

## 語彙知識ベースにおける記述量削減のための語彙規則の形式化

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

自然言語処理において、語彙知識ベースの果たす役割は、理論的にも実用的にもますます増大してきている。さらに、意味理解を含む処理の高度化は、必然的に大量の語彙知識の記述を要求する。したがって、これらの知識をいかに体系づけて、冗長な記述を削減するかが、重要な課題となる。本稿では、既存の語彙項目から新たな語彙項目を派生させる語彙規則 (lexical rule) を類型的に形式化することによって、語彙知識ベースの記述量を削減する試みについて述べる。ここで、明らかにした規則のパターンは、“coercion” などの動的な処理にも応用することができる。

## Formulation of Lexical Rules for a Reduction of Descriptions in Lexical Knowledge Bases

Akira Oishi Yuji Matsumoto

Graduate School of Information Science, Nara Institute of Science and Technology

In natural language processing (NLP), the importance of the lexical knowledge base (LKB) has become increasingly emphasized both in theoretical and practical viewpoints. Natural language understanding systems require rich semantic representation for each lexical entry, as well as coherent relation between those entries so that they can perform deeper inference necessary for discourse understanding. In this paper, we describe a formulation of lexical rules to simplify our descriptions in the lexical knowledge base. Given the lexical rules, the output (inflected, derived, or “shifted”) form can be predicted from the input (base, stem, or “unshifted”) form. The patterns of the rules can be used in dynamic applications such as “coercion”.

## 1 はじめに

自然言語処理システムにおいて、辞書を含む知識ベースが重要な役割を果たすことは言うまでもない。意味処理、文脈理解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。一方、HPSG(Head-driven Phrase Structure Grammar)[5]をはじめとする最近の言語理論は、多くの個別的な情報を辞書中の語彙項目に持たせることによって、句構造規則の数を減らし、普遍的な原理を確立している。すなわち、簡潔な文法記述のために、複雑な語彙記述をせざるをえないという事情がある。そこで、何らかの方法で複雑で冗長な語彙記述を削減する必要があるが、このために、従来からさまざまな手法が提案されている。なかでも、代表的なもの、継承 (inheritance) と語彙規則 (lexical rule) を利用するものである。前者は、情報の一般性 (generality) に基づいて階層構造をなすソート (タイプ) を指定することによって、個別の語彙項目に記載する情報を削減するものである。上位ソートに記載されている情報は下位ソートに継承され、下位ソートはそのソートに特有の情報のみを記載すればよい。後者は、二つ以上の語彙またはその語彙が属するクラス間の関係を記述するものである。継承が垂直的 (vertical) な方向の記述量削減手法であるとすれば、語彙規則は水平的 (horizontal) な方向の削減手法であるといえる。多義語に対して、それぞれの意味を語彙規則の出力とすることについては、辞書中の語彙項目の数を増大させるという理由から批判があり、Pustejovsky(1991)[6]では、これに代わる生成的なメカニズム (generative device) が提案されている。その中で、中心的な役割を果たしているのが、語彙項目に含まれる qualia structure と呼ばれる構成的な要素と、type coercion である。これらのメカニズムによって、一つの語が特定の文脈の中で用いられるときに、さまざまな意味を持ち得ることが説明される。

本稿では、語彙規則という形でこれらのメカニズムを形式化する試みについて述べるが、その記述と実際のシステムでの使用は独立に考えられるべきものである。すなわち、ここで語彙規則として静的に処理しているものを、coercion rule として動的に適用することも可能である。また、以下では、接辞付加や複合語についても語彙規則として記述しているが、語幹、接辞等を独立に辞書登録しておいて、実行時に派生・複合させることも考えられる。どちらのやり方を選択するかは、経済性 (economy) と効率性 (efficiency) のトレードオフであり、個別のシステムごとに最適な方法を検討すべきである。

本稿の目的は、

- ・規則に内在するメカニズムは何か
- ・規則が適用される語彙に対する制約は何か

を明らかにすることである。これによって、その規則が適用されるべき意味クラスを構成する意味要素 (meaning component) の発見を含めた言語的知識の最適な一般化 (generalization) が可能になる。たとえば、Levin1993[4]では、英語の動詞について、Conative, Part-possessor ascension, Contact Locative, Middle, Anticausative という五つ語彙規則によって、“hit”, “cut” などの典型的な他動詞を四つのクラスに分類し、それぞれのクラスを構成する意味要素を定義している。

