

格パターン分析に基づく動詞の語彙知識獲得

大石 亨[†] 松本 裕 治[†]

動詞がある事象を表すために、論理的に最低限必要な名詞との関係を「格」という。格構造 (case frame) は、自然言語処理をはじめとする人工知能分野において、文の意味を表示するために、必要不可欠なものとして取り扱われてきた。特に、意味主導型の言語である日本語文の解析では、動詞とそのとりうる格との意味的關係を表示する深層格 (deep case) が重要な役割を果たしている。しかし、格関係の分析は意味的な問題であるだけに、どうしても個別のかつ主観的にならざるをえない。本論文では、文の表層に現れる格助詞およびそれと置換される語句のパターンに基づいて動詞を細かく分類し、この分類に基づいて、動詞の語彙知識を獲得する手法 (格パターン分析法) を提案する。この手法を用いることにより、意味的な情報を客観的にしかも類型化して取り扱うことができる。この手法を解析済みコーパスから得られた共起情報に適用して行った深層格獲得実験の結果と評価、ならびに、実験を通して得られた格パターンの組合せから動詞の意味構造 (semantic structure) を抽出する方法について述べる。

Lexical Knowledge Acquisition for Japanese Verbs Based on Surface Case Pattern Analysis

AKIRA OISHI[†] and YUJI MATSUMOTO[†]

In artificial intelligence, including natural language processing (NLP), case frames of verbs have been regarded as essential in expressing the meaning of sentences. Particularly, in the analysis of sentences of semantic-oriented languages like Japanese, deep cases expressing semantic relations of verbs and their possible arguments play an important role. However, treatment of semantic problems tends to be subjective and idiosyncratic. In this paper, we describe a method for lexical knowledge acquisition for Japanese verbs. The verbs are divided into small classes according to surface case patterns which consist of postpositional particles together with possible substitutions by other postpositional particles. This method makes it possible to deal with semantic information objectively on the basis of general principles. After proposing the Surface Case Pattern Analysis of Japanese verbs, we present the result of experiments for acquisition of deep case frames of verbs and evaluate this method. Then, by examining the data acquired through the experiments, we explain a way to obtain part of semantic structures of verbs.

1. はじめに

機械翻訳や文章理解などの自然言語処理システムにおいて、辞書をはじめとする知識ベースが重要な役割を果たしていることは言うまでもない。辞書中にどれだけの知識が含まれているかがシステムの枠を規定するのであり、意味処理、知識処理の導入を含む処理の高度化は、必然的に豊富な辞書記述を要求することになる。システムの持つ辞書の量と質が、そのシステムの性能と有効性を決定するといっても過言ではない。このために、人手による電子化辞書の開発がいくつか

行われている^{9),10),15),16)}。しかし、人手による辞書の作成には膨大な労力とコストが必要であるうえに、人間の主観的な判断の影響を受けやすく、記述の一貫性を保つのが困難であるなど、多くの問題がある。さらに、システムの新たな分野への適用や改良は、単語の新規登録や辞書記述の見直しを必要とすることから、辞書作成のコストや管理の容易さは実用上非常に大きな問題となる。これらの問題を解決するためには、辞書の記述基準を明確にするとともに、大規模辞書を作成、管理するための効果的なソフトウェアを開発する必要がある。特に、通常のデータベースとは異なり、辞書の場合には、その作成時に最大の難関がある。辞書作成時の困難さは、記述基準の設定の困難さにある。したがって、なんらかの客観的基準に基づいて、辞書

[†] 奈良先端科学技術大学院大学情報科学研究科
Graduate School of Information Science, Advanced Institute of Science and Technology, Nara

に記述されるべき情報を自動的に獲得する研究が重要になってくる。

本稿は、辞書に記載されるべき意味情報を客観的な基準で獲得する技法を提案するものである。次章では、本研究の基礎となる分析手法である「格パターン分析」について説明する。この手法は、個別的な語彙の意味を類型化して取り扱うためのものである。第3章では、第2章で述べた手法を解析済みコーパスから得られた名詞と動詞の共起情報に適用して行った動詞の深層格構造の獲得実験の内容と結果および評価を述べる。実験の結果得られた深層格は、あらかじめ人手によって辞書中に記載されている情報と照合することによって、その精度を評価している。第4章では、実験を通して得られたデータに基づいて動詞を分類し、その一つ一つに対してその分類に含まれる動詞が持つ意味構造 (semantic structure) を考える。現在の言語研究においては、深層格は動詞がとる名詞句に与えられる単なるラベルではなく、動詞の意味構造を形成する各要素の構造的な位置が持つ属性の一般化であると考えられている³⁾。さらに、一つの動詞が複数の格構造を持つ場合に、その組合せを考えることにより、より詳細な意味構造を獲得する手法 (syntactic bootstrapping) についても述べる。最後に、本論文で提案した方法に対する考察を簡単にまとめ、今後の課題などを明らかにする。

2. 格パターン分析

2.1 深層格

動詞がある事象を表すために、「何が何をどうする」というような名詞の参画が必要である。このように、動詞の語彙的な意味を実現するために論理的に最低限必要な名詞との関係を「格」という。

機械翻訳、自然言語理解をはじめとする言語情報処理において、格構造 (case frame) は、文の意味を表示するために必要不可欠なものとして取り扱われてきた。特に、意味主導型の言語である日本語文の解析では、動詞とそのとりうる格との意味的關係を表示する深層格 (deep case) が重要な役割を果たしている¹³⁾。

一方で、格の概念はあいまいなものであり、どれだけの異なる深層格を用意し、どのような基準で認定するかについての一致した見解はないこと¹⁴⁾、外部世界の事実関係を反映する真理条件的な関係にもとづく深層格と、認知のプロセスをダイナミックに反映する複合的視点による統合的な格の意味規定である認知格を区別する必要があること¹⁷⁾、などが指摘されている。すなわち、同じ格の関係と考えられる場合でも、視点

や文脈などによって格の解釈が一律に決められないこと、解釈する主体の判断がほぼ一致するプロトタイプの格の解釈と、視点や文脈によって判断がゆれる周辺的な格の解釈がみられることなどである。

