

## 格パターン分析に基づく日本語動詞の語彙知識獲得

大石 亨 松本 裕治

奈良先端科学技術大学院大学 情報科学研究科

### 1 はじめに

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 (syntactic argument structure) を持つかは、その動詞の語彙的な意味 (lexical semantic structure) と密接な関連があるということが指摘されている [1, 4]。言い替えば、同じ格構造を持つ動詞には意味的な共通性があり、逆に、同一の動詞でも格のとり方によってその表す意味に違いが出てくるということである。我々は、かつて、この格のパターンによる意味をあらかじめ定義しておくことによって、表層的な情報から動詞の深層格を獲得する試みを述べたことがある [3]。

本稿では、この深層格獲得実験を通して得られたデータに基づいて、動詞の意味構造 (semantic structure) を抽出し、その構造間の関係を明らかにする。

### 2 意味構造 (semantic structure)

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を持つ属性の一般化であると考えられている [2]。したがって、それぞれの深層格を動詞の意味構造に対応させることにより、深層格構造を動詞の意味構造に還元することができる。さらに、一つの動詞が複数の格構造を持つ場合には、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ことができる。

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」名詞をヲ格にとることから、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「ニタイシテ」や「ニムカッテ」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、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」のようには言えない。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。

以下では、動詞の持つ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当てはまる意味構造を考える。

### 3 格パターンの組み合わせによる動詞分類

我々は、EDR 共起辞書 [5] を利用して、次のような構造を持つデータを 158,888 件蓄積している。

|          |   |    |                               |    |                                 |    |                               |
|----------|---|----|-------------------------------|----|---------------------------------|----|-------------------------------|
| 動詞: 開設する |   |    |                               |    |                                 |    |                               |
| 格要素:     | <table border="1"> <tr> <td>が:</td> <td>                 名詞: 会社<br/>                 概念関係子: <i>agent</i> </td> </tr> <tr> <td>を:</td> <td>                 名詞: 事務所<br/>                 概念関係子: <i>object</i> </td> </tr> <tr> <td>に:</td> <td>                 名詞: 大阪<br/>                 概念関係子: <i>place</i> </td> </tr> </table> | が: | 名詞: 会社<br>概念関係子: <i>agent</i> | を: | 名詞: 事務所<br>概念関係子: <i>object</i> | に: | 名詞: 大阪<br>概念関係子: <i>place</i> |
| が:       | 名詞: 会社<br>概念関係子: <i>agent</i>   |    |                               |    |                                 |    |                               |
| を:       | 名詞: 事務所<br>概念関係子: <i>object</i>   |    |                               |    |                                 |    |                               |
| に:       | 名詞: 大阪<br>概念関係子: <i>place</i>   |    |                               |    |                                 |    |                               |

このデータから格助詞のボタンを抽出し、すべてのボタンに対しあらかじめ設定してあるボタン区分(表1)を割り当て、動詞ごとに、その動詞がどのボタンをいくつとるかを登録したインデックスファイルを作成した(表2)。

表 1: 表層格ボタン

| 格ボタン     | 記号 | 格ボタン     | 記号 |
|----------|----|----------|----|
| が        | a  | が、に      | b  |
| が、で      | b1 | が、において   | b2 |
| が、へ      | b3 | が、に対して   | b4 |
| が、によって   | b5 | が、から     | c  |
| が、と      | d  | が、から、に   | e  |
| が、から、へ   | e1 | が、から、と   | e2 |
| が、と、に    | f  | が、と、に対して | f1 |
| が、を、と    | g  | が、を、に    | h  |
| が、を、で    | h1 | が、を、において | h2 |
| が、を、へ    | h3 | が、を、に対して | h4 |
| が、を、から   | i  | が、を、から、に | j  |
| が、を、から、へ | j1 | が、を、から、と | j2 |
| が、を      | k  | が、について   | k1 |

表 2: 動詞別格ボタンファイル(一部)

| 動詞    | 用例数 | 格ボタンの区分と数  |
|-------|-----|--|
| 走る    | 480 | k=169 a=139 b=110 c=16 b3=12 b2=12 e1=6 h=5 e=3 h3=3 h2=2 b5=2 |
| 逃げる   | 95  | a=42 b3=18 b=14 c=14 e1=2                                      |
| 飛ぶ    | 195 | a=68 b=35 k=33 b3=20 c=16 b2=7 e=6 e1=4 b5=3 h=2               |
| 関係する  | 111 | b=66 a=26 d=18   |
| 関連する  | 114 | b=83 a=18 d=13   |
| 結びつく  | 107 | b=58 d=33 a=16   |
| 駆けつける | 54  | b=21 a=18 c=7 b3=5 b2=3  |
| 出かける  | 205 | b=129 b3=41 a=26 c=3 e=3                                       |