Alternation	Subclass
Conative	+motion,+contact
Part-possessor ascension	+contact
Contact Locative	+motion, +contact, -effect
Middle	+effect
Anticausative	+effect, -contact, -motion

Verb	Elements in semantic structure defining subclass membership
hit	motion, contact
cut	motion, contact, effect
break	effect
touch	contact

日本語の動詞について同様の処理を行なうためにも、まず、適用すべき語彙規則を明らかにする必要がある。以下では、動詞を中心とする語彙規則を次のように分類し、それぞれの規則が語彙の構造の中でどのような役割を果たしているかを、具体例を挙げながら説明する。

- 形態変化を伴うもの (word formation)<sup>1</sup>
  - 接辞付加 (affixation)
  - 複合動詞 (compounding)
- 形態変化を伴わないもの (meaning transfer)
  - 認知的焦点変化 (cognitive focus change)
  - 抽象化 (abstraction)

なお、以下で用いる素性名およびソート等は、概略、HPSG[5]に則ったものであるが、日本語用に、一部 JPSG [10]の用語を取り入れている。

<sup>1</sup>多くのシステムでは、活用語尾変化を語彙規則で取り扱っており、我々も現在のところ、語彙規則としてインプリメントしているが、郡司 (1994)[10]や Krieger and Nerbonne(1993)[3]で用いられている “distributed disjunction” を用いれば、規則の必要はなく、ここでは取り上げない。

## 2 接辞付加

この節では、接辞による語の派生 (derivation) について述べる。語形成規則については、多くのシステムが処理の都合上、手続き的に規則を適用しているが、Krieger and Nerbonne(1993)[3] のように、文レベルと同じような原理を語レベルに導入する試みもある。我々は、現在のところ、パーサーおよび論理プログラミング言語として ALE(Attribute Logic Engine)[2] を用いているので、そこで用いられている手続き的な方法で語彙規則を記述しているが、記述する情報量に変わりはない。なお、以下で例示する規則等は、ALE の記法を用いている。

ここでは形容詞を動詞化する接尾辞「める」を例に挙げて、我々の形式化を説明する。「高める」「深める」「速める」「弱める」「強める」などの動詞は、形容詞の語幹に接尾辞「める」が付いたものと考えられる。この種の動詞の意味の特徴は、働きかけを受けて変化をするもの、すなわち、「を」格の名詞が具体的な物や人ではなく、それらの動き、状態、特徴、あるいは関係であるということである。実際、動きや状態などに働きかけて、そこに変化を引き起こすということは、その持ち主である具体的な物や人に対して直接的または間接的に働きかけることを媒介にしなければ成り立たないが、これらの動詞は、変化を表す全過程のうちから、働きかけの具体性は切り捨てて、変化を受ける側面だけを取り出している [9]。

我々は、HPSG に基づいて、語や句の意味内容を表す素性 'SYNSEM:LOC:CONT' の値がとるソート *sem\_obj* の下位ソートを図1のように指定している。イタリック体で示したものがソート名であり、“ [ ] ” 中のローマン体で示したものは、そのソートにおいて導入される素性名である。コロンの後には、各素性がとる値の持つソートを指定している。下位ソートは、その上位ソートで導入された素性を引き継ぐ [1]。

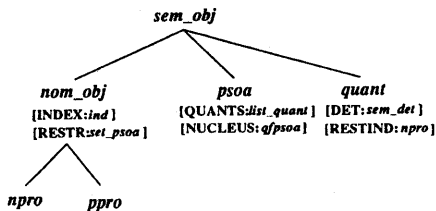


図 1: ソート 'sem\_obj' の下位区分

図 1 において、*nom\_obj* (nominal object) は体言、*psoa* (parametrized states-of-affairs) は用言、*quant*

(quantifier) は量子子の意味内容に対応する。(それぞれの下位ソートは、さらに区分されるが、ここでは省略している。)

以下に、この派生を記述した規則を示す。ALE では、入力と出力の間で変更されない素性についても同一性を保証するためにすべて記述する必要があるが、ここでは、関連のある素性のみを抜き出している。規則は、「規則名 'lex\_rule' 規則の記述 'morphs' 形態変化」という形式であり、「\* \* >」の前後に入力と出力の語彙に対する制約を記述してある。入力の記述と単一化可能な語彙見出しのみに、この規則が適用される。

```

adj_verb2 lex_rule
(word,
  synsem:(loc:(cat:(head:(v,
    vform:root,
    infl:adj),
    subcat:[(@ np(Ind2), @ case(ga))])),
  cont:nucleus:(Qf_psoa,
    scale:psoa),
  ...
**>
(word,
  synsem:(loc:(cat:(head:(v,
    vform:root,
    infl:vv),
    subcat:[(@ np(Ind1), @ case(ga),
      (@ np(Ind2), @ case(wo))])),
  cont:(nucleus:(bin_change,
    changer:Ind1,
    changee:Ind2,
    result:Qf_psoa),
    quants:[]),
  ...

morphs
  X becomes (X,m,e).
  
