

格情報を制約条件とした動詞階層化の試み

川田 亮一^{*1}
kawada@edr.co.jp

乾 裕子^{*2}

井佐原 均^{*3}
isahara@etl.go.jp

^{*1}日本電子化辞書研究所

^{*2}計量計画研究所

^{*3}電子技術総合研究所

あらまし

日本語では曖昧性の解消に動詞が重要な役割を果たしている。筆者らは現在、漸進的自然言語理解で用いるための動詞概念辞書を開発中であり、この辞書には、格情報を制約条件として記述する。その第一段階として動詞階層をボトムアップに作成した。本稿は動詞階層化に伴った幾つかの問題点を報告する。

キーワード

漸進的自然言語理解 動詞 シソーラス 制約条件 深層格

Verb Classification Based on the Constraints of Case Frames in Japanese

Ryoichi KAWADA^{*1}

Hiroko INUI^{*2}

Hitoshi ISAHARA^{*3}

^{*1}Japan Electronic Dictionary Research Institute

^{*2}The Institute of Behavioral Sciences

^{*3}Electrotechnical Laboratory

Abstract

In processing the Japanese sentences, verbs play important roles in disambiguation. We are developing a concept dictionary for verbs to be used for the Incremental Natural Language Understanding. This Dictionary describes information regarding constraints on case frames. As the first step of the development, We have classified verbs in a bottom up way. This paper reports some problems encountered in this step.

Keywords

Incremental Natural Language Understanding, Verb, Thesaurus, Constraint, Case Frame

^{*1} 日本電子化辞書研究所入社以前の研究に基づく。

1 はじめに

形態・構文・意味を統合的に処理する自然言語理解システムにおいては、解釈の組み合わせの爆発的な増大を避けることが重要な課題である。このため、文の完結を待たずに、文の要素の入力によって解析を進め、無用な曖昧性を消去する必要がある。これを漸進的自然言語理解 (Incremental Natural Language Understanding) と呼ぶ。

日本語の場合、述語が文あるいは節の最後に位置する。従って、述語に係る連用修飾句が現れた段階で述語を予測することが可能である。例えば、次の文を解析する場合を考えると、図1のようになる (図中、左側が入力文で、右側が述語の階層、点線で文の要素が入力された段階で候補となる述語の範囲を表している)。

- ・その折も折、法務当局が平沢の身柄を仙台拘置支所から東京の八王子医療刑務所に移送した。
(朝日紙、85.5.2 社説)

図1で「法務当局」が現れた段階では「法務当局」を主体/対象のガ格にとれる述語 (動詞・形容詞・形容動詞・述語名詞) 全体が候補である。次の「身柄を」が現れると、対象のヲ格をとる述語、すなわち他動詞全体が候補となる。この時点で、「身柄」という名詞の特殊性から、「移送する、留置する、拘束する、保護する…」などの動詞を候補に挙げることが出来るかもしれない。次の

「仙台拘置支所から」で始点のカラ格をとる動詞 (状態変化他動詞) に候補が絞られる。同時に「仙台拘置支所」という名詞の所性から物理的な移動を表す動詞に候補が絞られる可能性もある。

このように漸進的自然言語理解では述語を予測し、その範囲を限定することが必要であるが、このためには図に示したような、述語や名詞の辞書を持つことが前提となる。現在、この目的で利用するための動詞概念辞書を開発中であり、本稿ではその途中成果と問題点を報告する。

2 動詞概念辞書

動詞概念辞書とは、いわゆる動詞のシソーラスを拡張したものである。広義のシソーラスには様々なものがあるが^{[1] [2] [3] [4]}、ここでは次の前提に基づいたものを動詞概念辞書と呼ぶ。

- (1) 動詞の表層形をノード (エン트리) とするネットワーク形式の辞書である。
- (2) 各ノード (エン트리) には制約条件として格情報を記述する。
- (3) 各ノード (エン트리) は自分の親ノードの制約条件の論理積を継承する。

まず、動詞概念辞書は動詞の表層形をエン트리とし、その表層形によって表される概念を成立させる制約条件を記述する。同義あるいは概念的に関連する語であっても異なる表層形を持つ語は別

