

## 語義文からの動詞間の上位-下位関係の抽出†

富 浦 洋 一†† 日 高 達†† 吉 田 将†††

本格的な意味処理を行うためには、単語の意味に関する知識が必要であることは言うまでもない。その中で、語彙間の上位-下位関係は、最も基本的な知識の1つである。本研究では、動詞間の上位-下位関係を国語辞典から抽出する手法を開発することを目的とする。動詞は第一階述語論理では $n$ 項述語に対応する。また、動詞は一般に多義であり、語義が異なれば、上位-下位関係にある動詞も異なる。そこで、動詞をその語義ごとに述語に対応させ、動詞間の上位-下位関係を述語間の関係として捉える。したがって、動詞間の上位-下位関係の抽出では、単に見出し動詞と上位-下位関係にある動詞(定義動詞)を抽出するだけでなく、定義動詞の語義の選択、および、見出し動詞と定義動詞の変数の対応も考慮しなければならない。本稿では、まず、(1)動詞間の上位-下位関係を論理的に定義し、(2)見出し動詞とその語義文の論理的関係、および、語義文の統語構造と論理的性質について述べ、(3)動詞間の上位-下位関係を示す情報は語義文の統語構造中のどこに現れるかについて述べる。さらに、(4)定義動詞の語義、および、見出し動詞との変数の対応を適切に選択するための必要条件とヒューリスティックについて述べ、最後に、(5)抽出結果について述べる。

### 1. はじめに

本格的な意味処理を行うためには、単語の意味に関する知識が必要であることは言うまでもない。その中で、語彙間の上位-下位関係は、必然的な関係であり、推移律が成立するため効率良く表現でき、最も基本的な知識と言える。また、文の言い換えや、文と文の近さを求めるのに利用できる。名詞間の上位-下位関係に関しては、構文解析中で、意味処理に先行して用いられることもあって、幾つかの研究<sup>1),2)</sup>があるが、動詞に関してはほとんど研究されていない。本研究では、鶴丸<sup>1)</sup>と同様、国語辞典の語義文をその見出し語に関する語意知識収集のための一次情報と考え、これから、動詞間の上位-下位関係を抽出する手法を開発することを目的とする。

本来、国語辞典は人間が語彙の意味を調べるのに利用するものである。したがって、国語辞典には上位-下位関係が計算機に利用できるような形で、構造化されて陽に表現されているわけではない。そこで3章では、(1)見出し動詞とその語義文の論理的関係を定義し、(2)語義文の統語構造と論理的性質について述べ、これらから、上位-下位関係を導出することで、(3)語義文がどのような統語構造のとき、見出し動詞

とその語義文との間にどのような動詞間の上位-下位関係が抽出できるかについて述べる。

2章で述べるように、我々は、動詞の各語義ごとに述語を対応させ、動詞間の上位-下位関係を、述語間の関係として捉えている。見出し動詞と上位-下位関係にある動詞として語義文より抽出される動詞(これを定義動詞と呼ぶ)が、その語義文においてどの語義で用いられているのかは語義文中には陽に示されていない。また、2章で上位動詞と下位動詞の変数の対応が重要であることを述べるが、見出し動詞と定義動詞の変数がどのように対応しているのかも語義文中には陽に示されていない。そこで、上位-下位関係の抽出の際には、

#### 定義動詞の語義

#### 見出し動詞と定義動詞との変数の対応

を適切に選択しなければならない。この完全自動化は困難であるが、4章で述べるように、必要条件を用いて、その候補を絞り、さらにヒューリスティックを用いてその候補に優先順位を付けることができる。これにより、適切な選択であると自動的に判定される順に人間に提示し、人間が自分の語意知識に基づいて正しい選択を判断することで、抽出支援システムとして使用できる。5章では、このようにして作成したシステムを用いて行った抽出実験の結果について述べる。

本稿では、第一階述語論理では表せない補文を要求する動詞、内包的動詞は取り扱わない。また、語義文の主節の述部が動詞である語義文を対象とする。なお、国語辞典としては、情報処理振興事業協会の計算機用日本語基本動詞辞書<sup>3)</sup>を用いた。

† Extracting the Superordinate-Subordinate Relation between Verbs from Definition Sentences in Japanese Dictionaries by YOICHI TOMIURA, TORU HITAKA (Department of Electronics, Faculty of Engineering, Kyushu University) and SHO YOSHIDA (Department of Artificial Intelligence, Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology).

†† 九州大学工学部電子工学科

††† 九州工業大学情報工学部知能情報工学科

## 2. 動詞間の上位-下位関係

まず、本稿で用いる略記法について触れておく。

$$x : x_1, \dots, x_n$$

$$P(x) : P(x_1, \dots, x_n)$$

$$P(x) : P_1(x_1) \wedge \dots \wedge P_n(x_n)$$

$$\exists xy : \exists x_1 \dots \exists x_n \exists y_1 \dots \exists y_m$$

$$\forall xy : \forall x_1 \dots \forall x_n \forall y_1 \dots \forall y_m$$

さらに表記の簡略化のために、具体例を除き、述語の変数の位置は意味を持たず、たとえば、 $P(x, y)$  は、‘ $P$  は  $x, y$  に関する述語である’ という意味のみを表すものである。

