

形態素解析用辞書における動詞の形態・意味関係

加藤 恒昭^{1,a)} 林 良彦^{2,b)} 伊藤 たかね^{1,c)}

概要:

形態素解析で用いられる辞書に含まれる動詞について、その形態・意味関係の記述を行った。異表記関係にある項目をまとめて語(動詞)を定義し、それらの間に、可能、使役、受動等の文法的態、自他対応という語彙的態に関する関係を記述した。加えて、動詞を語構成の観点から分類し、その構成要素と関連づけた。本稿では、これらの関係記述の方針を述べると共に、その基本的な統計量を報告する。加えて、別の言語資源との対応関係を調査して得られた、それらとの整合性について報告する。

Morpho-semantic relations between verbs in a dictionary for morphological analysis

TSUNEAKI KATO^{1,a)} YOSHIHIKO HAYASHI^{2,b)} TAKANE ITO^{1,c)}

Abstract: Morpho-semantic relations between verbs in a dictionary used for morphological analysis were annotated. A word was defined as a collection of its orthographical variations, and words were associated with each other in terms of relations on grammatical voices such as potential, causative and passive, and lexical voice of transitive-intransitive correspondences. Verbs were also classified according to their composition, and associated to their components when those were compound or derivative verbs. This paper explains the policy of this relation annotation, and reports its basic statistics. In addition, based on the investigation on its correspondences to other language resources, the consistency among them and possibility of coordination of these resources are discussed.

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。それらの情報はシ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辞書項目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配慮からか、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^{*1}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的関係とそれに関与する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し、他の言語資源との比較を行う^{*2}。

2. 動機

形態素解析用辞書への関係記述の目的は、形態素解析の結果を構成する語や形態素(辞書項目)の間に存在する様々な形態的・意味関係を明らかにした言語資源を構築することで、これを通じて、語義の記述や、含意等、語義間の意

¹ 東京大学 The University of Tokyo

² 大阪大学 Osaka University

a) kato@boz.c.u-tokyo.ac.jp

b) hayashi@lang.osaka-u.ac.jp

c) ito@boz.c.u-tokyo.ac.jp

^{*1} <http://chasen-legacy.sourceforge.jp/>.

<http://mecab.sourceforge.net/>.

^{*2} 本稿の前半は [4] に加筆を行ったものであるが、その後の見直しによって、何点かの仕様や判断の変更がなされている。

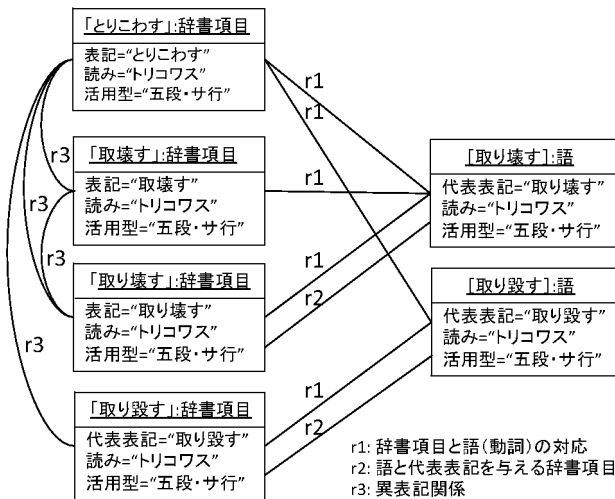


図 2 語と見出し語の関係

名交じりの表記を持たない語の存在が考慮されておらず、ひらがな表記を持つ辞書項目は、それと同じ読みと活用型の、漢字を含む表記を持つ辞書項目があれば、その語と異表記の関係にあるとされてしまう。この問題を解決するために、5節で述べる2つの言語資源からひらがな表記しか持たない辞書項目を抜き出し、その語義を持つような(漢字仮名交じり表記あるいはひらがな表記を持つ)語が存在しているかを人手で確認し、存在していない場合は新しい語として追加した。結果、[ときめく][ませる][させる]の3語が加わっている(この処理をしないとこれらのひらがな表記は「時めく」「増せる」「刺せる」等の異表記となる)。

3.2 動詞間の形態・意味関係

寺村は態を「補語の格と相関関係にある述語の形態の関係」^{*3}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沒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[9]。

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3にその定義を示す7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]-[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

^{*3} p.208. ここでの「述語」は本稿における動詞に対応する。

されている。使役関係3、能動受動関係2、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、後述のように形態関係が可能関係と同一で、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。

