で、対応する動詞の数は 10 である。

記述された動詞間の形態・意味関係の関係数を、意味関係があるもの、意味関係がある中での異表記であるものの数と合わせて、表 1 に示す。可能関係は数が多く、子音動詞の可能形はかなり網羅的に辞書に含まれていることがわかる。一方、能動受動関係 1 は数が少ない上、その大半がひらがな表記された受動形についてのものであり、原則として辞書に含めない方針であったと想像される。

語構成の観点からの分類を表 2 に示す。「その他」は「サ動+」「句」「古語」「不明」に分類されたものを示す。森田は、現代語の小型国語辞書を調査し、単純動詞、動詞+動詞の複合動詞（本記述での「動詞+」）、その他の複合・派生動詞の割合が 47%、40%、13%であったと報告してい

表 3: 複合・派生動詞の構成要素の分類

	token 数	type 数
動詞	5,810	949
動詞亜種	491	281
不明	89	70
(合計)	6,390	1,300

る [4] が、本記述では 42%、40%、18%で、ややその他の複合・派生動詞が多くなっている。

「その他」以外の動詞の「動詞」と分類された構成要素について、本記述の動詞、動詞亜種、対応づけられなかったもの（不明）のいずれであったかを表 3 に示す。それぞれの複合動詞の構成要素として現れた数の合計（token 数）と同じ動詞等をまとめた数（type 数）を示している。1,300 種類の構成要素からその 3 倍近い 3,500 程度の複合動詞が得られている。

## 5 おわりに

形態素解析用辞書 IPADIC 中の動詞を対象として行った形態・意味関係の記述について報告した。本記述が言語処理にとって有益なものとなることを期待している。今後、他の言語資源との関係、特に意味の記述を持つ国語辞書等との比較や対応づけを行うと共に、複合動詞の意味分析等の実際の場面で利用していく予定である。

## 謝辞

本研究は、挑戦的萌芽研究「大規模語彙知識を融合した語彙概念構造体系の構築」および財団法人電気通信普及財団より助成いただいている「日本語意味理解のための動詞語彙概念構造の研究」の一部として進められている。ご理解とご支援をここに深く感謝する。

## 参考文献

- [1] 伊藤たかね. 日本語自他交替動詞の完結性と意図性—大規模辞書構築の現場からの予備的考察—, 今西典子編 言語研究の宇宙—長谷川欣佑先生古稀記念論文集, 開拓社, 2005.
- [2] 加藤恒昭, 林良彦, 伊藤たかね. 語釈文を用いた複合動詞の特徴分類, 第 17 回言語処理学会年次大会, F2-3, pp.568-571, 2011.
- [3] 益岡隆志, 田窪行則. 基礎日本語文法—改訂版—, くろしお出版, 1992.
- [4] 森田良行. 動詞・形容詞・副詞の事典, 東京堂出版, 2008.
- [5] NTT コミュニケーション科学基礎研究所監修. 基本語データベース, 学習研究社, 2008.
- [6] 寺村秀夫. 日本語のシンタクスと意味 I, くろしお出版, 1982.
- [7] 由本陽子. 複合動詞・派生動詞の意味と統語, ひつじ書房, 2005.