上位-下位関係は必然的かつ推移的な関係であり、第一階述語論理では含意に相当する。たとえば、「動物」と「人間」は上位-下位関係にある（「動物」は「人間」の上位語である）。名詞は個体の集合を示す一項述語に対応するので、これは、

$$\forall x[\text{人間}(x) \supset \text{動物}(x)]$$

と表現できる。ただし、人間( $x$ )は「人間」に対応する論理式、動物( $x$ )は「動物」に対応する論理式である。名詞の場合は、このように述語間の関係として表現しても、「動物」は「人間」の上位語である’ というように単に単語間の関係として表現しても、お互いの表現に容易に変換できるので、表している情報は同じである。

動詞は個体間の関係を示す  $n$  項述語に対応する。通常、名詞間で定義される上位-下位関係を動詞に拡張し、

$$\forall x[V^L(x) \supset V^U(x)] \quad (1)$$

なる関係を動詞間の上位-下位関係とする。ただし、 $V^L(x)$ ,  $V^U(x)$  は、下位動詞、上位動詞に対応する論理式である。たとえば、「自衛隊員は道路の亀裂を土砂で埋めた」の「埋める」と「彼は財布をポケットに入れた」の「入れる」に対して、「 $x$ が $y$ を $z$ で埋める」が成立すれば、「 $x$ が $z$ を $y$ に入れる」が成立する。これは次のように表現できる。

$$\forall xyz[\text{埋める}(x, y, z) \supset \text{入れる}(x, z, y)]$$

埋める( $t_1, t_2, t_3$ )は「 $t_1$ が $t_2$ を $t_3$ で埋める」に対応する論理式、入れる( $u_1, u_2, u_3$ )は「 $u_1$ が $u_2$ を $u_3$ に入れる」に対応する論理式である。この例でも分かるように、上位動詞と下位動詞の同一の格が必ずしも対応していない（「埋める」のデ格（手段格）が「入れる」のヲ格（対象格）に対応している）ので、格の対応まで表現しなければならない。本稿ではこれを変

数の対応（この例では、第1変数が第1変数に、第2変数が第3変数に、第3変数が第2変数に対応する）と呼んでいる。また、2つの述語間に対応のとれない変数が存在するが、その変数を存在限量すると、含意で結ばれる関係が成立するものがある。たとえば、「雨で畑が潤った」の「潤う」と「水が電気を帯びる」の「帯びる」との間には、

$$\forall x[\exists y \text{ 潤う}(y, x) \supset \exists z \text{ 帯びる}(x, z)]$$

すなわち、

$$\forall xy[\text{潤う}(y, x) \supset \exists z \text{ 帯びる}(x, z)]$$

が成立する。潤う( $t_1, t_2$ )は「 $t_1$ で $t_2$ が潤う」に対応する論理式、帯びる( $u_1, u_2$ )は「 $u_1$ が $u_2$ を帯びる」に対応する論理式である。このように2つの述語間に対応しない変数がある場合でも、本稿では上位-下位関係として扱う。ただし、文の言い換えなどに利用することを考えた場合、「雨で畑が潤う」を「畑がある何かを帯びる」では余りに情報が落ちるので、一般には次のように、存在限量された変数の値が存在する範囲を名詞で限定して強化したものを扱う。

$$\forall xy[\text{潤う}(y, x) \supset \exists z \text{ 水分}(z) \wedge \text{帯びる}(x, z)]$$

このように、名詞の場合と異なり、変数の対応が重要で、特に上位動詞と下位動詞の間で対応の取れない変数がある場合は、変数の対応はさらに重要となり、下位動詞に対応しない上位動詞の変数の値の存在域を名詞で限定していることも合わせて、動詞間の上位-下位関係は単に単語間の関係として表現するのではなく、述語間の関係として表現しなければならない。

語義が異なれば、上位あるいは下位の動詞も異なるので、語義ごとに述語を対応させ、上述のように(1)を拡張して動詞間の上位-下位関係を次のように定義する。

### 【定義1】 動詞間の上位-下位関係

動詞  $v^U$  の語義  $j$  が動詞  $v^L$  の語義  $i$  の上位語義であるとき、次式が成立する。

$$\forall xy[V^L_i(x, y) \supset \exists z[N(z) \wedge V^U_j(x, z)]]$$

ただし、 $V^L_i(x, y)$ ,  $V^U_j(x, z)$ ,  $N(z)$  は、それぞれ、 $v^L$  の語義  $i$  に対応する論理式、 $v^U$  の語義  $j$  に対応する論理式、名詞に対応する論理式の論理積である。 ■

なお、このように拡張しても推移性は失われない。たとえば、 $V^2_j$  が  $V^1_i$  の上位語義であり、 $V^3_k$  が  $V^2_j$  の上位語義とする。

$$\forall xy[V^1_i(y, x_1, x_2) \supset$$

$$\exists z[N_i(z) \wedge V^2_j(x_1, x_2, z)]]$$

$$\forall xy[V^2_j(y, x_1, x_2) \supset \\ \exists z[N_2(z) \wedge V^3_k(x_1, x_2, z)]]$$

これから,

$$\forall xy[V^1_i(y_1, y_2, x) \supset \\ \exists z[N_1(x_1) \wedge N_2(z_2) \wedge V^3_k(x, z_1, z_2)]]$$