以上のような問題意識のなかで、ここでは、カテゴリカルなアプローチを考える。カテゴリカルなアプローチとは、個々の動詞の個々の格のとりかたを見ていくのではなく、全体としてどんな格のパターンがあるかを事前に分類しておいて、動詞句としての意味がどれに属するかを考えていくやり方である。このアプローチを格パターン分析^{*}と呼ぶ。この方法を利用することによって、無数の組み合わせになる動詞と格との関係を整理して扱うことが可能になる。総合的な語彙的な意味のあり方が類型として整理されるからである。

2.2 格の類型

近年、動詞の構文的な振舞いは、その動詞の意味によってある程度まで予測されること、すなわち、ある動詞がどのような項構造 (argument structure, 以後、「格パターン」と呼ぶ) を持つかは、その動詞の語彙的な意味と密接な関連があるということが指摘されている^{1),5)}。また、同じ動詞が異なる格パターンを持つときに生じる意味的な差異についても考察がなされている^{6),8)}。たとえば、

- (1a) ペンキを壁に塗る
- (1b) 壁をペンキで塗る
- (2a) リボンを部屋に飾る
- (2b) 部屋をリボンで飾る

のような言い換えが行われた場合、(1a), (2a) のタイプでは、あくまで「ペンキ」、「リボン」が取り上げられているのであって対象の接着の意味になっているが、(1b), (2b) では、「壁」「部屋」が変化の対象として取り上げられている。後者において、「壁全体が塗られる」あるいは「部屋全体が飾られる」というようなニュアンス (holistic effect) が出るのは、この違いによるものである。「壁の変化」「部屋の変化」を取り上げる構造なので、少なくとも変化になるだけの「ペンキ」「リボン」が必要となる。このような現象は、「塗る」「飾る」という語彙的な意味とは別のレベルで、動詞句全体としての意味があり、そのような格パターン自体の持つ意味を考える必要を示唆するものである。

このような格のパターンそれぞれが持つ意味をあらかじめ設定しておくことによって、同一の動詞でも格

^{*} この分析法は、連語論的分析とも呼ばれ、森山 1988¹⁸⁾ にその多くを拠っている。

の取り方によってでてくる意味の違いが明らかになるばかりでなく、同じ格関係を持つ動詞の意味の共通性も明らかにすることができる。ただ、カテゴリカルな意味の類型だけでは、意味的な問題でもあり、客観的に意味をどう整理していくかが重要な課題となる。

2.3 格助詞置換分析

格関係を中心的に表すのは格助詞である。しかし、格助詞は必ずしも固定的ではない。同じ動詞が同じ名詞をとり、格助詞だけが異なる場合がある。これらの格助詞がどのように置換されるかということが、格関係を意味的に探る際の大きな手がかりとなる。

たとえば、格助詞の「に」は、次のような格助詞に置換される。

奈良 に/で 泊まる
 奈良 に/へ 着く
 彼 に/に対して 抵抗する
 彼 に/に向かって 投げる
 失敗 に/によって 悩む
 彼 に/と 会う

山梨 1994¹⁷⁾に述べられているように、これらのパラフレーズの間には認知的な意味の違いがあるが、その違いは真理条件的な同値関係に及ぶものではない。すなわち、格助詞が変わることで、もちろん、何らかの意味の変化は生じているが、その意味の変化は、動詞句のカテゴリカルな意味に及ぶものではなく、あくまで格助詞の意味の変化に留まるということである。意味的な問題の扱いは、どうしても主観的になりがちであるが、このような(複合格助詞的なものも含めて)格助詞の置換を利用することによって、格のパターンの分析を客観的なものにすることができる。この方法を格助詞置換分析と呼ぶ。

以上のような考察に基づいて、格のパターンを整理したものが表 1 である。

3. 深層格獲得実験

3.1 深層格の認定および利用資源

前章で述べたように、どれだけの異なる深層格を用意し、どのような基準で認定するかについての一致した見解はない。本研究では、表層格フレームを抽出する資料として、「EDR 日本語共起辞書」¹⁰⁾を使用した。そこで用いられている「概念関係子」を深層格として採用した。実験の結果を評価するためである。概念関係子は、表層において係り受け関係にある二単語のそれぞれが対応する概念の間で成立する意味的な関係を示すものである。EDR 電子化辞書では 24 種類の概念関係子が設定されているが、そのうち、動詞と

名詞の関係を表すと考えられる以下の 11 種類の概念関係子を対象として実験を行った*。

agent	動作を引き起こす主体
object	動作・変化の影響を受ける対象
cause	原因
material	材料または構成要素
source	起点
goal	終点
place	事象の成立する場所
purpose	目的
basis	比較の基準
beneficiary	受益者
quantity	動作・変化の量

「EDR 共起辞書」は、EDR コーパスから得られた語と語の間の共起情報が記述されたものである。共起情報とは、ある単語が、どのような関係で、どういう単語と同一の文中で同時に用いることができるのかという共起関係に関する情報である。EDR 共起辞書では、二つの単語に関する共起関係が、以下に示すような関係子を介する関係として表現されている。

[受け側単語-共起関係子-係り側単語]

たとえば、「先生が辞書を引く」という文がコーパスに現れた場合、

[引く-が-先生] [引く-を-辞書]

の二つが、共起関係として表現される。EDR 共起辞書では、このような表層における共起関係に加え、意味的情報、文法的情報を付加したものが一つの単位(共起項目と呼ばれる)として記述されている。我々は、約 116 万件にのぼる共起項目から、受け側単語品詞が動詞で、係り側単語品詞が名詞であるデータを抽出した。なお、この段階で、例文中に「れる」、「せる」、「てもらう」など格関係に変更をもたらず助動詞を含んでいるもの、処理対象以外の任意的な概念関係子を含むもの、「出」「回」など、語幹のみでは自動詞と他動詞の区別がつかないもの**を除外した。

次に、各共起項目が EDR コーパスのどの文から採られたかという情報を格納してある共起辞書コーパス対応データと、各共起項目に付与されているコーパス文番号をキーとして、同一のコーパス文番号を持つデータをまとめることにより、236,157 件の表層の格

* 任意的な格関係を表すと考えられる scene, manner, time, time-from, time-to などを除外している。

** 各単語には、概念識別子が付されており、これを用いればそれぞれを区別することは可能であるが、本研究では、できるだけ表層的な情報から深層的な情報を得ることを目的としているので、概念識別子は使用しなかった。

表1 格パターンによる動詞分類

Table 1 Verbal categories.

