

# HPSG 入門

## 目次

### HPSG 文法入門

- 素性構造文法から HPSG へ
- HPSG 文法の概要
- 文法の拡張

今回は、最終課題で扱う二つの問題について解説する。一つ目は、第5回で扱った素性構造による文法をさらに発展させ、HPSG (Head-driven Phrase Structure Grammar) という枠組みに沿って日本語文法を記述することを目指す。

## 1 HPSG 入門

### 1.1 素性構造文法から HPSG へ

#### 1.1.1 補語素性の導入

第5回の課題5-2で見たように、動詞は前側の格助詞句とのくっつき方に関して、多様な振舞いをする。例えば動詞“住む”と“買う”は、

- きつねが山に住む
- きつねがとんびに油揚げを買う

というように、前者ではが格とに格をとるのに対し、後者ではが格、に格、を格をとっている。

そこで、このような「まわりとのくつつき方」を区別するために、*sign* の中に新たに COMPS<sup>1</sup> という素性を作ることにしよう。そうすれば、第5回と同じ方針で次のように動詞の区別を表現できる。

$$\left[ \begin{array}{l} \textit{sign} \\ \text{POS : 動詞} \\ \text{COMPS : 補語-が格_に格} \end{array} \right] \rightarrow \text{“住む”}$$

$$\left[ \begin{array}{l} \textit{sign} \\ \text{POS : 動詞} \\ \text{COMPS : 補語-が格_に格_を格} \end{array} \right] \rightarrow \text{“買う”}$$

例えば、型「補語-が格\_を格」<sup>2</sup>は、動詞が「が格の格助詞句 (が格)」と「に格の格助詞句 (に格)」を前にとることを示している。それに応じて、文法規則はつぎのようになるだろう (各規則の COMPS 素性に注目して欲しい)。

$$\left[ \begin{array}{l} \textit{sign} \\ \text{POS : 動詞句} \\ \text{CONJU : ①} \\ \text{COMPS : ②} \end{array} \right] \rightarrow \left[ \begin{array}{l} \textit{sign} \\ \text{POS : 動詞} \\ \text{CONJU : ①} \\ \text{COMPS : ②} \end{array} \right]$$

$$\left[ \begin{array}{l} \textit{sign} \\ \text{POS : 動詞句} \\ \text{CONJU : ①} \\ \text{COMPS : 補語-が格_に格} \end{array} \right] \rightarrow \left[ \begin{array}{l} \textit{sign} \\ \text{POS : を格_格助詞句} \\ \text{CONJU : ①} \\ \text{COMPS : 補語-が格_に格_を格} \end{array} \right]$$

$$\left[ \begin{array}{l} \textit{sign} \\ \text{POS : 動詞句} \\ \text{CONJU : ①} \\ \text{COMPS : 補語-が格} \end{array} \right] \rightarrow \left[ \begin{array}{l} \textit{sign} \\ \text{POS : に格_格助詞句} \\ \text{CONJU : ①} \\ \text{COMPS : 補語-が格_に格} \end{array} \right]$$

$$\left[ \begin{array}{l} \textit{sign} \\ \text{POS : 動詞句} \\ \text{CONJU : ①} \\ \text{COMPS : 補語-*empty*} \end{array} \right] \rightarrow \left[ \begin{array}{l} \textit{sign} \\ \text{POS : が格_格助詞句} \\ \text{CONJU : ①} \\ \text{COMPS : 補語-が格} \end{array} \right]$$

さて、この時点で COMPS には「補語-*empty*」と「補語-が格\_に格\_を格」、「補語-が格\_に格」、「補語-が格」という4通りの型が現れており、それぞれに対応する文法規則が必要に

<sup>1</sup>補語 complements より

<sup>2</sup>なお、LiLFeS でこのような'-'を含む型を使う時は、'補語-が格\_を格'のように、'で囲む必要がある。

なっている。しかし、例えば「補語-が格\_に格\_を格」と「補語-が格\_に格」、「補語-が格」は、型名に現れている通り「格助詞句とくつつく」という性質を共有している。この共通性をうまくとらえられれば、文法規則の数を減らすことができそうだ。