が推論され、 $V^3_k$  は  $V^1_i$  の上位語義であることがわかる。

### 3. 上位-下位関係記述部の抽出

#### 3.1 見出し動詞と語義文の関係および抽出方針

動詞は一般に多義であり、国語辞典には各語義ごとに、見出し動詞の示す個体間の関係を語義文として記述している。したがって、見出し動詞とその語義文の関係は、

$$\forall x[V^*(x) \equiv S(x)] \quad (2)$$

と考えられる。 $V^*(x)$  は見出し動詞の1つの語義に対応する論理式、 $S(x)$  はその語義文に対応する論理式である。以後、単に“見出し動詞”だけでその特定の1つの語義を示すものとする。さらに、語義文中の動詞の語義の選択、見出し動詞との変数の対応は適切になされているものとして、議論を進める。

#### 【例1】

建つ  $\forall$  […に…が] 建物ができる。「駅前にビルが建った。」

ローマ数字は語義番号を、[ ] は受け得る格助詞のパターンを、「」は用例を示す。語義文中の「できる」は次の意味で用いられている。

できる  $\forall$  […が…に] それまでは存在しなかった何かが完成・成立する。「最近新しい法律ができた。」

名詞は個体の集合を示す一項述語に対応し、動詞は個体間関係を示す  $n$  項述語に対応することを述べたが、名詞と動詞から成る単文は、名詞の示す集合の要素である個体が、動詞の示す関係にあることを表す。

たとえば、「建つ  $\forall$ 」の語義文は、

$$\text{建物}(x_2) \wedge \text{できる}_4(x_2, x_1)$$

に対応する。建つ  $\exists$  ( $t_1, t_2$ ) が「個体  $t_1$  に個体  $t_2$  が建つ  $\forall$ 」を表し、できる  $_4(u_1, u_2)$  が「 $u_1$  が  $u_2$  にできる  $\forall$ 」を表すとする、(2)の関係、

$$\forall x[\text{建つ}_\exists(x_1, x_2) \equiv \text{建物}(x_2) \wedge \text{できる}_4(x_2, x_1)]$$

は我々の言語的直観に合う。

ところが、見出し動詞に対応しない自由変数が語義文中に現れる場合がある。

#### 【例2】

受ける  $\forall$  […が…に] 人気を得る。「その芝居は老人に受けた。」

語義文中の「得る」は次の意味で用いられている。

得る  $\exists$  […が…から…を] 相手の同調を受ける。

「知事は県民から圧倒的支持を得た。」

この場合、(2)の関係、

$$\forall x[\text{受ける}_\exists(x_1, x_2)]$$

$$\equiv \text{人気}(x_3) \wedge \text{得る}(x_1, x_2, x_3)]$$

は我々の言語的直観に合わない。なぜなら、上式より、

$$\forall x[\text{受ける}_\exists(x_1, x_2) \supset \text{人気}(x_3)]$$

が推論されるが、これは明らかに我々の直観に反する。

「人気を得る」を単独の文として考えると、

$$\exists x[\text{人気}(x_3) \wedge \text{得る}_\exists(x_1, x_2, x_3)]$$

となる。このように、日本語の場合、限量子に相当する表現がない場合は存在限量を表す「ある」を補って考えるべきであることがわかる。したがって、見出し動詞の変数に対応しない変数は存在限量を補って、

$$\exists zS(x, z)$$

と考え、これと、見出し動詞との間に同値関係があると考えられる。すなわち、見出し動詞とその語義文の関係は、次のように一般化できる。

【定義2】 見出し動詞とその語義文の論理的関係

$$\forall x[V^*(x) \equiv \exists zS(x, z)]$$

$V^*(x)$  は見出し動詞に対応する論理式、 $S(x, z)$  は語義文に対応する論理式である。

動詞の語義文には固有名詞は現れず、「すべての」のような全称限量に相当する表現はない。したがって、先の例で示したように、単文でかつ連体修飾部および副詞などの連用修飾句のない語義文は、

$$\exists z[N_x(x) \wedge N_z(z) \wedge V^d(x, z)] \quad (3)$$

に対応する。 $N_x(x)$ 、 $N_z(z)$  は名詞に対応する論理式の論理積、 $V^d(x, z)$  は述部の動詞に対応する論理式である。ただし、定義2のように、存在限量される  $z$  は見出し動詞と対応しない変数であり、変数  $x$  は見出し動詞との関係において全称限量されている。このような単純な語義文の場合、(3)と定義2から、見出し動詞  $V^*(x)$  と  $V^d(x, z)$  との間の上位-下位関係、

$$\forall x[V^*(x) \supset \exists z[N_z(z) \wedge V^d(x, z)]]$$

が得られることが容易に分かる。

語義文が必ずしも(3)の形式の論理式に対応するわけではないが、語義文  $S$  に対して、(3)の形式で表現可能な  $S$  の部分構造  $S'$  のうち、

$$S \supset S' \quad (4)$$

なる関係が成立するものがあれば、修飾部のない単文の場合と同様に、定義2と(4)から上位-下位関係が得られる。このような  $S'$  を上位-下位関係記述部と呼ぶ。