型	格の交替	他の情報	深層格	意味の分類	動詞の例
無格型	[ガ] 格さえとらないことがある	場所 上位概念	場所格 対象格	自然現象 自然現象	停電する 春めく
[ガ]型	自他対応なし 自他対応あり	生物 場所	主体 場所格 対象格 対象格	単独動作 単独自然現象 単独自然現象 変化	生きる 踊る 悩む 晴れる 暮れる 生まれる ほける 壊れる 膨れる
[ガ, ニ]型	ニ格の交替なし [ニ/ア (ニオイテ)] [ニ/ヘ] [ニ/ニタイシテ] [ニ/ニヨッテ/ア] [ニ/ト] [ニ/ト]	場所 場所以外 同一 相互	主体, 場所格 主体, 対象格 主体, 場所格 主体, 終点格 主体, 対象格 主体, 原因格 主体, 結果格 主体, 相互格	存在 関係 存在様態 接着 態度, 働きかけ 原因・変化 主体改変型変化 対等な関係	ある いる 存在する 関与する 因る 要る 浮かぶ 沈む あたる 埋まる 憧れる 干渉する 驚く 呆れる 悩む なる 改まる 変わる 会う 関わる 別れる
[ガ, カラ]型	[カラ/ヲ] カラ格の交替なし		主体, 起点格 主体, 起点格 主体, 起点格 主体, 起点格	出発 出現 消滅 成立, 生成	出る 離れる 去る 生まれる 発生する なくなる 消える できる 成り立つ
[ガ, ト]型	[ト/ニ] ト格の交替なし	既出 相互 引用 引用	主体, 相互格 主体, 引用 主体, 引用 対象, 引用	相互動作 非伝達の発話 思考・感情 自発	結婚する 喧嘩する わめく 叫ぶ 感じる 思う 聞こえる 見える
[ガ, カラ, ニ]型	[ニ/ヘ/ヘト] [ニ/ト]	同一	主体, 起点, 終点 主体, 起点, 終点	出発 単純移動 移動形態 主体改変	離れる 出発する 上がる 移動する 歩く 急ぐ 泳ぐ 改まる 変わる なる
[ガ, ト, ニ]型	[ニ/ニタイシテ] [ニ/カラ]	引用 引用	主体, 引用, 終点 主体, 引用, 起点	伝達 伝聞	言う 怒る 話す 聞く 承る 教わる
[ガ, ヲ, ト]型		引用 相互	主体, 対象, 引用 主体, 対象, 相互	判断 対象交換 対象競争	思う 考える 感じる 交換する 取り引きする 争う 競う
[ガ, ヲ, ニ]型	[ニ/デ (ニオイテ)] [ニ/ヘ] [ニ/ト] [ニ/ニタイシテ] [ニ/ト]	相互 同一	主体, 対象, 場所 主体, 対象, 終点 主体, 対象, 相互 主体, 対象, 終点 主体, 対象, 結果	発見 対象接着 対象姿勢方向 相互型対象変化 適応 対象改変	見る 発見する 浴びせる 当てる 倒す 向ける 合わせる 重ねる 合わす 試みる 変える 直す
[ガ, ヲ, カラ]型			主体, 対象, 起点	対象出現 対象消滅 対象移動 対象影響	出す 生む 消す なくす 引用する 買う 分ける 受ける
[ガ, ヲ, カラ, ニ]型	[ニ/ヘ] [ニ/ト]	同一	主体, 対象, 起点, 終点	対象移動 対象改変	出す 遠ざける 改める 変える
[ガ, ヲ]型	ヲ格の交替なし [ヲ/(ヲトオッテ)] [ヲ/カラ] [ヲ/ニタイシテ] [ヲ/ニツイテ] [ヲ/ニ] [ヲ]	場所 同族目的 量	主体, 対象 主体, 経路 主体, 起点 主体, 対象 主体, 対象 主体, 対象 主体, 対象 主体, 量	対象消滅 対象変質 対象生産 感情 経路 起点の経路 働きかけ 知的動作 ? 単独動作 単独動作	砕く 殺す 焦がす 乾かす 書く 築く 愛する 諦める 越える 過ぎる 降りる 出る 攻撃する 怒る 考える 検討する ご馳走する 落ちる 生きる 歌う 走る 黙る

ただ、「が」「を」「に」のような基本的な格については、その取り得る深層格の範囲が非常に多岐にわたるため、名詞と動詞の意味マーカの組合せをあらかじめすべて指定しておくことは困難である。そこで、**STEP-1**、**STEP-2**で作成したファイルを利用することによって、深層格を決定した。以下に、その手順を示す。

A. 「に」「を」「と」格の処理

- 1. 動詞別格パターンファイルの参照** 当該データが「に」「を」または「と」格を含む場合、動詞をキーとして動詞別格パターンファイルを参照し、その動詞が「に」「を」または「と」格の置き換わったパターンを含むかどうかチェックする。
- 2. 格助詞置換参照処理** その動詞が「に」「を」または「と」格の置き換わったパターンを含むとき、動詞・格助詞パターン・置換された格助詞をキーとして、格助詞置換参照用ファイルを検索し、処理対象データともっとも近い意味マーカを含むものを選択することにより、深層格を決定する。
- 3. 意味マーカの利用** その動詞が「に」「を」または「と」格の置き換わったパターンを含まないとき、または置換された格が近接した意味マーカを持たない場合（後述）、**STEP-3**で設定した動詞の意味マーカと格助詞の組合せによって深層格を決定する。
- 4. デフォルト値** 上記の処理によって深層格が決定できない場合、あらかじめ設定してあるデフォルト値（「を」= “object”, 「に」= “goal”, 「と」= “object”）を割り当てる。

B. 「が」格の処理

- 1. 意味マーカの集計** 格助詞置換参照用ファイルから、処理対象データと同じ動詞・格助詞パターン・「が」格をキーとして持つデータを検索する。このデータの意味マーカを上位3桁ごとに集計する。
- 2. 深層格の割り当て** 集計された意味マーカのうち、もっとも多数のものにより深層格を以下のとおり割り当てる。

意味マーカの上位3桁	意味	深層格
117,125,126,144,147,152	(場所・地名など)	place
121,122,123,124,127,128,	(人間活動の主体)	agent
上記以外のすべて		object