そこで「この sign はどんな sign とくつつくか」を、COMPS 素性の中に直接書き込んでしまい、文法規則にはそれを見て sign 同士をくつつけるかどうかを判別させる。そのようにすると、「くつつき方」毎に文法規則を用意する必要がなくなる。例えば、“住む”は POS が「が格\_格助詞句」であるような sign とくつついたあと、POS が「に格\_格助詞句」であるような sign とくつつくので、“住む”に対応する sign は次のように表すことにする。

$$\left[ \begin{array}{l} \textit{sign} \\ \text{POS : 動詞句} \\ \text{COMPS : } \langle \left[ \text{POS : に格_格助詞句} \right], \left[ \text{POS : が格_格助詞句} \right] \rangle \end{array} \right]$$

COMPS 素性のところに  $\langle \rangle$  で示したのは、素性構造のリストの意味である。こうすれば、文法規則二つで動詞の全ての振舞いを記述でき、文法が大幅に簡略化される。

$$\left[ \begin{array}{l} \textit{sign} \\ \text{POS : 動詞句} \\ \text{CONJU : } \boxed{1} \\ \text{COMPS : } \boxed{2} \end{array} \right] \rightarrow \left[ \begin{array}{l} \textit{sign} \\ \text{POS : 動詞} \\ \text{CONJU : } \boxed{1} \\ \text{COMPS : } \boxed{2} \end{array} \right]$$

$$\left[ \begin{array}{l} \textit{sign} \\ \text{POS : 動詞句} \\ \text{COMPS : } \boxed{1} \end{array} \right] \rightarrow \boxed{2} \left[ \begin{array}{l} \textit{sign} \\ \text{POS : 動詞句} \\ \text{COMPS : } \langle \boxed{2} \mid \boxed{1} \rangle \end{array} \right]$$

$\boxed{2}$  はリストの先頭、 $\boxed{1}$  はリストの尾部を表している。

この考え方は例えば格助詞についても適用できる。格助詞は前に名詞句をとるので次のように表し、

$$\left[ \begin{array}{l} \textit{sign} \\ \text{POS : が格_格助詞} \\ \text{COMPS : } \langle \left[ \text{POS : 名詞句} \right] \rangle \end{array} \right]$$

これが前の名詞とくつつくと格助詞句になる、という規則は、次のように書くことができる。

$$\left[ \begin{array}{l} \textit{sign} \\ \text{POS : が格_格助詞句} \\ \text{COMPS : } \boxed{1} \end{array} \right] \rightarrow \boxed{2} \left[ \begin{array}{l} \textit{sign} \\ \text{POS : が格_格助詞} \\ \text{COMPS : } \langle \boxed{2} \mid \boxed{1} \rangle \end{array} \right]$$

### 1.1.2 Head(主辞) の導入と文法規則の分解

さて、上の規則では格助詞句と格助詞が区別されている。CFG ではこの区別によって、「格助詞」と「格助詞が名詞句とくつついてできた格助詞句」を分けていた。しかし、「す

でにくっついたか、くっついていないか」の情報は COMPS 素性に分離したので、POS 素性でこのような区別をする必要はもうない。したがって、格助詞の語彙項目と文法規則は次のように変更してしまおう。

$$\left[ \begin{array}{l} sign \\ POS : \text{が格\_格助詞} \\ COMPS : \text{①} \end{array} \right] \rightarrow \text{②} \left[ \begin{array}{l} sign \\ POS : \text{が格\_格助詞} \\ COMPS : \langle \text{②} \mid \text{①} \rangle \end{array} \right]$$

動詞についても、同様にする<sup>3</sup>。

$$\left[ \begin{array}{l} sign \\ POS : \text{動詞} \\ COMPS : \text{①} \end{array} \right] \rightarrow \text{②} \left[ \begin{array}{l} sign \\ POS : \text{動詞} \\ COMPS : \langle \text{②} \mid \text{①} \rangle \end{array} \right]$$