3.2 節では、語義文の統語構造とその論理的性質について述べ、3.3 節では、定義2と3.2 節で述べたことから、語義文がどのような統語構造のとき、上位-下位関係記述部を持ち、それは語義文中のどの部分構造であるか、そしてそれからどのような上位-下位関係が抽出されるかについて述べる。

### 3.2 語義文の論理的性質

まず、単文である語義文とそれから修飾部を取り除いた文との間の論理的な関係について考察する。前節でも述べたように、動詞の語義文には固有名詞は現れず、全称限量に相当する表現はない。したがって、単文の語義文は次のような論理式に対応する。

$$NP_1(x) \wedge \dots \wedge NP_n(x) \wedge VP(x)$$

ただし、変数の限量に関しては見出し動詞との関係において決るので、ここでは省略している。 $NP_i(x)$  は名詞句（「ある」のような存在限量に相当する表現があれば、名詞句からそれを除いた部分）、 $VP(x)$  は動詞句に対応する論理式である。

#### 【仮定1】

$$(a) \quad \forall x [NP(x) \supset N(x_i)]$$

$$(b) \quad \forall x [VP(x) \supset V(x)]$$

$N(x_i)$  は名詞句  $NP(x)$  から連体修飾部を取り除いた部分に対応する論理式、 $V(x)$  は動詞句  $VP(x)$  から副詞などの連用修飾句を取り除いた部分に対応する論理式である。

例えば、「とても好きだという気持」「しっかりつかむ」に関して、

$$\forall x [\text{とても好きだという気持}(x_1, x_2, x_3) \supset \text{気持}(x_3)]$$

$$\forall x [\text{しっかりつかむ}(x_1, x_2) \supset \text{つかむ}(x_1, x_2)]$$

が成立する。とても好きだという気持  $(x_1, x_2, x_3)$  は、「 $x_3$  は  $x_1$  が  $x_2$  をとても好きだという気持である」ことを意味する。

単文  $s$  から連体修飾部、連用修飾句を取り除いてできる文を  $s$  の骨格文と呼ぶ。仮定1より、次の性質が成り立つ。

#### 【性質1】 単文とその骨格文の関係

$$\forall xy [S(x, y) \supset S^k(x)]$$

$S(x, y)$  は単文に対応する論理式、 $S^k(x)$  はその骨格文に対応する論理式である。

次に、文を構成する文の性質について考える。1つの文内に主語述語の関係が複数存在するものは以下の場合である。

① 連体修飾部

② 補文

③ 連用中止による複数の文の連接

④ 接続詞、接続助詞を伴うもの

（「…時」「…ために」「…目的で」も拡張された接続詞として扱う。）

①は仮定1(a)で述べた。②は第一階述語論理では表現できないので本研究では取り扱わない。③④は、「事の解決を図るためにその力に頼る。」のように、2つの文が主従の関係にある複文と、「相手を讃え、栄誉を与える。」や「他から何かを貰ったり、引き継いだりする。」のように、個々の文が対等の関係にある重文とがある。

複文に関して次の仮定をする。

#### 【仮定2】

複文に対応する論理式を  $S(x, y)$ 、その主節に対応する論理式を  $S'(x)$  とすると、

$$\forall xy [S(x, y) \supset S'(x)]$$

が成立する。

語義文中の従属節は、条件節ではなく、原因、理由、手段、目的などを示していて、どれも主節の成立を制限するものではない。例えば、「自分の求めるものを見つけるためにある場所を調べる」が成立するならば、「ある場所を調べる」が成立し、従属節「自分の求めるものを見つけるために」は目的が何であるかを付加するだけで、主節の成立を制限していない。

また、重文に関して次の仮定をする。

#### 【仮定3】

重文に対応する論理式を  $S(x, y)$ 、その重文を構成する個々の文に対応する論理式を  $S_i(x)$  とすると、次式を満たすものがある。

$$\forall xy [S(x, y) \supset S_i(x)] \quad (5)$$

(5)が成立するのは、「かつ」的な意味で結ばれる重文で、これを連言的重文と呼ぶ。

文  $s'$  が複文  $s$  の主節または重文  $s$  を構成する文であるとき、 $s$  を  $s'$  の親の文と呼ぶ。 $s'$  と親の文の推移的閉包の関係にある文を  $s'$  の祖先の文と呼び、逆に  $s'$  を  $s$  の子孫の文と呼ぶ。また、 $s$  が単文のとき、 $s$  は  $s$  の祖先の文であり子孫の文であるとする。仮定2、仮定3より、次の性質が成り立つ。

【性質 2】 文  $s$  とその子孫の単文  $s'$  の関係

$s'$  の祖先が単文、複文、連言的重文のみの場合

$$\forall xy[S(x, y) \supset S'(x)]$$

が成立する。ただし、 $S(x, y)$ ,  $S'(x)$  は、それぞれ  $s$ ,  $s'$  に対応する論理式である。 ■

### 3.3 上位-下位関係記述部の抽出

語義文の子孫の単文  $s'$  の祖先が、単文、複文、連言的重文のみの場合を考える。以下のように記号を定義する。

$$y = y_1, y_2, \quad t = t_1, t_2$$

$V^e(x, y)$  : 見出し動詞に対応する論理式,

$S(x, y, z, t)$  : その語義文に対応する論理式,

$S'(x, y_2, z, t_2)$  :  $s'$  に対応する論理式,

$S^k(x, z)$  :  $s'$  の骨格文に対応する論理式,  
 $= N_x(x) \wedge N_z(z) \wedge V^d(x, z)$

これらの間に以下の関係が成立する。

$$\forall xy[V^e(x, y) \supset \exists ztS(x, y, z, t)]. \quad (\text{定義 2})$$

$$\forall xyz[S(x, y, z, t) \supset S'(x, y_2, z, t_2)].$$

(性質 2)