たとえば、「警官が現場に駆けつける」というデータの「に」格の深層格を決定する場合を考える。動詞別

格パターンファイルを参照することによって、「駆けつける」という動詞は、「が、に」という助詞パターン(b)の「に」格の置き換えられたパターンとして、「が、へ」(b3)と「が、によって」(b5)の二つの可能性を持つことがわかる(表3)。

次に、それぞれのパターンごとに、格助詞置換参照用ファイルを検索する。「駆けつける」「が、へ」「へ」をキーを持つデータには、「東京(12590520)」「法務省(12710320)」「病院(12650920)」などが登録されており、「駆けつける」「が、によって」「によって」をキーを持つデータには、「通報(13123470)」「知らせ(131234130)」が登録されている。

したがって、処理対象データである「現場(12620340)」にもっとも近い意味マーカを持つ「病院(12650920)」が選択され、この「に」格が「へ」と置換され得ると判断し、深層格は“goal”であると決定される。

この処理において問題となるのは、本来置換されるべきではない任意的な格要素まで置換されることである。たとえば、「出かける」という動詞の「に」格は「外国に出かける」のように“goal”を表すものと、「買物に出かける」のように“purpose”を表すものがあるが、置換対象のパターンとして「映画館へ出かける」など、「が、へ」という例しかなかったために、すべて“goal”になってしまう。この場合の意味マーカは以下のようにになっている。

外国	12530 2 150 (国)
買物	13761 8 20 (売買)
映画館	12650 11 20 (店・旅館・病院・劇場など)

したがって、格助詞置換の範囲を、名詞句の意味マーカがシソーラスの根節点から2レベルまで一致するものに制限することによって、このような誤った置換を防止した。シソーラスの根節点の一つ下のレベルは、「抽象的關係」、「人間活動の主体」、「人間活動—精神および行為」、「生産物および用具」、「自然物および自然現象」という五つのカテゴリーからなっており、このカテゴリーを越えて格助詞の置換が行われることは稀であると考えたからである。

一方、「が」格の処理には、個々の用例ごとの意味マーカによってではなく、その動詞があるパターンで典型的にとる名詞の意味マーカによって深層格を決定した。これは、もの名詞が擬人的に用いられている場合や、意味マーカが得られない場合にも対処するためである。

これまでに述べた処理の概要を図1に、**STEP-4**で使用した規則を表6に示す。

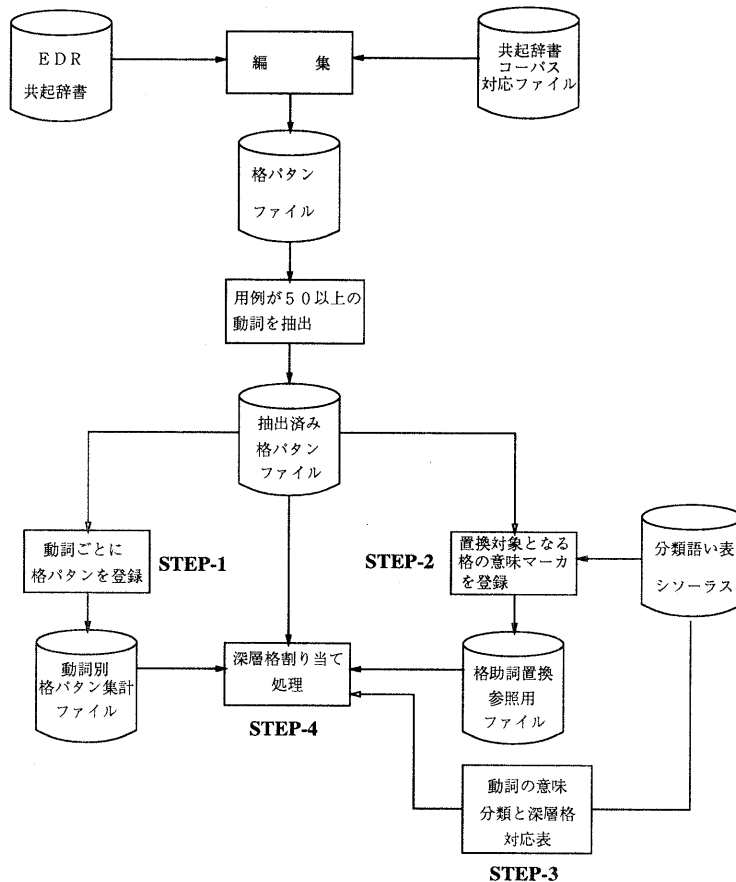


図1 処理の概要

Fig. 1 Process of experiments for deep cases acquisition.

3.3 結果および評価

実験は、最初に名詞シソーラスと動詞パターンのみを用いて行った。その結果、80.32%の項が、EDR 共起辞書に付されていた概念関係子と一致した。(総数 210,242 件、一致件数 168,859 件)

次に、この処理で不一致となったデータをもとに、先に述べた動詞の意味分類と深層格の対応表を作成し (STEP-3)、これを利用して実験を行った結果、一致率は 81.97% (一致件数 172,345 件) に向上した。

動詞シソーラスを利用しても、著しい一致率の向上が得られなかった理由は、同一の動詞の同じ用法の同じ格に対しても、人間の判断のゆれによって、異なる概念関係子が付されている例が多数あったことにある。本実験では動詞のとり格パターンごとに深層格を一意に決定してしまうので、人間にも決定し難いような場合には対処できない。これは、深層格の概念自体が明確な境界を持たないことを反映している。このように、同一の動詞の同じ用法に複数の深層格が付加されてい

る場合に、どちらかの深層格が得られれば正解とすれば、不一致件数は 2,933 件となり、一致率は 98.6% に向上する。

また、実験において、一致しなかった件数をもっとも多かったのは「が」格である (16,751 件、不一致全体の 44.2%)。本実験では、「は」「も」「まで」などの副助詞をすべて「が」格とみなして処理を行ったことが主な原因であると考えられる。これらは、本実験の結果得られる表層の格のパターンと意味マーカを利用して、どの格と対応しているかを判断すべきであるが、今回は実装できなかった。

さらに、シソーラス中に存在しない名詞が 52,913 件にのぼったことも一致率に悪影響を与えていると考えられる。このうち、不一致の件数は 12,669 件であり、全不一致件数の約 30% を占めている。これを除くと、一致率は 83.96% に向上する。意味マーカが得られない場合には、シソーラスのコードによる制限をせず、格助詞のパターンにのみ基づいて置換処理を

表6 深層格獲得に用いた規則
Table 6 Rules used in experiments for deep cases acquisition.