上の動詞と格助詞の二つの規則を比べれば、どちらも親の POS と右側の子の POS が同じになっているので、次のような二つの条件にまとめ直した方がよいだろう。

句をくっつける規則:

$$\left[ \begin{array}{l} sign \\ COMPS : \text{①} \end{array} \right] \rightarrow \text{②} \left[ \begin{array}{l} sign \\ COMPS : \langle \text{②} \mid \text{①} \rangle \end{array} \right]$$

POS の情報を伝える制約:

$$\left[ \begin{array}{l} sign \\ POS : \text{①} \end{array} \right] \rightarrow \dots \left[ \begin{array}{l} sign \\ POS : \text{①} \end{array} \right]$$

これまで一つの句構造に一つずつ文法規則を書いていたのを、「正しい句構造であるためには、この規則と制約を両方満たす必要がある」というふうに、二つの条件に分解して書いているわけである。

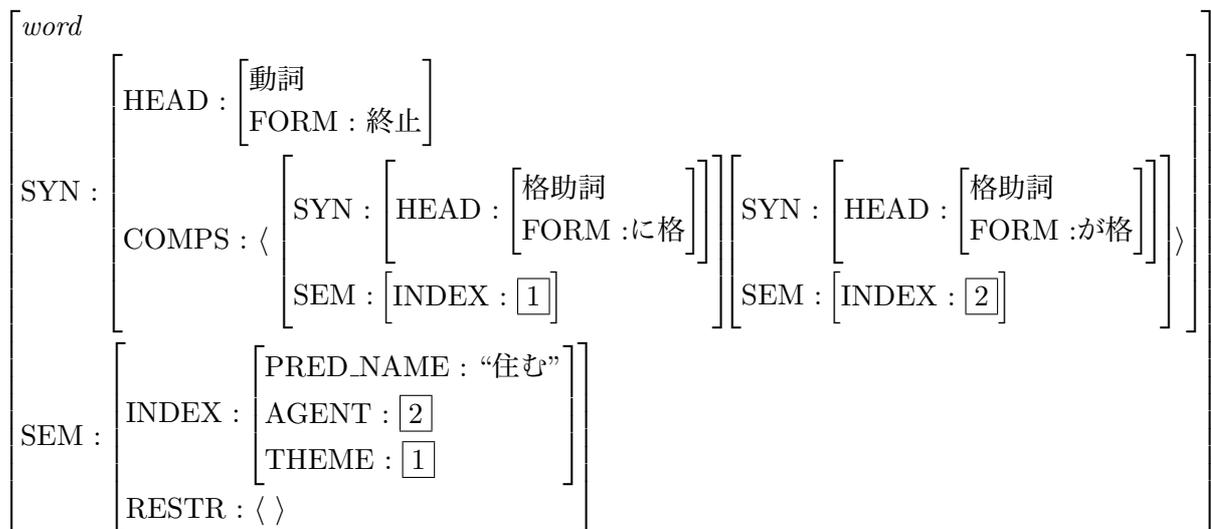
さて、上の文法規則の例では、句ができるときに POS の情報が右側の子によって決められている。このように、一般に句の性質は、その中の一つの単語によって強く支配されていることが多い。そのような単語を、句の主辞 (**head**) と呼ぶ。句が主辞から受け継ぐ性質は、動詞、格助詞、といった品詞の区別だけではない。例えば格助詞句の場合は格の種類は格助詞から受け継がる。また格助詞句と動詞句から動詞句ができるときは、活用形や時制が動詞句から受け継がれる。

従って、「単語が句の主辞になった時、句に提供する情報」を sign の中で HEAD 素性の下を書くように分離し、

$$\left[ \begin{array}{l} sign \\ HEAD : \left[ \begin{array}{l} \text{動詞} \\ \text{CONJU} : \text{終止} \end{array} \right] \end{array} \right] \rightarrow \text{“買う”}$$

$$\left[ \begin{array}{l} sign \\ HEAD : \left[ \begin{array}{l} \text{格助詞} \\ \text{CASE} : \text{が格} \end{array} \right] \end{array} \right] \rightarrow \text{“が”}$$