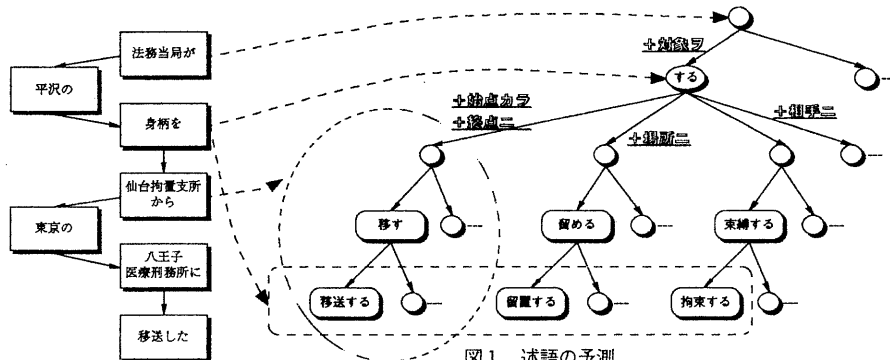


図1 述語の予測

のエントリとなる。例えば「減ぶ」と「減びる」は同義であるが、表層形が異なるので別エントリであり、「動く」と「動かす」（自他の対応）、「死ぬ」と「殺す」（深層格の対応）は後者が前者の概念を含んでいるが、表層形が異なるのでやはり別エントリとなる。

次に各エントリには格情報を制約条件として記述する。ここで、格情報とは動詞が要求する必須格、任意格、状況成分を問わず、該当する動詞によって表現される概念を他の動詞によって表される概念と識別するために必要となる情報である。従って、日本語では動詞の下位範疇化 (Subcategorization) としての表層格 (ガ格・ヲ格・ニ格等)、項の意味役割としての深層格 (主体・対象・相手・始点・終点など) を中心として、項に現れる名詞の範囲、その動詞が用いられる状況を限定する語用論的な情報 (位相など) が含まれる。

また、この格情報を制約条件としたネットワークを構成するために必要な仮想ノードを設定しても良い。例えば状態変化他動詞が一般的に次の制約条件を継承するならば、これに該当する表層の言語表現が存在しなくてもノードを設定できる。

ノード名：状態変化他動詞

制約条件：

- (A) 主体-ガ格 (人 組織)
- (B) 対象-ヲ格 (*)
- (C) 始点-カラ格 (*)
- (D) 終点-ニ/ヘ格 (*)

*は任意の名詞をさすこととする。

さらに、これらの制約条件は親ノードから子のノードに自動的に継承されるため、該当する動詞には差分の制約条件のみを記述すれば良い。また、格情報を制約条件とするために、同一の表層形であっても格情報が異なれば (多義語の場合)、別のエントリとなる。

以上のように、ここでの動詞概念辞書とは、概念の体系ではなく、動詞の体系である。また、概

念を直接記述するのではなく、概念を成立させるための制約条件を記述する。極端に言えば、例えば動詞概念辞書のほぼ最上位のレベルに「アル、ナル、スル」という言語形式が在ったときに、図2で言語形式を中心として右側が概念体系とすれば、左側がここで言う動詞概念辞書である。

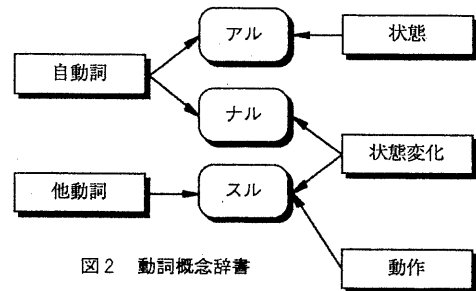


図2 動詞概念辞書

3 動詞概念辞書開発の概要

動詞概念辞書は次の4段階に分けて開発する。ここで、動詞階層とは木構造形式 (親ノードを一つしか持たない) の辞書を言う。

- (1) 対象動詞を選定し、その対象動詞に、ボトムアップに直上の上位語を付加し、部分的な動詞階層を作成する。
- (2) 部分的な動詞階層を統合し、より上位の上位語を付加することにより全体的な動詞階層を作成する。
- (3) 動詞階層の各エントリに制約条件を付加する。
- (4) 制約条件に基づき、動詞階層を調整し、必要があればネットワーク構造にする