格	pattern	格の交替	名詞の意味	動詞の意味	格パターンファイル	深層格
ガ			場所 人間活動の主体 その他		a > k	place agent object
ヲ	k k,h k k k	カラ	場所 場所 量	移動・上がり下がり 出発・隔離 探求・応接 過不足・優劣	e,e1,(b,c)を含み, i,j,hを含まない cを20%以上持つ kが20%以下 上記以外のすべて	place source goal basis quantity object
カラ	c,e,i,j c,e,i,j		原材料	成立・生産	上記以外のすべて	material source
ニ	b,h b,h b,h b,h b b,h b,h b,h,e,j b,h b,h b,h b,e,h b,e,h b,f,h,	なし なし なし デ デ ニオイテ ニヨッテ ヘ ニタイシテ ニタイシテ ニムカッテ ト ト カラ	場所 出来事 場所 場所以外 場所 場所以外 人物	存在・建設 関係・異同 移動・使用 気分・情緒 気分・情緒 変化 変化以外 伝聞・受領	bが70%以上 e, jを含まず, kを含む 上記以外のすべて	place object purpose place cause place cause goal goal object goal goal object source goal
ト	d,g,e2 d,g d,f,g	ニ		変化・思考 異同・比較	上記以外のすべて	goal basis object
デ			場所 場所以外			place cause
ヘ	b3,e1,h3,j1					goal
ニオイテ	b2,h2					place
ニヨッテ	b5,h5					cause
ニツイテ	k1					object
ノタメ			人 出来事	授受・製造		beneficiary purpose
ヨリ			場所	移動 比較		source basis
ダカラ						cause

行ったため、特に“place”や“cause”を表す「で」格のような任意的な格によって誤った深層格を与えてしまう例が目立った。さらに、名詞・動詞ともに多義性を考慮せず、シソーラスから得られるすべてのコードを意味マーカとして用いたために、正確な置換処理ができなかったことも実験の精度を下げる一因である。名詞の多義性は、本研究の置換処理を利用することによって、動詞の多義性は、用法の違いとして現れるものであれば、格パターンをもとに判断することができる可能性もある。今後の課題である。

今回の実験では、格助詞の置き換え可能性に基づい

て分類した格パターン一つ一つの意味を考えることに重点を置いたために、格パターン間の関係を十分利用することはできなかったが、実験で得られた格パターンの組み合わせをもとにして動詞を分類してみると、表7に示すように動きの認知的なとらえかたを反映した結果が浮かび上がってくる。次章では、このデータを利用して、動詞の意味構造の一部分を付与する試みを述べる。

表7 格パターンの組み合わせによる動詞分類 (一部)
Table 7 The classification of verbs based on the combinations of case pattern (partial).

格パターン	動詞
b2 a b	寝る 暮らす 遊ぶ 勤務する あふれる 眠る
k i c a	引き出す 取り出す 検出する
k i c h	持ち出す 引く
d b2 a	会う 出会う
b	対する 挑戦する 反する 基づく
k h b a	開設する 建設する 飾る 設置する
h k b a	贈る 派遣する
b a d	ぶつかる 関係する 関連する 結び付く 似る

4. 意味構造の付与

4.1 意味構造 (Semantic Structure)

この章では、動詞の格構造は、その動詞が本来的に持つ意味構造 (semantic structure) から投射されるという仮定に基づいて、格構造に基づいて動詞の意味構造を付与する試みを述べる。ここで用いる表現形式は、Jackendoff によって提案されている語彙分解理論 (Lexical Decompositional Theory) に則ったものであり、語の意味を major ontological category と呼ばれるいくつかの基本的な要素の組み合わせによって表現するものである。たとえば、“Bill went into the house.” という文は、

$$[s[{}_{NP}Bill][{}_{VP}[{}_{V}went][{}_{PP}[{}_{P}into][{}_{NP}the\ house]]]]$$

という構文構造を持つが、これに対応する概念構造 (conceptual structure)[☆] は、

$$[Event\ GO([{}_{Thing}BILL], [{}_{Path}\ TO([{}_{Place}\ IN([{}_{Thing}HOUSE])])])]$$

となる。この中で用いられている角括弧 “[”, “]” は、概念の構成要素 (conceptual constituents) を示すものであり、それぞれの構成要素には、Event, Thing, Path などの major ontological category (“semantic part of speech” と呼ばれる) を表すラベルが付されている。

また、角括弧の中の大文字で表現されているのは、概念内容 (conceptual content) と呼ばれ、それぞれの構成要素の内容を表している。BILL と HOUSE は Thing の内容であり、名詞句に対応している。それ

以外のものは、関数を表すものである。IN は、物体 (Thing) をその物体の内部の場所 (Place) に写像する一項関数である。TO は、Thing または Place を引数にとり、それらを経路 (Path) に写像する一項関数であり、その Path が引数の Thing または Place を終点とすることを表示する。GO は、Thing と Path を引数にとる二項関数であり、第一引数の Thing が第二引数の Path を通って移動するという Event に写像する。したがって、先の表現は、「BILL が HOUSE の内部を終点とする経路を通って移動する」と読むことができる。

ここで注意すべきは、深層格や thematic role の名前で呼ばれてきたものは、この構造の中の要素が占める位置に対応しているということである。すなわち、Theme (先の実験では “object” と呼んでいる) は、関数 GO の第一引数であるし、Goal は、関数 TO の引数である。ここでは、BILL と [IN([HOUSE])] がそれぞれに対応している。このように、深層格は、それ自身独立した体系を持つのではなく、意味構造における構造的な位置であるとみなすことができる。

なお、この構造は、文を構成する語彙ごとに考えることができる。たとえば、“into” と “go” に対する lexical entries は、それぞれ、次のように与えられる。

$$\left[\begin{array}{l} \text{into} \\ P \\ [{}_{Path}\ TO([{}_{Place}\ IN([{}_{Thing}A])])] \end{array} \right]$$

$$\left[\begin{array}{l} \text{go} \\ V \\ [{}_{Event}\ GO([{}_{Thing}A], [{}_{Path}A])] \end{array} \right]$$

上から順に、音韻構造、構文構造、意味構造を表している。すなわち、“into” という語は、構文構造において前置詞 (P) であると同時に、ある物体の内部へ向かう経路を表す意味構造を持つ。また、この中で用いられている添字の A は、“linking subscript” とよばれ、表層に項として現れるべき要素を表している。