さらに、動詞ごとに格のボタンを用例の多い順に並び替え、用例数が全体の5%以下の格ボタンを削除したうえで、同一の格ボタンの並びを持つ動詞をまとめることにより、表3に示すようなデータを得た。

表 3: 格ボタンの並びと動詞の対応表(一部)

| 格ボタン       | 動詞   |
|------------|--|
| a          | 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 足りる 薄れる 変化する   |
| a b        | おさまる そろろう たまる のぼる みえる 近づく 現れる 広がる 困る 収まる 察わる 出現する 出来る 成長する 存在する 達する 運れる 定まる 登場する 努力する 燃える 発展する 表れる 落ち着く 劣る |
| a b b1     | でる 育つ 泣く 広まる 降る 合意する 止まる 出回る 進む 成功する 続く 裏生する 定着する 動く 普及する 儼れる 落ち込む   |
| a b b1 c   | 生まれる   |
| a b b1 k   | 光る 生じる   |
| a b b3     | 動き出す   |
| a b b3 c   | 延びる  |
| a b c      | うかがえる くる 見える 上がる 落ちる   |
| a b c b1   | あがる 成り立つ   |
| a b c e    | 届く   |
| a b d      | からむ 共通する 思える 対立する 分かる 変わる 絡む   |
| a b k      | よみがえる 完成する 輝く 驚く 欠かず 吹く 漂う 付く  |
| a b k b1   | 咲く 生きる 浮く  |
| a b k b3   | 移動する   |
| a b k b3 c | 飛ぶ   |
| a b k c    | 流れる  |
| a b k c b3 | 飛び出す   |
| a b k h    | にじむ  |
| a b1       | ふえる まとまる 悪化する 安定する 下がる 行われる 高まる 死ぬ 死亡する 成立する 生活する 増える 低下する 売れる 発足する 表面化する 崩れる 亡くなる                         |

次に、同じ格ボタンの並びを持つものをその比率によって分割した。同じ<a,b>というボタンの並びを持つ動詞であっても、a(ガ型)がb(ガ、二型)に対して極端に大きな値を持つものと、ほぼ同じ値を持つものではその意味構造に違いがあると考えられるからである。同様に、<b,a>という並びを持つ動詞も、bの値が大きく、<b>と同類と考えられるものと、bとaがほぼ同じ値を持つものに分割される。後者は、<a,b>でaとbが似た値を持つものと類似していることになる。表3からも明らかなように、一つの動詞がとる格ボタンの数が多いものは、そのグループに含まれる動詞が少数であるのに対し、<a,b><a,b1>のように、格ボタンの数が少ないものほど多くの動詞が集中する傾向がある。そこで、この分割処理は、二つのボタンしか持たないグループに対して行なった。

最後に、類似した格ボタンの並びを持ち、意味が近いと考えられるものをまとめる処理を行なった。また、任意格による影響と考えられるものを修正し、5%以下で切り捨てたボタ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ない、それぞれのグループごとに意味構造を考えた<sup>1</sup>。以下に例を示す。

### 単独変化動詞

これは、格のパターンとしてa(ガ型)のみを持つものである。ガ格には、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。「ない」→「なくなる」、「深い」→「深まる」、「薄い」→「薄れる」など、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ように、変化の結果は語彙の中に含まれている。それで、semantic structure は、次のようになる。

$$[\text{Event GO}_{\text{Ident}}([\text{Thing } \alpha]_A, \left[ \begin{array}{l} \text{Path} \\ \text{TO}([\text{State BE}([\alpha], [\text{AT}([\beta]]))]) \end{array} \right])] \quad (1)$$

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いるために項としては実現されず、変化の主体である GO<sub>Ident</sub> の第一引数のみに linking subscript の A が付されている。これがガ格として表層に現れる。

### 単独動作動詞

a(ガ型)を最も典型的にとるが、その他の格パターンも同時に持つ動詞は、ガ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となないものがある。