<sup>3</sup>動詞については「終止形の動詞は動詞句になれる」「未然形の動詞と終止形の助動詞がくっつけば動詞句になる」などの区別があった。これらの区別は 1.3.1 節で扱う。



→ “住む”

図 1: 文法の語彙項目

“POS の情報を伝える” 制約を，次のように書き換えて “主辞の情報を伝える” 制約にすれば，

**head\_feature\_principle**

$$\left[ \begin{array}{l} \textit{sign} \\ \text{HEAD : } \boxed{1} \end{array} \right] \rightarrow \dots \left[ \begin{array}{l} \textit{sign} \\ \text{HEAD : } \boxed{1} \\ \text{(この sign は句の主辞)} \end{array} \right] \dots$$

この制約は動詞句や格助詞句，名詞句など，さまざまなタイプの句構造に適用できる。

このように sign の情報を，独立した役割を果たす素性に分けていき，それぞれの素性に別々の制約をかけることによって，文法規則の数を少なくすることができる。そうすると素性の具体的な値（が格の格助詞，など）は文法規則に書かなくてもよくなり，単語の sign だけに対して指定すれば良くなっていく。

このような，「文法が各単語に対して指定している情報」のことを語彙項目と呼ぶ。ここまで述べたように，文法規則を整理することによって文法規則の数は少なくなっていく，それに応じて語彙項目がより詳細な情報を持つようになる。そのような文法を語彙化文法と呼ぶ。今回は語彙化文法の一つである，HPSG (Head-Driven Phrase Structure Grammar) にもとづいて，日本語の文法を記述することにする。

## 1.2 HPSG 文法の概略

今回は，HPSG にそった日本語文法の簡単な実装をこちらで用意して配布するので，課題ではそれをもとにして，いくつかの拡張をしよう。

まず最初に，今回扱う文法の語彙項目の例を見てもらおう。図 1 は，単語 “住む” に対する語彙項目である。詳しい説明に入る前に，各素性がどのような役割を持っているか，簡単に説明しておこう。

**SYN 素性** この素性の値は、その句/単語の構文的な性質を示している。具体的には HEAD 素性、COMPS 素性と、前節で述べたような制約や規則がくみ合わさって構文木が作られる。

**HEAD 素性** この句/単語が主辞になった時に親の句に与えるような情報をここに書き込む。“住む”は動詞なので、ここでは“動詞”となっている。FORM 素性は活用形を示し、この場合“終止”となっている。

**COMPS 素性** 単語が左側に義務的に要求する補語の sign のリストを指定する。“住む”は左側に「が格」と「に格」の格助詞句をとるので、HEAD\格助詞と指定されている。

**SEM 素性** 単語の持つ意味的情報を示す。“住む”の場合、「が格」でくつつく名詞句が“住む”の「主体」なので AGENT と関連付けられ、「に格」でくつつく名詞句が“住む”の「対象 (場所)」なので THEME と関連付けられている (図の [1], [2] による構造共有)。

### 1.2.1 構文構造

さて、前述の SYN 素性の値によって、どのように構文構造が組み立てられるかを見ていこう。

**Schema** 1.1.2 節では、動詞句と格助詞句がくつついて動詞句に、名詞句と格助詞がくつついて格助詞句になるのを、“句をくつつける”という1つの規則で表現した。HPSG では、このような「くつつけ方」のことを schema (スキーマ) と呼ぶ。

1.1.2 節の“句をくつつける”規則で扱ったのは、「右の子が主辞になっており、主辞の方に指定された制約で、左の子どもが選ばれる」句構造だった。例えば、格助詞“油揚げを”の主辞は“を”で、“を”の COMPS 素性に書かれた制約によって、名詞句“油揚げ”が選ばれている。しかし「句がくつつく」パターンは他にもあるので、それに応じて schema を用意してやる必要がある。