全体的にボトムアップに辞書を開発するために、動詞概念辞書の中で下層に属すると予想されるサ変動詞を初期エントリとして、新聞記事から選定した。現在、(1)の作業が終了し、(2)の作業を行っている段階である。また、動詞概念辞書では制約概念が中心的な役割を果たすため、一部の動詞について (3)も試験的に行っている。以下、(1)

の作業を中心に報告する。

4 動詞階層の作成

動詞概念辞書開発の第一段階として、動詞階層の作成を行った。初期エントリーは、新聞記事から選定した漢語サ変動詞1436語である。

作業にあたり、以下の作成手順を設定した。基本的な方針は単独の作業者により、ボトムアップに作業を行い、途中成果をトップダウンに評価することにより一つ前の手順を修正し、これを交互に繰り返すことである。実際の作業の流れは図3に示した。

- (1) 各エントリーについて、自由連想(ただし、辞書等の参照は認める)によって上位概念の候補語をなるべく多く記述する。
- (2) 候補語を参考に、直上の上位語として最も妥当であると思われるものを記述する。
- (3) 動詞階層を作成する。
- (4) 3を基に2を修正し、3~4を繰り返す。
- (5) 各見出し語に制約条件を記述する。
- (6) 制約条件を含めた動詞階層を作成する。
- (7) 6を基に5を修正し、6~7を繰り返す。

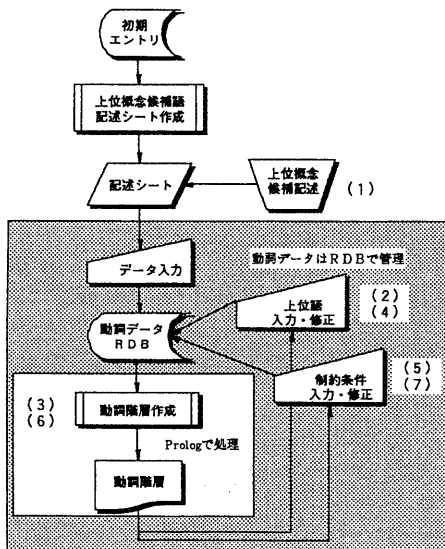


図3 動詞階層作成フロー

動詞の各エントリーに記述される情報は、動詞データとしてデータベースで一元的に管理した。また、これを基に動詞階層としての木構造の出力はProlog上の処理で行った。動詞データベースの項目一覧を図4に示す。

項目	内容
ID	初期データの連番、≤1436
見出し語	動詞表層形
読み	見出し語の読み
分類語彙表番号	「分類語彙表」の該当番号
上位語	見出し語の上位語
上位概念候補語	見出し語の上位概念の候補語
(下位概念)	(使用せず)
メモ	上位語や制約条件記述の際に気づいたこと
制約条件1~4	制約条件の記述
岩国品詞1~2	「岩波国語辞典」での品詞の記述(自他)
岩国意味	「岩波国語辞典」でのサ変の範囲内の意味記述
語構成	漢語の語構成情報

図4 動詞データベース項目一覧

上位語の選定及び制約条件の記述(2)(5)では、記述者の言語直感だけに頼るのは問題がある。そこで、国語辞典^{[5][6][7]}の記述を参考にすることにした(より一般的な動詞の意味を記述するために、この段階では新聞記事のKWICは使用しなかった)。図4では岩国品詞1~2・岩国意味の項目が相当する。また、(4)(7)で、動詞階層を評価する客観的な基準が必要となる。このための基礎資料として分類語彙表^[11]の番号を記述した(ただし、上位語選定の判断には利用しない)。

動詞データベースに入力され、管理された情報は必要に応じて、Prologの述語の形式に変換され、動詞階層作成(3)(6)のために利用される。Prologでの表現は図5に示すようにアリティ6の述語である。動詞階層の作成では、初期エントリーの上位語を参照し、これが初期エントリーの中に含まれて