```

この記述において、「>」の左側が素性名、右側がその値を表している。「HEAD」素性は、品詞を指定するものである。*v* (verbal) は用言を表し、いわゆる動詞、助動詞だけでなく、形容詞、形容動詞と呼ばれているものを含む。我々は、JPSG にしたがって、これらの区別は品詞の区別ではなく、活用型の区別であるとみなしている。「INFL」は活用型を示す素性であり、*adj* は形容詞、*vv* は語幹が母音で終わる動詞の活用 (学校文法の一段活用に相当) を表す。「VFORM」は活用形を

示す素性であり、*root*は語幹そのものの形を表す。大文字で始まる文字列は変数であり、同じ変数名が指定されている時は、値の共有を示している。‘@’はマクロを示しており、‘@ np(Ind1), @ case(ga)’は、

```
loc:(cat:(head:n,
          subcat:[],
          marking:ga),
      cont:index:Ind1)
```

と展開される。

この規則によって、入力形容詞の「が」格は出力動詞の「を」格となり、出力される動詞の意味内容は、「が」格の名詞によって表される指示対象(Ind1)が「を」格の名詞によって表される対象(Ind2)の状態を、もとの形容詞が表していた状態(Qf<sub>psoa</sub>)に変化させるということになる。さらに、入力形容詞の意味内容は、ソート *qf<sub>psoa</sub>* で、素性として‘SCALE’を持つ。この素性は、状態を含む出来事 *event* を属性 *property* に変換するオペレーターの役割を果たしており、これが、この規則が適用される形容詞の意味的な制約として働く。例えば、物の属性を表す形容詞「赤い」と、状態の程度を表す形容詞「高い」の素性構造は、次のようになっている。(「山が高い」という場合の「高い」は前者と同様の構造を持っており、これが、後述する抽象化によって、後者のような構造を持つと考えられる。)

形容詞「赤い」の素性構造の一部

```
subcat:[loc:(cat:(head:n,
                  subcat:[],
                  marking:ga),
            cont:index:Ind)],
      nucleus:(red,
              inst:Ind))
```

形容詞「高い」の素性構造の一部

```
subcat:[loc:(cat:(head:n,
                  subcat:[],
                  marking:ga),
            cont:restr:(elt:Restr,
                       elts:e_set)),
      cont:(quants:[],
            nucleus:(high,
                    scale:Restr))
```

前者の意味内容の値はソート *red* であり、その素性として‘INST(instance)’を持つ。この素性の値が、「が」格の名詞句の‘INDEX’と共有されており、その名詞句

の実体が「赤い」という属性を持つことを示している。一方、後者の意味内容の値はソート *high* であり、その素性として‘SCALE’を持つ。この素性の値のソートは *psoa* であり、「が」格の名詞句の‘RESTR(restriction)’と共有されている。すなわち、「高い」ものは具体的な物や人ではなく、それらの動き、状態、特徴、あるいは関係であるということを示している。規則とマッチするのは後者だけであるので、「高める」は出力されるが、「\*赤める」は出力されない。また、「感度を高める」が言えて、「山を高める」が言えないことも説明できる。

### 3 複合動詞

複合動詞とは、動詞が複数個集まって、音韻的に全体として一つの動詞のような振舞いをするものである。図3は、複合動詞と意味クラスの関係を示したものである。矢印の起点と終点がそれぞれ複合動詞の前項と後項の動詞の属する意味クラスに対応している。

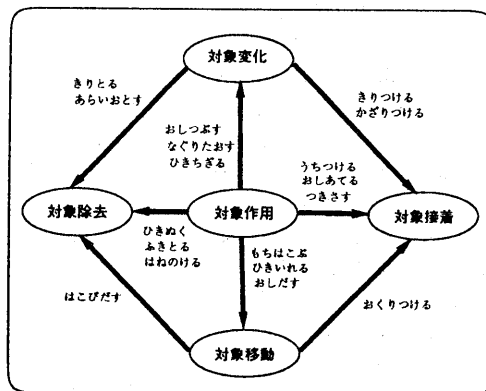


図3: Relations between categories

これらの複合動詞の前項は作用(様態)を表し、後項は移動(方向)を表している。前項の意味内容は、

```
cont:(nucleus:(bin_causing,
              manner:Manner)
```