実際に IPADIC の語を調べてみると、可能関係では語幹+eru、使役関係1では語幹+asu、使役関係2では語幹+aseruのものだけが登録されていることがわかった。つまり、母音動詞(一段活用動詞)については、その可能形や使役形は語として登録されていない。このため、寺村の自発態は、形態関係が可能関係と完全に重なることになる。

また、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。ちなみに、自発関係のほとんどの対は[割る]-[割れる]のように意味的に可能関係との曖昧さを持つ。自発関係のみ成り立った対は[擦りむく]-[擦り剥ける]のみであった。
- 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。
- 使役3は他動詞どうしの関係であるので、自他関係は付与しない。
- 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

可能関係 動詞 v1 と v1 の語幹+(ra)reru の形態^{*4}をもつ動詞 v2 との関係。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 にも使役関係 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。

使役関係 3 形態的に使役関係 1,2 にない共に他動詞の v1,v2 で、使役関係 1 と同じ意味関係が成り立つもの。実際には [着る] - [着せる] 等, x+ru, x+seru の形態をもつ動詞の対に限られている。

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 3 形態・意味関係の定義

しない。

使役と自動詞・他動詞の関係について、森田 [7] は他動詞であるのに「～せる」の形にまで拡大使用することになった A 型、本来「～せる」を伴って使役表現とするところを A 型に引かれて「～す」の形となった B 型あるいは使役動詞 ([9] では「使役のつづまったもの」と呼んでいる) という区別をしているが、本記述において使役関係 2 と自他関係の両方が付与されたものが A 型に、使役関係 1 のみが付与され自他関係は付与されないものが B 型に対応する。

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 [7]。この観点に基づき、動詞を図 4 に示す 14 種類に分類した。「品詞」がその品詞である語に由来する派生動詞、「品詞+」がその品詞である語と動詞からなる複合動詞である。ただし動詞中心の分類名となっており、「接辞+」は一般には動詞由来の派生動詞とされるものであるが、接頭辞と動詞からなる複合動詞であるかのような分類名になっている。「接辞」には、単独では語にならない接辞や擬態語等と接尾辞からなると考えられる語を分類している。その他に一般には現代語の語とはされないものを整理する幾つかの分類がある。

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 IPADIC 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も含まれない場合と、(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]) は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 IPADIC 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 (ちなみに「咲き始める」という辞書項目があれば「さきはじめる」は同じ語の異表記としてまとめられるのでこのような問題は生じない)。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。理論的には不適切かもしれないが、これらについても、ある程度系統的な扱いが期待できるものについては複合動詞と分類している。

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と動詞を関

*4 益岡らの日本語文法 [6] 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。[買う] [言う] 等、ワ行五段活用動詞は語幹に w を含むとしている。

- 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+ [すっ飛ぶ (すっ, 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 4 語構成に基づく動詞分類の定義

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.

4.1 辞書項目と語

対象とした IPADIC には, 動詞-自立の品詞分類を持つ辞書項目が 14,819 項目あり, そこから 7,364 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辞書項目の数) の平均は 2.2 で, 735 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. 逆にある辞書項目が対応する動詞の数の平均は 1.1 で, 13,765 の辞書項目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辞書項目は「かける」「たつ」「つく」「ひく」で, 対応する動詞の数は 10 である.

4.2 形態・意味関係

記述された動詞間の形態・意味関係の係数, 意味関係があるもの, 意味関係がある中で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ある程度網羅的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, 原則として含めない方針であったと想像される.

意味関係を考慮して, いずれの動詞の可能形, 使役形, 受動形でもないような動詞 (ただしその間に自他関係が成り立つ場合は除く. 以下, このような動詞を基本動詞と呼ぶ) は, 6,148 語であった. ただし, この中には [あぶれ

表 1 付与された形態・意味関係係数

	係数	意味関係	異表記	自他関係
可能	1,382	1,164	35	51
使役 1	161	132	7	74
使役 2	100	69	17	26
使役 3	4	-	0	-
能動受動 1	65	61	47	2
能動受動 2	9	-	0	-
自他	612	-	24	

る] のように他の語 ([炙る] [焙る]) の可能形等でありながら, それとは異なる語義 (「仕事にあふれる」) を持つものは含まれていない (ちなみに [溢れる] は「あふれる」「こぼれる」の読みしか持っていない).

4.3 自他対応

他動詞と自動詞の対応は, 形態的 [9] に, あるいは活用形式 [7] から, 幾つかの型に分けることができる. それらの分類に属するもので本記述での頻度が大きいものを例とあわせて表 2 に示す. また, 「動詞+」の複合動詞はふたつの動詞を構成要素として持つため, 独特の自他対応の形態となる場合がある. その頻度を表 3 にまとめる. ここで, v1 は前方, v2 は後方の構成要素を表わす.