いれば、そのエントリの下位語のリストを埋め、無ければその語をエントリとして追加するという処理を行い、(部分)動詞階層を(部分)木構造として出力する。

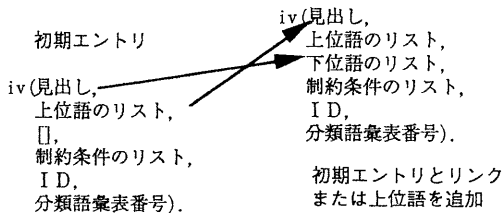


図5 Prologでの表現

5 成果

動詞階層の作成において次の結果を得た。初期エントリの総てに上位語を記述し、初期エントリの約20%に制約条件を記述した。

初期エントリ	1436語
追加上位語	500語
制約条件の記述	276語

全体のノード数は初期エントリと追加上位語の合計約2000である。追加された500の上位語は、現段階では部分木の最上位のノードである(上位語の上位語は追加していないため)。従って、500の部分木が存在し、一つの部分木は平均約4個のノードを持つことになる。なお、初期エントリの上位語が初期エントリということもあるため、最大の深さは最上位を1として4であった。

制約条件の記述は、今後の作業に役立てるために試験的に行ったものであるが、表層格の記述は行わなかった(付表を参照)。

6 分析

動詞階層の作成の過程で幾つかの問題点が明らか

かになった。大きく分けると、上位語選定に関わる問題点と、制約条件の記述に関わる問題点があげられる。

6-1 上位語選定の問題点

6-1-1 複合動詞

今回の作業では、国語辞典の語義記述を参考にして上位語の選定を行った。このため、国語辞典の語義に基づいて上位語を記述する傾向が現れた。ここで、問題となるのは小型の国語辞典では、漢語サ変動詞の意味がその漢語を訓読した和語系複合動詞として、あるいはその漢語を構成している漢字を用いて記述してあることが多いことである。例えば、

開放する あけはなすこと

出現する あらわれ出ること

和語系複合動詞の方が意味が広く、上位語となる場合は一応問題ないが、同義、あるいは下位語の可能性もあり^[13]、また複合動詞をどう扱うかという方針を決めていなかったため、今回の作業ではなるべく複合動詞は上位語としないことにした。

6-1-2 語構成

また、上の問題とは一応独立であるが、二字漢語の構造(語構成、要素となる漢字の意味とその構造の種類)は、その漢語の意味・制約条件と関連している。^{[11][12][14]}例えば、次のように

客体 遷都する →上位→ 移す(都を)

接続 持参する →上位→ 持つ/行く

並立類義 移動する →上位→ 移る

並立対義 上下する →上位→ 動く

客体関係の構造を持つ漢語は前要素が上位概念と結び付き、後要素が制約概念になる。接続関係の漢語は上位概念が二つに分かれ易い。並立類義の場合は上位概念が一つに決め易いが、並立対義では、前後要素の各々が継起的に行われることを制約条件で記述し、上位語としては、前後要素のさ

らに上位の語が結び付く、などの傾向がある。

このため、動詞データベースに語構成の項目を設け、一部の動詞について「客体、連体、連用、接続、並立類義、並立対義、曖昧、不明」の記述を試験的に行った。その結果、語構成の判定は上位語の選定と同程度に困難であることが分かった。語構成と意味の関係は漢字形態素^[8]を見出しとする辞書が利用出来ないと上位語を判定する材料には出来ないであろう。

6-1-3 漢字表記と訓読

和語系動詞を上位語とする時に、漢字表記をどこまで認めるかという問題がある。例えば、日本語では次のように和語動詞を漢字でかき分けることにより、曖昧性を減少させることが出来る。^[9]