$$\forall xyzt[S'(x, y_2, z, t_2) \supset S^k(x, z)]. \quad (\text{性質 1})$$

ゆえに,

$$\forall xy[V^e(x, y) \supset \exists zS^k(x, z)]. \quad (6)$$

これから,

$$\forall xy[V^e(x, y) \supset \exists z[N_x(x) \wedge V^d(x, z)]] \quad (7)$$

を得る。

【抽出規則】 上位-下位関係記述部

$s'$  を語義文の子孫の単文とする。 $s'$  の祖先が、単文、複文、連言的重文のみの場合、 $s'$  の骨格文が上位-下位関係記述部である。 ■

上位-下位関係記述部の述部の動詞を定義動詞と呼ぶ(前述の記号の定義では  $V^d(x, z)$  に対応)。語義文から上位-下位関係記述部を抽出し、定義動詞の語義、見出し動詞と定義動詞の変数の対応を適切に選択すれば、(7)式に対応する上位-下位関係が抽出できる。

## 4. 抽出の際の必要条件とヒューリスティック

前章では,

(a) 定義動詞の語義

(b) 見出し動詞と定義動詞の変数の対応

が適切に選択されているものとして、議論を進めてきた。実際の上位-下位関係の抽出に当たっては、(a)(b)の選択を適切に行わなければならない。ところが、1章でも述べたように、国語辞典には(a)(b)に関する

情報は明示されていない。そこで本章では、(a)(b)の選択が適切であるための必要条件とヒューリスティックについて述べる。

### 4.1 必要条件

必要条件を述べるために、まず、述語の有意味定義域を導入する。

【定義 3】 有意味定義域

述語  $P(x)$  の  $x_i$  に関する有意味定義域  $D_{Pi}(x_i)$  は

$$\forall x[P(x) \supset D_{Pi}(x_i)] \quad (8)$$

を満足するものである。 ■

【例 3】 有意味定義域

‘建つ<sub>s</sub>( $x, y$ )’, ‘学生( $x$ )’ を, ‘建つ<sub>v</sub>’, ‘学生’ に対応する述語とする。‘建つ<sub>s</sub>( $x, y$ )’ の  $x$  に関する有意味定義域(ニ格の有意味定義域)は ‘場所( $x$ )’ であり,  $y$  に関する有意味定義域(ガ格の有意味定義域)は ‘生産物( $y$ )’ である。また, ‘学生( $x$ )’ の有意味定義域は ‘人間( $x$ )’ である。 ■

有意味定義域は、名詞句の意味素性、意味マーカと呼ばれてきたものに相当する。

従来、構文解析で用いられてきた名詞と動詞の共起関係による選択制約により、定義動詞の語義の候補を絞る。選択制約は、次の必要条件 1 に相当する。

【必要条件 1】

上位-下位関係記述部の定義動詞に係る名詞の格助詞とその名詞の有意味定義域の組を

$$\langle CN_1, DN_1(x) \rangle, \dots, \langle CN_m, DN_m(x) \rangle \quad (9)$$

とし、定義動詞の語義  $\alpha$  での受け得る格助詞と有意味定義域の組を

$$\langle CD_1, DD_1(x) \rangle, \dots, \langle CD_n, DD_n(x) \rangle \quad (10)$$

とする。定義動詞の語義  $\alpha$  の選択が適切であるための必要条件は、次の条件を満たす(9)から(10)への 1 対 1 対応  $\phi$  が存在することである。

任意の  $i$  ( $1 \leq i \leq m$ ) に対して、ある  $j$  ( $1 \leq j \leq n$ ) が存在して、

$$\langle i, j \rangle \in \phi, \quad CN_i = CD_j,$$

$$\exists x[DN_i(x) \wedge DD_j(x)]. \quad \blacksquare$$

【仮定 4】 見出し動詞の有意味性

見出し動詞を真にする個体の組は存在する。 ■

定義動詞の語義の選択、見出し動詞と定義動詞の変数の対応の選択が適切であるならば(6)の関係が成立する。仮定 4 と(6), (8)より,

$$\exists x[De_k(x) \wedge DN_i(x) \wedge DD_j(x)] \quad (11)$$

を得る。ただし、見出し動詞の  $k$  番目の変数と定義動詞の  $j$  番目の変数が対応しており、上位-下位関係記

述部の  $i$  番目の名詞が定義動詞の  $j$  番目の変数に対応する格で係っているものとし、 $D_{ek}(x)$ ,  $D_{Ni}(x)$ ,  $D_{aj}(x)$  はそれぞれ見出し動詞の  $k$  番目の変数に関する有意味定義域、 $i$  番目の名詞の有意味定義域、定義動詞の  $j$  番目の変数に関する有意味定義域である。したがって、逆に (11) を、定義動詞の語義の選択、見出し動詞と定義動詞の変数の対応が適切であるための必要条件と考えることができる。