と表される。ソート *bin\_causing* は、二項関係で、働きかけを表し、素性‘MANNER’によってその働きかけの様態を指定している。一方、後項は、

```
cont:(nucleus:(bin_moving,
              path:Path)
```

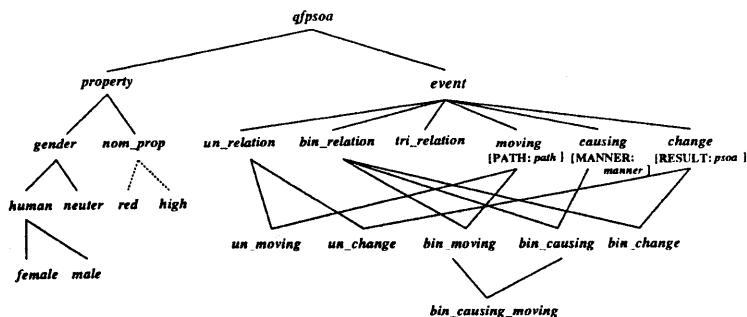


図 2: ソート 'qfpsoa' の下位区分

という意味内容を持つ。ソート *bin\_moving* は、二項関係の移動を表し、素性 'PATH' を持つ。'PATH' の値は、さらに 'TO', 'FROM' などの素性を持ち、その値が、「に」格、「から」格の名詞句と構造共有される。

出力される複合動詞の意味内容は、

```
cont: (nucleus: (bin_causing_moving,
               path: Path,
               manner: Manner),
```

となる。ソート *bin\_causing\_moving* は、*bin\_causing* と *bin\_moving* の両方のソートの下位タイプであり、双方の素性を引き継ぐ (multiple inheritance、図 2 参照)。一般に、一つの動詞が移動の様態と方向を同時に表すことはまれであるが [8]、二つが複合することによって、その語彙的ギャップを埋めている。

## 4 認知的焦点変化

前節で見たように、一つのソートは複数のソートを上位クラスに持ち、素性を多重継承することができる。一つの動詞が、このようなソートに属する場合には、その語彙的な意味の中に複数の上位ソートに属する意味を含んでいることになり、多義的になる。たとえば、

靴の紐を 結ぶ。

リボンを 棒の先に 結ぶ。

のように、「結ぶ」という動詞の意味には対象に働きかける動作 (*bin\_causing*) と、対象を移動させる動作 (*bin\_moving*) が含まれている (*bin\_causing\_moving*) が、どちらに焦点が置かれるかによって、表層に現れる格のパターンが変化する。

一方、二つのソートが表す意味の間に原因と結果の関係があり、動詞の表す意味が必然的にその両方を含

む場合には、それを関係付ける語彙規則が必要になる。たとえば、

ペンキを 壁 に塗る

壁 をペンキで塗る

のように、意味的に同じ役割を果たす名詞が、異なる格で表示される。我々の形式化では、これらの格の関係と意味の違いを、ソートの指定によって、簡潔に表現することができる。

```
alternation1 lex_rule
(word,
 synsem: (loc: (cat: (head: (v)),
                 subcat: [(@ np(Ind1), @ case(ga)),
                          (@ np(Ind2), @ case(wo)),
                          (@ np(Ind3), @ case(ni))]),
 cont: (nucleus: (Qfpsoa, bin_moving_causing,
                 path: to: Ind3,
                 manner: make_change),
 ...
**>
(word,
 synsem: (loc: (cat: (head: (v)),
                 subcat: [(@ np(Ind1), @ case(ga)),
                          (@ np(Ind3), @ case(wo)),
                          (@ np(Ind2), @ case(de))]),
 cont: (nucleus: (bin_change,
                 changer: Ind1,
                 changee: Ind3,
                 result: a_state,
                 by: Qfpsoa),
 ...
```

この規則は、ある対象(Ind2)をある場所(Ind3)に移動させることによって、その場所に変化が起こることを示している。入力の記事中の素性‘MANNER’のソートによって、変化の発生の有無を区別している。出力の記事中の素性‘RESULT’の値は *a\_state(after state)* で、その原因を素性‘BY’によって表示する。この値が、入力の‘CONT:NEUCLEUS’の値と共有されている。この素性‘BY’は、イベントを状態に写像するオペレーターである。

同様の例として、変化→生産(「水を湯に沸かす」「湯を沸かす」、人に対する働きかけ→モーダルな態度(「彼をうながす」「彼に決心をうながす」)などがある。また、この他の認知的焦点変化には、過程を表す具体的な作用動詞が、内在と構成といった論理的関係を表すのに用いられるといったアスペクチュアルな変化がある(「くみたてる」「かたちづくる」「もつ」「ふくむ」「おびる」など)。Pustejovsky(1995)[7]では、

John baked the potato.