・みる 見る←見学する／視る←視察する
 看る←看病する／観る←観覧する
 診る←診察する

・あらわす 表す←表示する／現す←具現する
 著す←著作する／顕す←顕彰する

漢字のかき分けが「みる・あらわす」で表される概念の下位概念と考えることが出来る。また、次のように、逆に漢字の下位概念として読みがあると考えられる場合もある。

・行 行く←先行する／行う←実行する

漢字表記と訓の問題は、語構成と意味の問題と同様に日本語にとって本質的なことであるが、この問題を持ち込むことは、現段階では動詞階層を複雑化することになるので、常用漢字表音訓の範囲内で記述することにした。上記の例では「見る・表す／表す／著す・行く／行う」を上位語として記述して良いことになる。

6-1-4 多義語の問題

当初は明確に意識していなかった問題に初期エントリの多義の問題がある。サ変動詞は下層に属するためにあまり多くないと予想していたためであるが、国語辞典の語義記述に従えば凡そ10%の初期エントリが多義性を持っている。例えば、

・裁断する

(1) 布・紙などを型に合わせて断ち切ること。

→裁つ

(2) 物事の善悪・正邪を判断して断定すること。

→断定する

・並行する

(1) 並んでいくこと。 →行く

(2) 同時に行われること。 →行う

上記の語は明らかに上位語が異なり、また、制約条件も異なる。そこで、これら多義性を持つ語は別エントリとして登録することにした。現在、上記の語を含む16語が多義語として二つあるいは三つに分けて登録されている(従って、動詞階層のノードが16個増加する)。今後、さらに見直しが必要である。

6-2 制約条件記述の問題点

制約条件記述については、試験的な記述を行ったに過ぎないので、まだ分析を終えていない。ここでは、国語辞典との関連で動詞の自他の問題についてだけ触れる。

制約条件には、既に述べたように格情報を記述する。今回の作業では表層格の記述は行っていないが、例えば対象のヲ格をとる他動詞「移す」とその自動詞形「移る」の部分をも動詞階層から抜き出すと次のようになる。

▼移す (ID:10023, BN:)

→遷都する (ID:828, BN:1.1521)

:対象 都

→移送する (ID:79, BN:)

:対象 事件の処理

:始点 裁判所

:終点 裁判所

→移植する (ID:78, BN:1.1521+1.3810)

:対象 (植物 生体組織)

→移管する (ID:74, BN:1.1521)

:対象 (管理 管轄)

▼移る (ID:10024, BN:)

→推移する (ID:755, BN:1.1521)

→移動する (ID:81, BN:1.1521)

: 終点 所

→移転する (ID:80, BN:1.1521)

: 対象 (組織 建物 権利)

: 終点 (場所 住所)

→移住する (ID:76, BN:1.1521)

: 動作主 人

: 終点 国

→移行する (ID:75, BN:1.1521)

: 対象 状態

日本語の動詞の自他については種々の議論^[15]がなされている。上に示した階層では、「移す」の下位にまとめられている語は、いずれも対象のラ格をとる（「移る」の下位にある「移転」の対象はガ格）。しかし、「遷都する」での対象の役割は特殊である。「遷都する」では、対象が既に語の中に含まれているため、「首都を遷都する」とは言えない。国語辞典の品詞も自動詞である。これは、対象の値が十分に語の中で特殊化されているために対象を言語表現として必要としないためと考えられる。ここではこれを対象の特殊化された他動詞と認めることにする（これと似た例に「出火する＝ガ格を取らない」がある）。

また、サ変動詞には自他両用に用いられる語が多く存在する。上記の例でも「移動する」「移転する」は自他両用である。また、次のような例をどう考えるかという問題もある。

心臓が機能_を停止した。(対象の属性か?)

さらに、サ変動詞は使用分野が限定される傾向があるので、これを多義性と関連させて制約条件で扱うことも考えられる。

7 まとめ

動詞概念辞書開発のための動詞階層作成の問題点を述べた。現在、動詞階層の部分木をまとめる

作業に入っており、500の部分木が半数程度にまとまりつつある。今回、十分に分析出来なかった制約条件の記述については今後の課題である。また、これまでボトムアップに進めて来た作業を、どこかで最上位のトップダウンから見直す必要がある。これらの問題を解決して動詞概念辞書が完成した際には、データを公開する予定である。