4.4 語構成

語構成の観点からの分類を表 4 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る [7] 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.

表 2 自他対応の形態的分類

他動詞	自動詞	(例)	頻度	備考
x-eru	x-aru	([上げる] - [上がる], [混ぜる] - [混ぜる])	135	
x-eru	x-u	([浮かべる] - [浮かぶ], [育てる] - [育つ])	87	
x-asu	x-u	([動かす] - [動く], [飛ばす] - [飛ぶ])	74	使役関係 1
x-u	x-eru	([折る] - [折れる], [炊く] - [炊ける])	52	可能関係 (自発)
x-asu	x-eru	([枯らす] - [枯れる], [蒸らす] - [蒸れる])	51	
x-su	x-ru	([移す] - [移る], [通す] - [通る])	47	
x-aseru	x-u	([浮かせる] - [浮かぶ], [濁らせる] - [濁る])	26	使役関係 2
x-su	x-reru	([隠す] - [隠れる], [壊す] - [壊れる])	23	

表 3 複合動詞の自他対応の形態的分類

分類	(例)	頻度
v2のみ自他対応	([持ち上げる] - [持ち上がる], [焼き付ける] - [焼き付く])	68
v1のみ自他対応	([出し切る] - [出切る], [流し込む] - [流れ込む])	7
v1,v2共に自他対応	([生み出す] - [生まれ出る], [折り曲げる] - [折れ曲がる])	7

表 4 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		

表 5 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

複合動詞(「品詞+」と分類されたもの)の構成要素のうち、動詞であるものが、IPADIC中の動詞(動詞)、IPADIC中の動詞と関係づけることができる動詞亜種(動詞亜種)、そのどちらでもなかったもの(その他)のいずれであったかを表5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語程度の複合動詞が得られていることがわかる。

5. 他の言語資源との対応

言語理解には様々な言語資源が必要となる。それらの言語資源の内容は一定の整合性を持ち、有機的に組み合わせ活用できることが望ましい。本節では、本記述を用いたIPADICと他の言語資源との対応づけの可能性や、本記述と他の言語資源との整合性の程度について検討する。

5.1 基本語データベースとの語の対応

基本語データベース[8]は、人間向けの国語辞書を基に作成された計算機可読辞書であり、一般の辞書に見られるような語義の記述である語釈文を備えている。この情報を利用して、形態素解析結果に語義をふるることができれば、その後の意味処理での活用が期待できる。これがどの程度まで可能かを見るために、IPADICと基本語データベースに含まれる語の重なりを調査した。

基本語データベースは、3,117語の見出し語(カタカナによる表記、つまり読み)を持ち、それらと表記の組み合わせは3,657項目となる。この見出し語と表記の組み合わせを語と考える。見出し語に較べて表記があまり多くないことからわかるように、形態素解析用辞書とは違って、漢字仮名交じり表記の揺れを網羅することは考えられていない。またひらがな書きの表記も多くない。

本記述で定義したIPADICの語はひとつの読みとそれに対応づけられたいくつかの表記を持つ。基本語データベースの語 W_L の読みがIPADICの語 W_I の読みと一致し、 W_L の表記が W_I の表記のいずれかと一致する時、 W_L と W_I が対応すると定義する。

IPADICの語がどの程度基本語データベースのそれと対応するかを、単純動詞とその他(複合・派生動詞等)、いずれの動詞の可能形、使役形、受動形でもない基本動詞とその他という2次元で分類して、まとめたものを表6に示す。ここで、「対応」は上記の意味で対応するものの数を示す。「対応?」は、読みもしくは表記のいずれかが一致する語が存在した語の数を示す。「対応せず」はそれ以外で、いずれかが一致する語も存在しないような語の数である。

「対応」の場合、形態素解析結果として得られた語の語義に関する情報を基本語データベースから得ることができ、「対応せず」の場合、それを得ることはできない。「対

表 6 IPADIC の語と基本語データベースの語との対応

		対応	対応?	対応せず	総数
基本動詞	単純動詞	1,235	673	587	2,495
	その他	1,464	334	1,855	3,653
その他	単純動詞	82	98	438	618
	その他	15	2	581	598