John baked the cake.

という例をあげて、状態変化と生産という意味の違いを、“cocomposition”と呼ばれるメカニズムで取り扱っている。そこでは、動詞のとる名詞(ここでは直接目的語)が動詞の意味に働きかけて、そのイベントタイプを動的に変更するとされている。このためには、動詞のみならず、名詞の意味構造を詳しく指定する必要があるが、その際に、双方の意味的なクラスを考慮しておかないと、あらゆる名詞と動詞の意味の側面(qualia structure)の組合せを試さなくてはならない。さらに、上の例のように、二つ以上の名詞句が関与する場合には、より複雑な操作が必要となる。むしろ、ここでは、原因と結果という推論的な要素が働いていると考えるのが自然であり、この関係が成り立つ意味クラスをあらかじめ指定しておくことによって、不必要な操作の可能性を除去することができる。

## 5 抽象化

抽象化とは、文字通り、具体的な語彙の意味に“ずれ”が起こって、より抽象的な意味を実現している場合を指す。抽象化は、動詞が格にとる名詞の素性によってもたらされる。2節で述べたように、HPSGでは、名詞の意味内容は、ソート *nom\_obj* であり、素性‘INDEX’と‘RESTR’を持つ。‘INDEX’の値は、ソート *ind* であり、指示対象の一致素性(‘PERSON’,‘NUMBER’,‘GENDER’)を表示する。‘RESTR’の値は、ソート *set\_psoa* で、その名詞に対す

る意味的な制限が記述される。我々は、人、物、事の違いをこの *psoa* の素性‘NUCLEUS’の値が持つソート *qfpsoa* のサブソートによって区別している(図2参照)。すなわち、ソート *nom\_prop* (*red,high* など形容的な属性を除く)によって物を、ソート *human* によって人を、ソート *event* によって事を表す。以下では、動詞の項に対するこれらのソートの制限と、それが変更されることによってもたらされる動詞の意味クラスの変更について述べる。

「うちのめす」「きずつける」「(精神的に)ころす」「へこます」「おしつぶす」「おいつめる」などの動詞は、物にたいして物理的に働きかけることを表す具体的な作用動詞が、人を示すヲ格の名詞と組合わさって、意味にずれ=抽象化をおこし、心理的な状態変化または心理的態度(働きかけ)を示す動詞に移行したものである。これらの動詞はその意味的な特徴として、対象にあるダメージを与える変化を表すということができる。前節で述べたように、対象に変化をもたらし作用は、*bin\_change* というソートによって表されるが、このソートは、別の次元「意味分野(semantic field)」によって、*bin\_phys\_change*、*bin\_psych\_change* という下位ソートに区分される。前者は、物理的(physical)な意味分野を、後者は心理的(psychological)な分野を表す。これを規則に対する制約として表せば、

```
abst(wo,nom_prop,bin_phys_change,
      result:bad_obj_state,
      human,bin_psych_change,
      result:bad_psy_state).
```

となる。引数は、順に、入力の動詞の‘SUBCAT’素性の中の名詞句の‘MARKING’素性の値(wo)、『CONT:RESTR:RLT:NUCLEUS’の値(nom\_property)、入力の動詞の‘CONT:NUCLEUS’素性の値(bin\_phys\_change)とその値が持つ素性及び値(‘RESURT:bad\_obj\_state’)、出力の動詞の‘SUBCAT’素性の中の名詞句の‘CONT:RESTR:RLT:NUCLEUS’の値(human)、出力の動詞の‘CONT:NUCLEUS’素性の値(bin\_psych\_change)とその値が持つ素性及び値(‘RESURT:bad\_psy\_state’)である。

同様の例として、「わたす」「とる」「あげる」などの動詞が、具体的移動から所有移動の意味に用いられる場合がある。このときには、二格の名詞が場所から人に変化する。

```
abst(ni,place,bin_locat_moving,
      pass:to:place,
      human,bin_poss_moving,
```

pass:to:human).