#### 【必要条件 2】

上位-下位関係記述部に対して、定義動詞の語義  $\alpha$  が必要条件 1 を満足し、対応  $\phi$  が必要条件 1 を満足させる対応とする。定義動詞の語義  $\alpha$  での各変数に関する有意味定義域を

$$D_{aj}(x) \quad (j=1, \dots, n),$$

骨格文の名詞の有意味定義域を

$$D_{Ni}(x) \quad (i=1, \dots, m),$$

見出し動詞の各変数に関する有意味定義域を

$$D_{ek}(x) \quad (k=1, \dots, l)$$

とすると、定義動詞の語義  $\alpha$ , 対応  $\phi$ , 見出し動詞の変数から定義動詞の変数への 1 対 1 対応  $\Psi$  の選択が適切であるための必要条件は、任意の  $\langle k, j \rangle \in \Psi$  に対して、

ある  $i$  に対して  $\langle i, j \rangle \in \phi$  ならば

$$\exists x [D_{ek}(x) \wedge D_{Ni}(x) \wedge D_{aj}(x)]$$

任意の  $i$  に対して  $\langle i, j \rangle \notin \phi$  ならば

$$\exists x [D_{ek}(x) \wedge D_{aj}(x)]$$

が成立することである。 ■

#### 4.2 ヒューリスティック

語義文を構文解析し、得られた上位-下位関係記述部に対して、必要条件 1 (選択制約) を用いて定義動詞の語義の候補を絞り、その候補の語義ごとに、必要条件 2 を満たす見出し動詞と定義動詞の変数の対応を求める。さらに、次に述べるヒューリスティックを用いて、その候補に優先順位を付ける。

ある動詞を他の動詞を用いて定義する場合、最も平易で明瞭な定義は、互いの動詞のすべての変数の対応が取れている場合である。辞書編集者はなるべく平易で明瞭な語義文を記述すると期待できるので、見出し動詞を定義する場合、そのような動詞があるならばそれを定義動詞として語義を記述すると考えられる。

#### 【ヒューリスティック】

必要条件 1, 2 を満足する定義動詞の語義  $\alpha$ , 見出し動詞の変数から定義動詞の変数への 1 対 1 対応  $\Psi$  のうち、見出し動詞と定義動詞の間で対応しない変数

が少ない、定義動詞の語義  $\alpha$ , 対応  $\Psi$  の選択が適切な選択である。 ■

## 5. 実験

### 5.1 実験で用いる辞書について

情報処理振興事業協会技術センターから計算機用日本語基本動詞辞書 (IPAL)<sup>3)</sup> が公開された。IPAL では日本語の動詞のうちで語彙体系上ならびに使用頻度上重要であると考えられる基本的な和語動詞 861 語を取り上げ、それを意味および統語的特徴に基づいて下位区分し、それを 1 つの単位 (これをサブエントリと呼ぶ) として、意味、形態、統語、文法的カテゴリ、慣用表現などに関する情報を記述してある。そのうち、サブエントリ (1 つの語義に相当) に対する日本語文による意味記述、表層格による文型、各表層格に係る名詞の意味素性に関する情報から、本手法で必要な各語義ごとの、

① 語義文、

② 受け得る格とその格に係る名詞の有意味定義域が機械的に抽出できる。本実験では国語辞典として IPAL を用いる。

構文解析用の辞書として、②の情報を持つ辞書は現在多くの機関で実働あるいは開発中である。今後、メンテナンスを考慮してその語義の自然言語文による説明 (①) まで付加した構文解析用辞書、あるいは、辞書作成に掛かるコストを考慮して、既存の国語辞典に対してその語義ごとに②の情報を求めて作成した構文解析用辞書が開発されることは十分考えられる。そのような辞書に対しては、本稿で述べた手法により上位-下位関係が抽出できる。

次の例で用いられる意味素性の意味と意味素性間になり立つ包含関係を示しておく。

|              |             |
|--------------|-------------|
| hum : 人間     | pla : 植物    |
| par : 生物の部分  | nat : 自然物   |
| pro : 生産物・道具 | act : 動作・作用 |
| lin : 言語作品   | cha : 性質    |
| loc : 空間・方向  | con : 具体名詞  |
| phe : 現象     | abs : 抽象名詞  |

$$\text{hum}(x) \vee \text{pla}(x) \vee \text{par}(x) \vee \text{nat}(x) \vee \text{pro}(x) \supset \text{CON}(x),$$

$$\text{act}(x) \vee \text{lin}(x) \vee \text{cha}(x) \vee \text{loc}(x) \supset \text{ABS}(x).$$

con, phe, abs 間および hum, pla, par, nat, pro, act, lin, cha, loc 間のどの 2 つの間にも交わりはないものとする。なお、名詞の有意味定義域に対応する知識