今回配布したプログラムには、2種類の schema の実装を示した。

- head-complement-schema : 右側の主辞が、くつつく相手を選ぶ
- right-head-modifier-schema : 右側の主辞が、くつつく相手に選ばれる

head-complement-schema は補語、つまり動詞句にくつつく格助詞句や格助詞にくつつく名詞句などを扱うための schema である。この schema が、1.1.2 節で使った文法規則である。

この schema を使った“きつねが山に住む”の構文解析結果を、図 2 に示した。(素性の一部を省略してある)

課題で用いる実装にはさらに、right-head-modifier-schema が実装してある。これは形容詞句が名詞句を修飾するときのものである。例えば“おいしい油揚げ”において、主辞は右側の“油揚げ”であるが、これを head-complement-schema で解析しようとする、“油揚げ”の COMPS に形容詞 (この場合“おいしい”) を指定する情報を書く必要がある。そうすると油揚げには、形容詞がかかる語彙項目 (COMPS 素性に形容詞が入っている) とかからない語彙項目 (COMPS 素性は空リスト) の二つを用意する必要がある。しかも、“大きなおいしい油揚げ”などのように、“油揚げ”にさらにいろいろな形容詞がかかる場合を考えれば、“油揚げ”の語彙項目はどんどん増えてしまうことになる。

そこでこの構造に関しては、“おいしい”や“大きな”の方に、「名詞句とくつつく」、という情報を持たせることにする。こうすれば、“油揚げ”には語彙項目を一つ用意するだけ

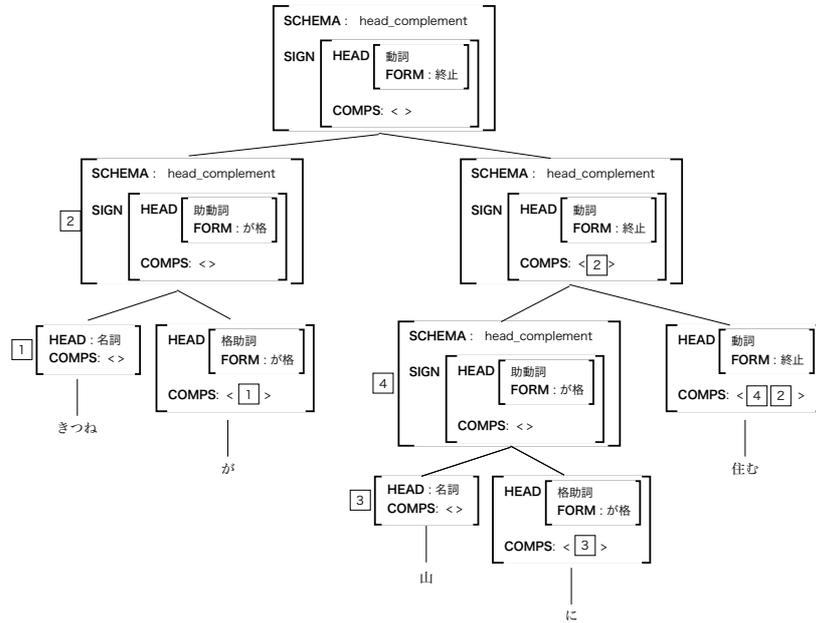


図 2: “きつねが山に住む” の構文解析結果

で、いろいろな形容詞がかかっている場合も解析できる。right-head-modifier-schema は、このような句を解析するためのものである。

**right-head-modifier-schema** 
$$\left[ \begin{array}{l} \textit{sign} \\ \text{COMPS} : \boxed{1} \end{array} \right] \rightarrow \left[ \begin{array}{l} \textit{sign} \\ \text{HEAD} : [\text{MOD} : \boxed{3}] \end{array} \right] \boxed{3} \left[ \begin{array}{l} \textit{sign} \\ \text{COMPS} : \boxed{1} \end{array} \right]$$