表 7 基本語データベースの語で IPADIC に含まれないものの内訳

		対応?	対応せず	総数
基本動詞	単純動詞	45	4	49
	その他	140	166	303
その他	単純動詞	8	23	31
	その他	4	1	5

「対応?」は以下のように位置づけられる。基本語データベースは、表記のバラエティはあまり多くなく、例えば「うたう」の読みを持つ語は「歌う」のみである。一方、IPADIC はそれに加えて「唄う」「唱う」等、6語を含んでいる。これらの語の幾つかには「歌う」と同じ語義を与えられると考えると、「対応」と「対応?」を合わせた語数は、語義を付与できる語の最も楽観的な予想数となる。

基本語データベースの語で IPADIC 中に対応する語を持たないものを同様の観点で分類して示したのが、表 7 である。この分類はその形態素解析結果を参照して人手で行っている。基本語データベースには「言い争い」「塗り残し」等、動詞とは考えにくい 16 語が動詞として含まれているが、それを除いて IPADIC の語に対応しない語の総数は 388 であった。その内訳を示している。

IPADIC に較べて基本語データベースに含まれる語の数は少ないので、ある程度は当然であるが、形態素解析で得られる語に対して語義をふるることができる割合は充分に多いわけではないことが観察される。

形態素解析の結果、基本動詞かつ単純動詞として解析されたものについて、悲観的にみても 50% (1,235/2,495)、楽観的にみても 74% (1,908/2,495) の語にしか語義が付与できないことが示されている。加えて、楽観的になるためにはその異表記語、異音語の対応づけが必要になる。対応しない語については（それらが使用頻度の低い語であることが予想されるとはいえ）、語義を新たに記述するしかなく、意味に関するより大規模な言語資源が待たれる。また、このような状況でありながら、比較的少数 (4-49 語) とはいえ、IPADIC に含まれない単語が基本語データベース中の存在するのは驚きである。網羅的な言語資源を構築することの困難さを示しているように思える。

基本動詞のうち、複合・派生動詞は、単純動詞以上に基本語データベースに含まれない IPADIC の語が多い。対応づけられるのは、楽観的にみても 50% 以下 (1,798/3,653) となっている。複合動詞の語義は、一部については構成要

素から構成的に得られるとはいえ、一般的にはそれらを構成要素の意味から推測する方法は確立されていない。これら複合・派生動詞の意味処理もしくは意味記述の充実が期待される。なお、IPADIC に含まれない基本語データベースの語が少なくない (166-303 語) ことも特徴で、一定の程度の生産性をもつ複合動詞の特徴を示している。

基本動詞でないものの多くが、国語辞書を基にした基本語データベースのような辞書に含まれていないのはむしろ当然で、これらについては、今回作成したような動詞関係の記述を利用してその意味を関連する基本動詞から得ることができる。基本動詞でない語がこのような辞書に含まれない場合、その語は関連する基本動詞から推測可能な語義しか持たないと考えられるからである。このような場面で本記述は語義を付与できる語の増大に貢献する。逆に基本動詞でない語が辞書項目になっている場合は、例えば「握らせる」のひとつの語義が「収賄の金などを渡す」であるように、ある語の使役形等ということだけからでは推測しづらい語義を持っていることが予想される。これらを考慮すると、形態素解析結果として得られた語だけを使って辞書からその語義を探すのではなく、動詞一語が得られた場合はその基となる基本動詞を、動詞と使役形等を作る接辞に分割された場合はそれらをまとめた一語をあわせて探す必要があることがわかる。

5.2 計算機用日本語基本辞書 IPAL との関係

計算機用日本語基本辞書 IPAL (以下、IPAL 辞書) [5] は、見出し語数は多くないが、動詞について、豊富な形態・意味情報を記述している。その中に、派生可能、派生使役、派生自他という、本記述の可能関係、使役関係、自他関係と類似した情報がある。これらの情報を対応づけ、整合性等の分析を行った*4。

IPAL 辞書は、861 語の見出し語 (読み) を持ち、それぞれの見出し語に複数の語義が結び付けられ、この語義毎に表記 (複数) や派生可能、派生使役、派生自他の関係が記述されている。なお、派生可能、派生使役、派生自他の関係先の語はひらがな表記で記述されている。読みと表記の対を語とすると、1,508 語が得られる。これと IPADIC の語との対応づけを行った結果、IPAL 辞書の語 1,284 語を IPADIC の語 1,201 語に対応づけることができた。対応づけられなかった 224 語のうち 103 語は対応づけられた語と漢字仮名交じり表記の揺れの関係にあった。残りの 121 語 (漢字仮名交じり表記の揺れを除くと 108 語) も、すべて同じ表記もしくは同じ読みを持つ語が IPADIC に含まれていた (前節の分類の「対応?」に相当する)。