は、IPAL 中から意味素性ごとの名詞の例を抽出し、一部こちらで IPAL に準じて用意した。

## 5.2 抽出例

「潤う I」を見出し動詞として抽出例を示す。その語義文、用例、受け得る格助詞とその有意味定義域の組は、

水分をたっぷりと帯びる。

「久しぶりの雨で田畑が潤った。」

{<デ, pro ( $t_1$ )  $\vee$  phe ( $t_1$ )  $\vee$  act ( $t_1$ )>},

<ガ, loc ( $t_2$ )  $\vee$  pla ( $t_2$ )>

である。語義文は単文なので、抽出規則より、その骨格文「水分を帯びる」が上位-下位関係記述部となる。「水分」の有意味定義域は「cha ( $t$ )」である。定義動詞「帯びる」の各語義に対する語義文、用例、受け得る格助詞とその有意味定義域の組は、

I ある性質・成分などを中心に含んで持つ。

「水は電気を帯びる。」

{<ガ, con ( $t_1$ )  $\vee$  nat ( $t_1$ )  $\vee$  lin ( $t_1$ )>},

<ヲ, phe ( $t_2$ )  $\vee$  cha ( $t_2$ )>

II ある役割や任務を引き受ける。

「代表が全権を帯びている。」

{<ガ, hun ( $t_1$ )>, <ヲ, abs ( $t_2$ )>}

III そういふ様子を持つ。

「彼女は顔が丸味を帯びている。」

{<ガ, hum ( $t_1$ )>},

<ガ, par ( $t_2$ )  $\vee$  phe ( $t_2$ )  $\vee$  abs ( $t_2$ )>},

<ヲ, cha ( $t_3$ )>}

IV 身体にその人の身分や権威を象徴する物を付ける。

「彼は胸に勲章を帯びている。」

{<ガ, hum ( $t_1$ )>, <ニ, par ( $t_2$ )>, <ヲ, pro ( $t_3$ )>}

である。必要条件 1 を満足する語義およびそのときの対応  $\phi$  は、

I, {<1, 2>} ...①

II, {<1, 2>} ...②

III, {<1, 3>} ...③

である。必要条件 2 を満たす見出し動詞と定義動詞の変数の対応  $\Psi$  は上の各場合に対して、

①  $\Psi = \{ \}, \text{or}, \{ \langle 1, 1 \rangle \}, \text{or}, \{ \langle 2, 1 \rangle \}$

②  $\Psi = \{ \}$

③  $\Psi = \{ \}, \text{or}, \{ \langle 1, 2 \rangle \}, \text{or}, \{ \langle 2, 2 \rangle \}$

である。このうちヒューリスティックにより最優先される定義動詞の語義と見出し動詞と定義動詞の対応は、

I,  $\Psi = \{ \langle 1, 1 \rangle \}, \text{or}, \{ \langle 2, 1 \rangle \}$

であり、人間による判断を加え次式を得る。

$\forall xy [\text{潤う}_1(y, x) \supset \exists z [\text{水分}(z) \wedge \text{帯びる}_1(x, z)]]$

## 5.3 抽出実験

抽出結果は以下のとおりである。

対象：見出し動詞と上位-下位関係記述部の組 1467  
得られた上位-下位関係：1030 (すなわち、真の解が必要条件を満たす)

残り 437 は、辞書中に定義動詞の適切な語義がないか、有意味定義域の不備か、今回のアルゴリズムが任意格に対処していないために抽出に失敗したものである。

1 組当りの 1 位解の平均の個数：4.1

1 位解に真の解が含まれる割合：70.1% (うち、語義文が単文の場合：81%)

1 位解とは先のヒューリスティックにより、最優先されるものである。1030 のうち単文は 533 あり、上位-下位関係抽出の支援システムとしては有効と言える。得られた上位-下位関係の性質は

最上位語義：353

上位-下位関係の最大の深さ：6

下位語義の多い最上位語義：

変える VII...55, 及ぼす I...31, 持つ V...25,

等である。

## 6. おわりに

動詞間の上位-下位関係を述語間の関係として定義し、語義文がどのような統語構造のとき、語義文のどの部分が見出し動詞とのどのような上位-下位関係を表しているのかについて述べ、抽出の際問題となる、定義動詞の適切な語義の選択、見出し動詞と定義動詞の変数の適切な対応をとるための必要条件とヒューリスティックについて述べた。最後に、IPAL を用いて抽出実験を行い、抽出支援システムとしての有効性を示し、得られた関係について若干の考察を行った。本格的な結果の調査、評価等に関しては改めて報告する予定である。

今後、今回の実験では行わなかった、「なる」、「する」、出現頻度の多い一部のサ変動詞を定義動詞とする語義文に対しても上位-下位関係の抽出実験を行う予定である。さらに、語義文から得られる上位-下位関係以外の関係についても研究を進めていきたい。

**謝辞** 本研究では、国語辞典として情報処理振興事業協会技術センターの計算機用日本語基本動詞辞書

IPAL を用いて実験を行った。担当者各位に感謝いたします。

なお、本研究は一部文部省科研費重点領域研究による。

### 参 考 文 献

- 1) 鶴丸弘昭, 兵頭竜二, 松崎 功, 日高 達: 語義を考慮した単語間の階層構造の抽出について, 情報処理学会自然言語処理研究会報告, No. 64, pp. 9-16 (1987).
- 2) 田中穂積, 仁科喜久子: 上位/下位関係シソーラス ISAMAP1 の作成(I)(II), 情報処理学会自然言語処理研究会報告, No. 64, pp. 25-44 (1987).
- 3) 計算機用日本語基本動詞辞書 IPAL (Basic Verbs) 説明書, 情報処理振興事業協会技術センター (1987).
- 4) 富浦洋一, 日高 達, 吉田 将: 語義文からの動詞間の階層関係の抽出, 情報処理学会自然言語処理研究会報告, No. 64, pp. 17-24 (1987).
- 5) 富浦洋一, 日高 達, 吉田 将: 語義文における動詞間の上位-下位関係, 九州大学工学集報, Vol. 62, No. 2, pp. 107-112 (1989).
- 6) 富浦洋一, 日高 達, 吉田 将: 語義文からの動詞の上位-下位関係の抽出, 九州大学工学集報, Vol. 62, No. 3, pp. 189-195 (1989).