なお、本稿で扱った初期データの記述に千田恭子氏（東京女子大、現、電力中央研究所）の協力を得た。

参考文献

- [1] 国立国語研究所編 『分類語彙表』(64 秀英出版)
 - [2] 大野・浜西 『角川 類語新辞典』(81 角川書店)
 - [3] 岡田直之 『語の概念の表現と蓄積』電子情報通信学会(91 コロナ社)
 - [4] 日本電子化辞書研究所 TR-041 EDR電子化辞書仕様説明書(93)
 - [5] 西尾・岩淵・水谷編『岩波国語辞典』第4版(86 岩波書店)
 - [6] 松村 明編 『大辞林』(88 三省堂)
 - [7] 森岡他編 『集英社国語辞典』第1版(93 集英社)
 - [8] 森岡健二 「文字形態素論」国語と国文学45・2(68)
 - [9] 森岡健二 「日本語と漢字」ソフィア35・4(87)
- 上記2論文は、森岡健二 『現代語研究シリーズ2 文字の機能』(87 明治書院)に所収
- [10] 森岡健二 『現代語研究シリーズ1 語彙の形成』(87 明治書院)
 - [11] 日向敏彦 「動詞性漢字の造語機能」上智大学国文学論集15(81)
 - [12] 日向敏彦 「漢語サ変動詞の構造」上智大学国文学論集19(85)
 - [13] 石井正彦 「漢語サ変動詞と複合動詞」日本語学6・2(87)
 - [14] 服部 隆 「漢字の表意性から見た2字漢語」福岡女学院大学紀要1(91)
 - [15] 鳥飼浩二 「自他動詞の認定をめぐる序論」月刊言語22・5(93 大修館)

付表.

制約条件として記述した内容の一覧を以下に示す。制約条件は「属性 値」の対として記述するが、以下の表では属性ごとにそこに現れた値（多くは国語辞典の語義記述に従って名詞あるいは連用修飾句そのものを記述している）の異なりを示した。

動作主 児童、生徒、生物、組織、代表、使節、主権者、取引所、人、人間、代表者、複数

対象 行為、規則、方式、資格、権利、集会、会議、編成、編集、裁判、事業、制度、業、運転、運航、営業、業務、映画、演劇、会、学校、概観、設備、装備、包装、管理、管轄、間柄、情勢、関係、取り決め、組織、願ひ、禁止行為、記録、契約、許可、金銭、物品、計画、予算、建築物、状態、建物、船、言葉、概念、論理、戸、窓、門、制限、誤り、不良箇所、考え、主張、行事、行動、範囲、集団、講義、講習、高貴な場所、特定の分野、作用、働き、研究、事件、任務、材料 ………

相手 人、貴人

内容 (対象の内容)

有り難い、競走相手、正当な権利、無い、高い、義務、承認、非承認、売買対象

始点 悪い (状態)、裁判所、危険な状態、学校、外国

終点 書物、書類、場所、程度、舞台、場面、強い (状態)、公開帳簿、所、所々、状態、全体の計算、帳簿、半分、人体、悪い (状態)、元、興奮状態、住所、国、公共団体、軟らかい、乗物、機械、誤解、非難、不安定、裁判所、激しい (状態)、良い (状態)、他国民、自生、繁殖、港、安全な状態、新聞、雑誌、内地、基地、郷里、自国、故郷、自宅、自分

手段 権限、意志、命令、力づく、権力、つなく、規律、計算、条件

道具 火薬

目的 祝う、祈る、達成、営利

程度 大

時 現在、後で

所 会議、日本、複数、舞台、さお、塔、
法廷

数量 1

様態 正式に、離れて、一緒に、前もって、周期的、強く、徐々に、専門的、客観的、統一的、親身に、静かに、声をそろえて (?)、共に、並べて、一つ一つ、活発に、互いに、きっぱりと、じっと、はっきり、まともに、円形に、確かに、飲んで、急激に、具体的に、繰り返す、激しく、公に、再度、思い通り、試みに、次第に、実際に、実地に、初めて、新たに、深く、生き生きと、先、続けて、第三者的、直接、夢中、明確に、目ざましく、連続的に

位相 漢語 (系)、文章 (語)