対応づけられた 1,201 語を持つ本記述での可能関係、使役関係、自他関係と、IPAL 辞書での派生可能、派生使役、

*4 GKS 配布版を用いている。
<http://www.gsk.or.jp/catalog/GSK2007-D/catalog.html>

表 8 動詞関係についての IPAL 辞書との比較

	本記述 総数	IPAL 辞書 総数	一致する 情報数	IPAL のみ (語の不在)	IPAL のみ (判断の相違)	本記述のみ
可能	454	594	433	156	5	21
使役	141	842	64	777	1	77
自他	559	558	456	66	36	103

派生自他の関係の数と一致の程度を表 8 に示す*5。IPAL 辞書にのみ存在すると分類される関係は、その関係先である語が IPADIC 中に含まれない場合と、含まれてはいるがその関係にあると判断されていなかった場合に分けられる。前者を(語の不在)、後者を(判断の相違)としている。前述のように IPAL 辞書の関係先は読みしか示されていないが、IPAL 辞書の関係先である語が存在するかの判断は、それと同じ読みを持つ IPADIC の語が存在するかどうかだけでなく、存在する場合はそれらの語の漢字仮名交り表記を手で確認することで行っている。一方、本記述の情報が IPAL 辞書中に存在するかは、それらが同じ読みの語に関係づけられていることを条件とした。

それぞれの関係で傾向は異なるが、一致していない関係がかなり多数見られることがわかる。

可能関係においては、IPADIC に母音動詞の可能形や子音動詞であっても使用頻度の低いと思われる表記については可能形が登録されていないことが(語の不在)の原因となっている。例えば、「逢える」「会える」は登録されているが、「遭える」「遇える」の登録はない。使役関係も IPADIC では網羅的な登録からほど遠いことがわかる。一方で、本記述のみの関係がある程度数あるのは、IPAL 辞書では語幹+(s)areru を使役形としていないことが主な原因である。例えば、IPAL 辞書での「持つ」の使役形は「持たす」のみで「持たせる」は記述されていない。

自他関係について、(語の不在)が、比較的少ないとはいえる程度の数存在するのは、形態的に使役 1 の関係にある語が IPADIC に登録されないことが多いこと等が理由となっている。(判断の相違)と本記述のみの関係が多いのは、3.2 節でふれたような自他関係の意味判断の基準の相違が原因で、自発性の有無をどう捉えるか、使役関係や能動受動関係との境界をどこに引くか等のズレが影響している。前者の例として、IPAL 辞書では [釣る] - [釣れる] は自他関係で [割る] - [割れる] は自他関係ではないが、本記述ではその逆になっている。後者の例としては、IPAL 辞書で自他関係とされている [捕える] - [捕えられる] や [食べさす] - [食べる] が、本記述では能動受動関係、他動詞の間の使役関係とされている。

*5 IPAL の派生自他がその動詞からの自動詞もしくは他動詞への派生となっているため、その形式に合わせて集計した。本記述のひとつの自他関係を自動詞から他動詞、他動詞から自動詞と 2 回数えることになる。

6. 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。形態素解析辞書中の語は様々な形態・意味関係で関係づけられることが確認された。得られた記述は、形態素解析結果に対する意味的な扱いを支援できるものと考えている。その点で、本記述が多く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。著者らは、今後、この記述を複合動詞の意味分析等の場面で利用していく予定である。一方で、複数の言語資源を比較してみると、それぞれの規模(網羅性)、判断や方針の違いは予想以上に大きく、これらを有機的に組み合わせ活用することは容易ではないこともわかった。これらの間の、特に形態素解析辞書と国語辞書等との橋渡しについても今後検討を行っていききたい。

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] 伝康晴, 小木曾智信, 小椋秀樹他. コーパス日本語学のための言語資源: 形態素解析用電子化辞書の開発とその応用, 日本語科学, 22 号, pp.101-122, 2007.
- [2] 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.
- [3] 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.
- [4] 加藤恒昭, 林良彦, 伊藤たかね. 形態素解析用辞書における動詞の形態・意味関係の記述, 第 18 回言語処理学会年次大会, B4-2, pp.919-922, 2012.
- [5] 桑畑和佳子, 橋本三奈子, 村田賢一. 計算機用日本語辞書の開発, 情報処理学会研究会 人文科学とコンピュータ, 18-4, pp. 27-34, 1993.
- [6] 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.
- [7] 森田良行. 動詞・形容詞・副詞の事典, 東京堂出版, 2008.
- [8] 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.
- [9] 寺村秀夫. 日本語のシンタクスと意味 I, くろしお出版, 1982.
- [10] 由本陽子. 複合動詞・派生動詞の意味と統語, ひつじ書房, 2005.