この章では、以上のような表現法を用いて、日本語の動詞の意味を表現することを考える。

4.2 Syntactic Bootstrapping

第2章では、動詞がとる表層的な格のパターンと格助詞の置換に基づいて、その格パターンをとる動詞が持つ典型的な意味を考え、動詞の中には複数の異なる格パターンを持つものがある。このような場合、その格パターンの組み合わせによって、さらに詳細な動詞の意味を予測することができる。この手法を

[☆] Jackendoff の一連の研究^{2)~4)}において、このような表現は、“conceptual structure” と呼ばれており、特に、語彙を表現するものは、“lexical conceptual structure” (lcs) と呼ばれている。本稿では、意味構造は概念構造そのものではないという考えにより、“semantic structure” という用語を用いる。

Syntactic Bootstrapping と呼ぶ*。

例えば、「言う」という動詞について言えば、「『君が好きだ』と言う」のように引用のト格を伴う補文をとることや、「早く帰るように言う」のように、引用内容を表す「ヨウニ」を伴ったり、「つまらないことを言う」のように「こと」名詞を「を」格にとることから、思考・感情などの知的動作を表すことがわかる。さらに、「彼に対して言う」のように「にたいして」や「にむかって」を伴うことから、相手に対して働きかける動作であることがわかる。また、「嬉しく思う」といえるのに対して、「*嬉しく言う」とは言えないことから、それが思考動詞ではなく、伝達動詞であることがわかる。これは、伝達動詞の場合は、引用成分の主体が引用動詞に対して独立的に関わり、「言う」人と伝達内容が無関係なのに対して、思考動詞の場合は、被引用成分つまり、思考内容の主体（主語）が引用動詞の主体（主語）と同一だと原則的に保証されているために、主体を消すことができ、副詞的な修飾への置換が可能であるからと考えられる。「私は（彼が）嬉しく思う」のように言えない。このようにして、「言う」という動詞の意味をかなりの程度まで推定することができる。ただし、本研究では、コーパスから得られる正例（positive evidence）のみを用いるので、「何々とは言えない」というような負例（negative evidence）を用いることはできない。また、先の実験では『名詞＋助詞＋動詞』の組のみを抜き出したので、補文や用言を伴うト格は含まれていない。そこで、朝日新聞社説の未解析コーパス九年分から、『用言＋「と」＋動詞』という例と『用言＋「よう」＋動詞』などの例を抽出して、分類の参考にしていく。

4.3 格パターンの組み合わせによる動詞分類

ここでは、第3章で述べた実験を通して得られた動詞別格パターンファイルを用いて動詞をいくつかのカテゴリーに分類し、それぞれのカテゴリーごとにあてはまる意味構造を考えていく。分類は、以下の手順で行った。

- (1) それぞれの動詞ごとに格のパターンを用例の多い順に並び替える。
- (2) 用例数が全体の5%以下の格パターンを削除する。
- (3) 同一の格パターンの並びを持つ動詞をまとめる。

* 発達心理学等の分野において、子供が言語を獲得する際に、構文と意味との規則的な対応を用いて、単語の意味を通してその言語の構文規則等を獲得するという仮定を“semantic bootstrapping”という⁷⁾。これに対し、構文的な情報から意味的な情報を得ることは、“syntactic bootstrapping”と呼ばれる⁸⁾。本稿では、この用語の意味を、動詞がとる複数の格パターンに基づいて、その動詞の意味を推定することに限定して用いる。

最後に、任意格による影響と考えられるものを修正し、5%以下で切り捨てたパターンでも有力な情報を提供するものや、未解析コーパスから得られた情報も考慮して最終的な分類を行った。機械的な手順だけでは、カテゴリーとしてまとめるべき動詞の境界が明示的に得られないことと、より正確な分類を行うためである。以下に、いくつかのカテゴリーの意味構造の例を示す。

単独変化動詞

これは、格のパターンとしてa（「が」型）のみを持つものである。「が」格には、変化の主体がくる。

なくなる、はっきりする、減る、減少する、上昇する、深まる、進展する、衰える、整う、切れる、絶える、増大する、薄れる、変化する、膨らむ

単純に変化することだけを表すのが、このタイプの意味である。変化の様子を起点（変化するもと）終点（変化のあと）のように表すことはない。それで、意味構造は、次のようになる。

$$[\text{Event GOIdent}([\text{Thing } \alpha], \left[\begin{array}{l} \text{FROM}([\text{StateBE}([\alpha], [\text{NOT AT}([\beta])]]) \\ \text{TO}([\text{StateBE}([\alpha], [\text{AT}([\beta])]]) \end{array} \right])] \quad (1)$$

ここで、Ident は semantic field と呼ばれる意味分野の一つであり、抽象的な移動、すなわち、状態変化を表している。FROM および TO の引数は、それぞれ変化の前後の状態を表すが、それらは語彙の中に含まれており、変化の主体である GOIdent の第一引数のみに linking subscript の A が付されている。これが、「が」格として表層に現れる。変化の前後の状態が語彙中に含まれていることは、これらの動詞の多くが形容詞から派生したものであることから推測できる。すなわち、「ない」→「なくなる」、「深い」→「深まる」、「薄い」→「薄れる」などのように、元の形容詞が変化の結果の状態を表しており、動詞はその状態に変化することを表す。そこで、この状態を β とし、変化主体である α が β でない状態から β である状態に移行することを GOIdent と [StateBE] を用いて表したものが (1) である。

また、このタイプに原因を表す「で」格がついて、「が、で」型 (b1) となり、 $\langle a, b1 \rangle$ という格パターンの並びを持つものもある。

ふえる、まとまる、悪化する、安定する、下がる、高まる、死ぬ、死亡する、成立する、増える、低下する、発足する、表面化する、崩

れる, 亡くなる, 急増する, 強まる, 固まる
「進行する」「本格化する」「増加する」のようなサ変動詞で, まれに他動詞として用いられるために, $\langle a, b1, k \rangle \langle a, k, b1 \rangle$ (k は「が, を」型) という並びを持つものも含まれよう. 意味構造は, 次のようになる.

$$\left[\begin{array}{l} \text{Event} \quad \text{GOIdent}([\text{Thing}]_A, [\text{Path} \quad \text{FROM}([\text{State}]_A) \\ \text{TO}([\text{State}]_A)]) \\ \text{[BY[Event]](A)} \end{array} \right] \quad (2)$$

BY は原因を表す一項関数であり, メインイベントを修飾する adjunct を表す. 引数のタイプは Event で, ここには, 出来事名詞句がくる. Path の中は省略している.