*bin\_locat\_moving*, *bin\_poss\_moving* は、それぞれ、場所 (locative)、所有 (possessive) という意味分野の違いを表す。この他の意味分野としては、生理的 (*physiology*)、状態変化 (*identificational*) などがある。

一方、「を」格の名詞が物から事へ変化する例としては、「おこす」「よぶ」「まねく」「よびおこす」など、人に対する働きかけを表す動詞や、「うむ」「うみだす」「きづく」などの生産を表す動詞が、状態の出現を表すこと、「意識する」「認識する」「みとめる」などの一般的な認識動詞が、具体名詞を伴えば/知覚する/という意味になり、抽象名詞を伴えば/考える、理解する/という意味になること、「つかまえる」「つかむ」「とらえる」「のみこむ」「くみとる」などの把握を表す動詞が抽象名詞をヲ格にとって、知的な活動を表すこと、「あたえる」「うける」という動詞が、「悲しみをあたえる」「衝撃をうける」のように状態を表す抽象名詞をとるときには、それによって示される状態を/つくりだす/という意味になり、「刺激をあたえる」「指導をうける」のように、動作性の抽象名詞をとるときには単に/する、される/といった助動詞的な意味になることなどが挙げられる。これらも、同様に、名詞のソート (*nom\_prop*→*event*) と、意味分野の組合せによって記述することができる。現在、これらは、語彙規則に付属する制約として記述しているが、名詞句と動詞のソートが一致しない時に動詞の意味を動的に変更するように記述することも可能であろう。

## 6 おわりに

本稿では、語彙知識ベースの記述量を削減するための、語彙規則の形式化について述べた。その中では、‘SCALE’や‘BY’のようなソートを変更するためのオペレーターとしての素性、および、意味分野を変更する関数 ‘*abst*’ を導入した。これらは、語彙規則ではなく、“coercion” [7] などの動的な処理においても重要な役割を果たすであろう。動的な処理は、「信頼をおとす」「誤解をとく」「空腹をみたく」「注意をひく」などのように、特定の単語と組合わさることで、語彙的な意味が特殊なものにずれてゆく場合や、「品質をおとす (悪くする)」、「評判をおとす (失う)」、「値段をおとす (安くする)」のように、語彙的な意味とこれに基づく文法的な結合能力が一般的であるために、物や人を表す具体名詞とも抽象名詞とも自由に組合わさることができる動詞の意味を理解する時に、特に必要となる。そのためには、“qualia structure” を含めた名詞

の意味構造を明らかにすることが不可欠である。さらに、動的な処理を行なう場合においても、ここで述べたようなパターンをあらかじめ明らかにしておくことは、不必要な計算を除く上でも重要である。

## 参考文献

- [1] Carpenter, B.: *The Logic of Typed Feature Structures*, Cambridge Tracts in Theoretical Computer Science, No. 32, Cambridge University Press (1992).
- [2] Carpenter, B. and Penn, G.: *The Attribute Logic Engine: User's Guide*, Carnegie Mellon University (1994).
- [3] Krieger, H.-U. and J.Nerbonne: *Feature-Based Inheritance Networks for Computational Lexicons*, Studies in Natural Language Processing, Cambridge University Press, chapter 7, pp. 90-136 (1993).
- [4] Levin, B.: *English Verb Classes and Alternations*, The University of Chicago Press (1993).
- [5] Pollard, C. and Sag, I. A.: *Head-Driven Phrase Structure Grammar*, Studies in Contemporary Linguistics, The University Chicago Press (1994).
- [6] Pustejovsky, J.: The Generative Lexicon, *Computational Linguistics*, Vol. 17, No. 4, pp. 409-441 (1991).
- [7] Pustejovsky, J.: Linguistic constraints on type coercion, *Computational Lexical Semantics* (Saint-Dizier, P. and Viegas, E.(eds.)), Studies in Natural Language Processing, Cambridge University Press, chapter 4, pp. 71-97 (1995).
- [8] Talmy, L.: *Lexicalization patterns: semantic structure in lexical forms*, Language Typology and Syntactic Description, Vol. 3, Cambridge University Press, chapter 2, pp. 57-308 (1985).
- [9] 奥田靖雄: を格の名詞と動詞のくみあわせ, 日本語文法・連語論 (資料編), 言語学研究会, むぎ書房, pp. 21-149 (1983).
- [10] 郡司隆男: 日本語句構造文法の概要, 日本語句構造文法に基づく効率的な構文解析の研究 (郡司隆男 (編)), 大阪大学言語文化研究科, pp. 1-37 (1994).