<a,b>という並びを持つ動詞 (任意的な原因を表すア格がついて<a,b,b1> 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、これらは二格で終点(goal)を表すものである。semantic structure は、

$$[\text{Event GO}([\text{Thing}]_A, [\text{Path TO}(\left\{ \begin{array}{l} \text{PLACE} \\ \text{THING} \end{array} \right\}]_{(A)}))] \quad (2)$$

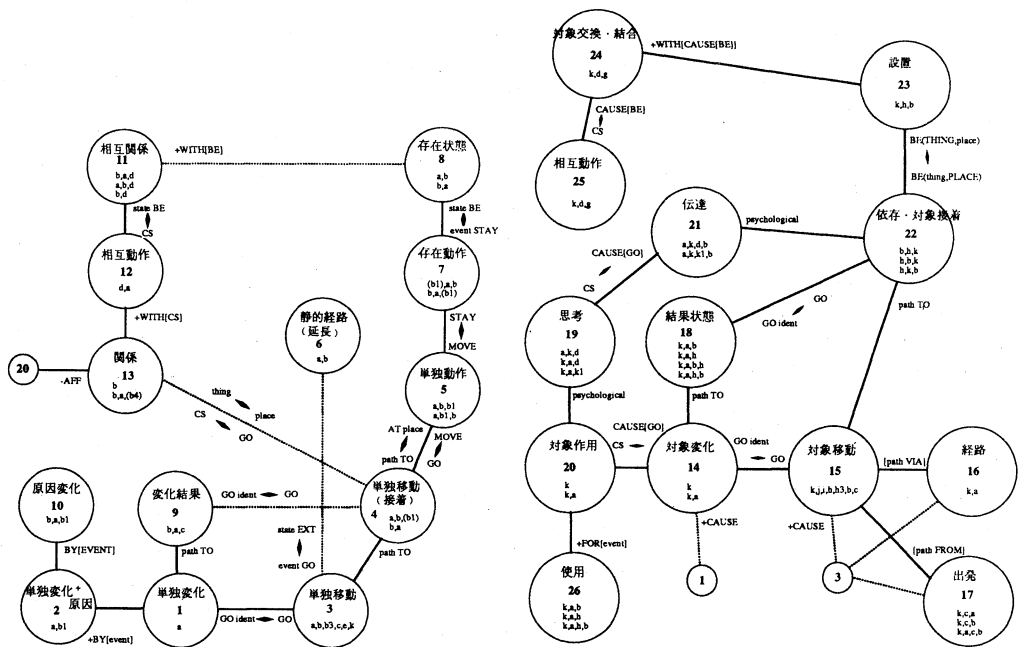
となる。

同様に、約30のグループに対して semantic structure を設定したが、紙数の都合ですべてを記述することはできない。詳細は大石 1995[6] を参照されたい。次節に、全体の構造を示す。

## 4 Semantic structure の関連図

次ページの右側の図は他動詞に、左側の図は自動詞に、それぞれ対応している。図中の○はカテゴリーを表し、カテゴリー名と、そのカテゴリーに属する動詞が持つ格パターンの並びを記載している。なお、○中の数字は各カテゴリーを識別するためのものであり、特に意味を持たない。カテゴリー同士をつなぐ線は、カテゴリー間の関係を表し、semantic structure の差異を付記している。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。

<sup>1</sup>以下で用いる表現法は、Jackendoff 1990[2] に則ったものである。



## 5 おわりに

本稿では、一つの動詞が複数の格パターンをとるときに、その組み合わせに基づいて動詞を分類し、その分類ごとに semantic structure を考える手法について述べた。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二格やデ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たと思われるものもある。「行なう」「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。

## 参考文献

- [1] Grimshaw, J. *Argument Structure*, Vol. 18 of *Linguistic Inquiry Monograph*, MIT Press (1990).
- [2] Jackendoff, R. *Semantic Structure*, MIT Press (1990).
- [3] 大石亨, 松本裕治 格パターン分析を利用した深層格獲得手法について, 情報処理学会研究報告, Vol. 94, No. 78 (94-NL-104) (Nov. 1994).
- [4] Pinker, S. *Learnability and Cognition: The Acquisition of Argument Structure*, MIT Press (1989).
- [5] (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).
- [6] 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, Master's thesis351202, 奈良先端科学技術大学院大学 情報科学研究科 (Feb 1995).