単独動作動詞

a (「が」型) を最も典型的にとるが, そのほかの格パターンも同時に持つ動詞は, 「が」格で動作の主体を表し, 無変化的動きを表すものが多い. これらの動詞は対応する他動詞を持たないという特徴がある. また, 主体に意志性があるものもないものがある.

$\langle a, b \rangle$ という並びを持つ動詞 (任意的な原因を表す「で」格がついて $\langle a, b, b1 \rangle$ というタイプになる場合もある) には,

のぼる, 近づく, 現れる, 出現する, 登場する, 表れる, 成長する, 達する, 落ち着く, 終わる, 進む, 動く, 落ち込む, 発達する
などがあり, これらは「に」格で終点 (goal) を表すものである. 意味構造は,

$$\left[\begin{array}{l} \text{EventGO}([\text{Thing}]_A, \\ \text{[PathTO}([\{ \text{PLACE} \\ \text{THING} \}]_A)]) \end{array} \right] \quad (3)$$

となる.

$\langle a, b, b1 \rangle$ または $\langle a, b1, b \rangle$ という並びを持つ動詞は, 「で」格で場所を表し, 行為の様態を伴う.

育つ, 泣く, 広まる, 降る, 止まる, 出回る, 誕生する, 定着する, 普及する, 光る, 生まれる, 咲く, 生きる, 浮く, 働く, 揺れる, 活躍する, 起きる, 鳴く, よみがえる, 完成する, 吹く, 漂う, 輝く

「生じる」「発生する」「笑う」のように, 自他両用の動詞や「人生を生きる」などのように同族目的語をとる場合には, $\langle a, b, b1, k \rangle \langle a, b1, k, b \rangle$ のような並びとなる. 意味構造は,

$$\left[\begin{array}{l} \text{Event} \quad \text{MOVE}([\text{Thing}]_A) \\ \text{[AT[Place]](A)} \end{array} \right] \quad (4)$$

となる. MOVE は, Path を伴わないイベントを表す一項関数であり, AT は, 行為が行われる場所を表す adjunct を表す. 引数のタイプは Place で, ここには, 場所名詞句がくる. 「に」格でマークされる場合には, 終点とも解釈される場合があり, この場合には, (3) と同じ意味構造とも考えられる. このように, 移動と存在は, 連続したカテゴリーであり, 「に」格は動きに焦点があれば終点を, 結果の状態に焦点があれば存在場所を表すことになる.

このほかに単独動作を表すものとしては, 移動を含む動詞がある. これらは, a, b のほかに, b3 (「が, へ」型) や c (「が, から」型), e (「が, から, に」型) を含むことによって他のタイプと区別することができる. また, このタイプの動詞がパターン k (「が, を」型) を含む時, その「を」格は経路を表す.

動き出す, くる, 上がる, 来る, 落ちる, あがる, 届く, 移動する, 飛ぶ, 流れる, 飛び出す, スタートする, 逃げる, 帰国する, 消える, 浮上する, 来日する, 抜ける, 浮かび上がる, 延びる

などがある. 起点, 終点, 経路のどれに重点があるかによってパターンの並びは変わってくる. 先に取り上げた $\langle a, b \rangle$ 型の動詞は, 終点に焦点のあるものであり, この中に含まれる. 意味構造は,

$$\left[\begin{array}{l} \text{EventGO}([\text{Thing}]_A, \\ \text{[Path} \left\{ \begin{array}{l} \text{FROM} \\ \text{TO} \\ \text{TOWARD} \\ \text{VIA} \end{array} \right\} ([\{ \text{THING} \\ \text{PLACE} \}]_A)]) \end{array} \right] \quad (5)$$

となる. FROM, TO, TOWARD, VIA は, それぞれ起点, 終点, 方向, 経路を表し, 「に」「から」「へ」「を」に対応する.

「延びる」は, 「ている」がついて状態を表す場合には,

$$\left[\text{StateEXT}([\text{Thing}]_A, [\text{Path}]_A) \right] \quad (6)$$

と表すこともできる. EXT は extention の略で, 移動経路の存在および延長を表すものであり, Jackendoff1990³⁾ で用いられている.

存在動詞

$\langle a, b \rangle$ または $\langle b, a \rangle$ という並びを持つ動詞の中には, 存在を表すものがある. この場合の「に」格は場所を表す.

おさまる, そろう, たまる, 広がる, 収まる, 存在する, 定まる, こもる, とどまる, 位置する, 陥る, 響く, 勤める, 傾く, 建つ, 座

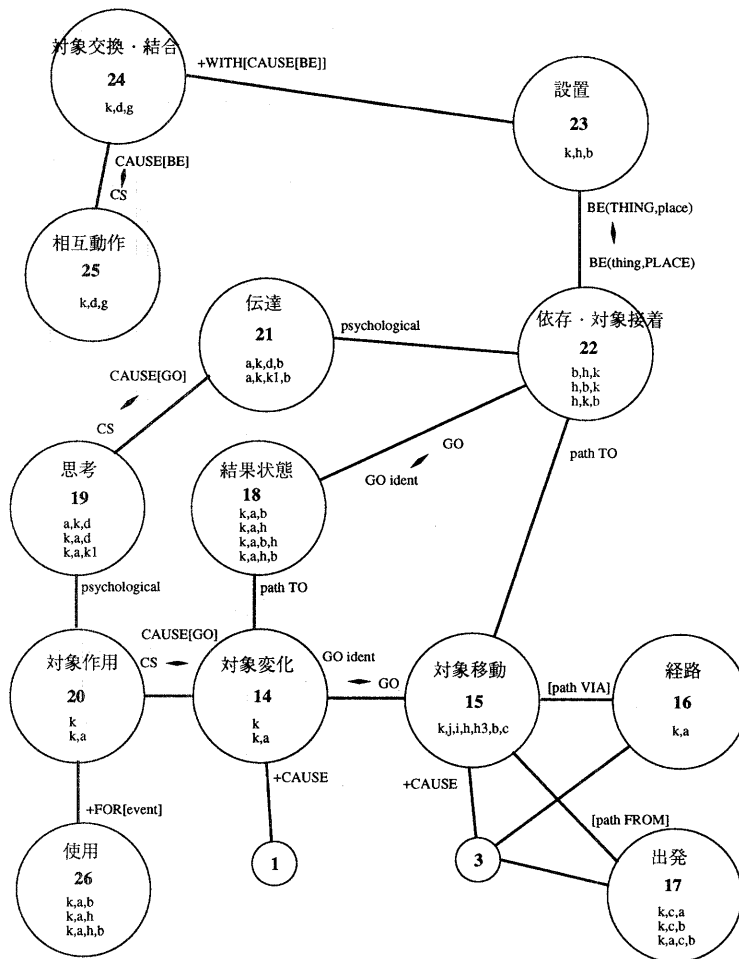


図3 意味構造の関連図(他動詞)

Fig. 3 Network of semantic structures (transitive verbs).