### 付録 抽出例

抽出した関係の一部を示す。述語の添字は IPAL の No の項目で語義番号に相当する。なお  $x_i, y_i$  は全称限量,  $z_i$  は存在限量されているものとする。

愛する<sub>1</sub>( $x_1, x_2$ ) $\supset$ 気持ち( $z_1$ ) $\wedge$ 持つ<sub>5</sub>( $x_1, x_2, z_1$ )

愛する<sub>2</sub>( $x_1, x_2$ ) $\supset$ 親しむ<sub>1</sub>( $x_1, x_2$ )

仰ぐ<sub>1</sub>( $x_1, y_1$ ) $\supset$ 上( $z_1$ ) $\wedge$ 向く<sub>1</sub>( $x_1, z_1$ )

上る<sub>6</sub>( $x_1, x_2, x_3$ ) $\supset$ 進む<sub>2</sub>( $x_1, x_2, x_3$ )

上る<sub>8</sub>( $x_1, x_2, x_3$ ) $\supset$ 出る<sub>1</sub>( $x_1, x_2, x_3$ )

上る<sub>14</sub>( $x_1, x_2$ ) $\supset$ 済む<sub>2</sub>( $x_1, x_2$ )

呆れる<sub>1</sub>( $x_1, y_1$ ) $\supset$

違和感( $z_1$ ) $\wedge$ 様子( $z_2$ ) $\wedge$ 感じる<sub>5</sub>( $x_1, x_2, z_1$ )

呆れる<sub>2</sub>( $x_1, x_2$ ) $\supset$ 違和感( $z_1$ ) $\wedge$ 感じる<sub>5</sub>( $x_1, x_2, z_1$ )

開く<sub>2</sub>( $x_1, x_2$ ) $\supset$ できる<sub>2</sub>( $x_1, x_2$ )

空く<sub>3</sub>( $x_1, x_2$ ) $\supset$ できる<sub>2</sub>( $x_1, x_2$ )

明ける<sub>1</sub>( $x_1$ ) $\supset$ 終わる<sub>1</sub>( $x_1$ )

空ける<sub>3</sub>( $x_1, x_2$ ) $\supset$ 作る<sub>11</sub>( $x_1, x_2$ )

上げる<sub>1</sub>( $x_1, x_2, x_3, x_4$ ) $\supset$ 移す<sub>1</sub>( $x_1, x_2, x_3, x_4$ )

上げる<sub>7</sub>( $x_1, x_2, x_3$ ) $\supset$ 入れる<sub>1</sub>( $x_1, x_2, x_3$ )

上げる<sub>10</sub>( $x_1, x_2$ ) $\supset$ 出す<sub>6</sub>( $x_1, z_1, x_2$ )

挙げる<sub>12</sub>( $x_1, x_3, x_2$ ) $\supset$ 出す<sub>14</sub>( $x_1, x_2, x_3$ )

挙げる<sub>19</sub>( $x_1, x_2$ ) $\supset$ 行う<sub>1</sub>( $x_1, x_2$ )

挙げる<sub>20</sub>( $x_1$ ) $\supset$ 纏まる<sub>1</sub>( $x_1$ )

上げる<sub>21</sub>( $x_1, x_2$ ) $\supset$ 吐く<sub>1</sub>( $x_1, z_1, x_2$ )

(平成2年2月16日受付)

(平成2年10月9日採録)



富浦 洋一(正会員)

昭和36年生。昭和59年九州大学工学部電子工学科卒業。昭和61年同大学院工学研究科電子工学専攻修士課程修了。平成元年同大学院博士課程単位取得退学。現在、九州大学工学部助手。自然言語処理の研究に従事。



日高 達(正会員)

昭和14年生。昭和40年九州大学工学部電子工学科卒業。昭和42年同大学院工学研究科電子工学専攻修士課程修了。昭和44年同大学院博士課程中退。工学博士。同年九州大学工学部助手。昭和48年同講師。昭和55年同助教授。昭和63年同教授。現在に至る。形式言語の方程式論, 自然言語処理, 手書文字認識の研究を行う。電子情報通信学会会員。



吉田 将(正会員)

昭和8年生。昭和33年九州工業大学電気工学科卒業。昭和35年九州大学大学院工学研究科修士課程修了。工学博士。昭和37年九州大学工学部講師。その後、九州工業大学教授。九州大学工学部教授を経て、昭和61年10月九州工業大学情報工学部長。この間、九州工業大学および九州大学情報処理教育センタ長、九州大学大型計算機センタ長を歴任。自然言語処理の研究に従事。人工知能学会、日本認知科学会、米国 ACL などの会員。