らに BY[event] に焦点が移れば、原因変化 (10) の動詞になるといった関係を表したものである。ここでは、わかり易さのために円で区切って線でつないだが、もちろん、各カテゴリーに明確な境界はなく、関連するカテゴリーは相互に連続したものである。

5. おわりに

本稿では、語彙の意味の分析に格成分を含めた動詞句全体の意味を利用する格パターン分析、および、意味的な分析を客観的に行うための格助詞置換分析について述べ、この手法を解析済みコーパスから得られた共起情報に適用して行った深層格獲得実験の内容と結果および評価を述べた。実験では、8割以上の格に対して、人間の与えたものと同じ深層格が得られ、本手法が格関係分析に有効であることを示している。さらに、実験を通して得られた格パターンの組み合わせに

よって動詞をいくつかのカテゴリーに分類し、そのカテゴリーごとに意味構造を考えた。詳細を述べることはできなかったが、対象とした858個の動詞のうち、約80%以上の動詞を網羅している。ただし、このうちの約200個は <k, a> という並びを持ち、対象作用動詞に含まれるものであり、その意味は非常に多岐にわたっている。これらは格のパターンのみでは判断できないものであり、本手法の限界を示している。また、分類できなかったもののうち、約100個は、場所や原因・目的を表す任意的な「に」格や「で」格などがつくことによって、本来あるべき位置とは別のところにまぎれこんだものである。さらに、動詞の多義性によって、複数の格パターンの並びが組み合わせられると思われるものもある。「行う」「する」「なる」のような機能動詞や、自他両方に用いられる漢語サ変動詞などである。今後、さらに詳しい分析をしていく必要が

あるが、統計的手法等による形態素解析および構文解析技術の向上にともない、タグ付きのコーパスが普及しつつある現状からみて、利用できる資源はますます増加すると考えられる。これらの資源によって得られたデータを用いて、完全ではなくともある程度自動的な分類ができていれば、その後の人手による分析が非常に楽になるであろう。

謝辞 EDR 電子化辞書データの使用を許可して下さった株式会社日本電子化辞書研究所ならびに著書をお貸しいただいた京都教育大学の森山卓郎氏に感謝いたします。

参考文献

- 1) Grimshaw, J.: *Argument Structure*, Linguistic Inquiry Monograph, Vol.18, MIT Press (1990).
- 2) Jackendoff, R.: *Semantic and Cognition*, MIT Press (1983).
- 3) Jackendoff, R.: *Semantic Structure*, MIT Press (1990).
- 4) Jackendoff, R.: Parts and boundaries, *Lexical and Conceptual Semantics* (Levin, B. and Pinker, S.(eds.)), Blackwell, chapter 2, pp.9-45 (1992).
- 5) Levin, B.: *English Verb Classes and Alternations*, The University of Chicago Press (1993).
- 6) Levin, B. and Rappaport, H. M.: Wiping the Slate Clean: A lexical Semantic Exploration, *Lexical and Conceptual Semantics* (Levin, B. and Pinker, S.(eds.)), Blackwell, chapter 5, pp.123-151 (1992).
- 7) Pinker, S.: *Language Learnability and Language Development*, Harvard University Press (1984).
- 8) Pinker, S.: *Learnability and Cognition: The Acquisition of Argument Structure*, MIT Press (1989).
- 9) Proctor, P.(ed.): *Longman Dictionary of Contemporary English*, Longman (1978).
- 10) (株)日本電子化辞書研究所: EDR 電子化辞書, 日本語共起辞書評価版(第2.1版)(1994).
- 11) 大石 亨: 格パターン分析に基づく動詞の語彙知識獲得に関する研究 (NAIST-IS-MT351202), 修士論文, 奈良先端科学技術大学院大学情報科学研

究科 (1995).

- 12) 国立国語研究所: 分類語彙表, 秀英出版 (1964, 1993).
- 13) 辻井潤一: 辞書の構成と機械翻訳, 情報処理, Vol.26, No.10, pp.1174-1183 (1985).
- 14) 辻井潤一, 山梨正明: 格とその認定基準, 情報処理学会自然言語処理研究会, Vol.52, No.3 (1985).
- 15) 情報処理振興事業協会: 計算機用日本語基本動詞辞書 IPAL 辞書編 (1987).
- 16) 情報処理振興事業協会: 計算機用日本語基本形容詞辞書 IPAL 辞書編 (1990).
- 17) 山梨正明: 日常言語の認知格モデル, 言語, Vol.23, No.1-12 (1994).
- 18) 森山卓郎: 日本語動詞述語文の研究, 明治書院 (1988).

(平成7年4月18日受付)

(平成7年9月6日採録)



大石 亨 (学生会員)

昭和37年生。昭和59年大阪大学文学部文学科卒業。同年奈良県庁入社。平成7年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了。現在、同大学院博士後期課程在学中。自然言語処理の研究に従事。言語処理学会会員。



松本 裕治 (正会員)

昭和30年生。昭和52年京都大学工学部情報工学科卒業。昭和54年同大学院工学研究科修士課程情報工学専攻修了。同年電子技術総合研究所入所。昭和59~60年英国インペリアルカレッジ客員研究員。昭和60~62年(財)新世代コンピュータ技術開発機構に出向。京都大学助教授を経て、平成5年より奈良先端科学技術大学院大学教授、現在に至る。専門は自然言語処理。人工知能学会, 日本ソフトウェア科学会, 言語処理学会, AAAI, ACL, ACM 各会員。