これを使った、名詞句 “大きなおいしい油揚げ” の解析を、図 3 に示した。

**Principle** 前節で述べた、「主辞の情報はそのまま親に引き継がれる」という性質は、上で見た 2 種類の句構造に共通しているので、このことは別に head\_feature\_principle という制約を使って表現している。この制約は「主辞の子の SYN\HEAD 素性値は親の句の SYN\HEAD 素性値と一致する」ことを表している。実装上は述語 head\_feature\_principle が、主辞の SYN\HEAD 素性の値と親の句の SYN\HEAD 素性の値と単一化することによって、主辞の情報を親に引き継がせている。

```
head_feature_principle($HEAD, $MOTHER) :-
    $MOTHER = (SYN\HEAD\ $X),
    $HEAD = (SYN\HEAD\ $X).
```

### 1.2.2 意味構造

SEM 素性は、構文構造が組み立てられた文から、文のもつ抽象的な意味を取り出すために使われる。

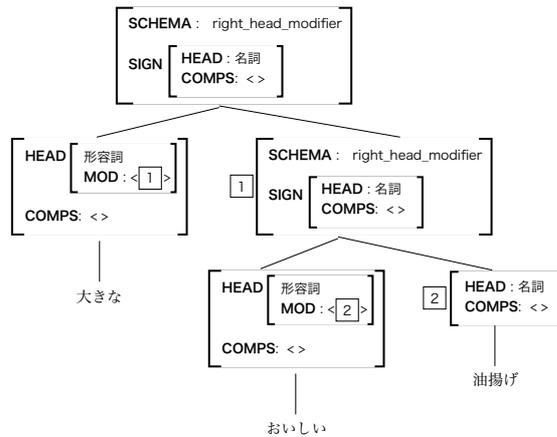


図 3: “大きなおいしい油揚げ” の解析結果

意味は次のような形の素性構造で表している。

$$\left[ \begin{array}{l} \text{INDEX: 意味構造} \\ \text{RESTR: } \langle \text{意味構造}, \dots, \text{意味構造} \rangle \end{array} \right]$$

INDEX 素性は、その語彙項目自身の意味内容を表す。例えば名詞“とんび”が表すのはとんびなので、ここでは INDEX にそのまま“とんび”と書き込むことにする。

動詞の場合もう少し複雑で、例えば取るという動詞の意味構造は、「取るという動作の主体」と「取るという動作の対象(物など)」が“取る”で表されるような関係にある、というように表す。具体的に、“とんびが油揚げを取る”の意味構造は、次のように表される。

$$\left[ \begin{array}{l} \text{PRED\_NAME: “取る”} \\ \text{AGENT: “とんび”} \\ \text{THEME: “油揚げ”} \end{array} \right]$$

この方針で、“おいしい油揚げ”の意味構造を書くと、次のようになるが、

$$\left[ \begin{array}{l} \text{PRED\_NAME: “おいしい”} \\ \text{ARG1: “油揚げ”} \end{array} \right]$$

“とんびがおいしい油揚げを取る”の意味構造を書くときに、取るの THEME 素性にこの意味構造を入れると取るの対象がおいしいという構造になってしまう。

このような時に使われるのが RESTR 素性である。“とんびがおいしい油揚げを取る”の意味構造は全体では次のようになる。

$$\left[ \begin{array}{l} \text{INDEX: } \left[ \begin{array}{l} \text{PRED\_NAME: “取る”} \\ \text{ARG1: “とんび”} \\ \text{ARG2: ① “油揚げ”} \end{array} \right] \\ \text{RESTR: } \langle \left[ \begin{array}{l} \text{PRED\_NAME: “おいしい”} \\ \text{ARG1: ①} \end{array} \right] \rangle \end{array} \right]$$

SEM 素性は semantic\_compositinality\_principle と semantic\_inheritance\_principle という別の principle を使って、子から親へ受け継がれるようになっている。

### 1.3 文法の拡張

#### 1.3.1 助動詞

拡張の一つの例として、“ない”などの助動詞を考えてみよう。第6回までの文法では、助動詞は動詞とくっついて動詞句となり、そのときの親である動詞句の活用形は、助動詞の活用形から受け継がれていた。今の文法では、活用形は主辞から親へ受け継がれるから、動詞と助動詞がくっつくときは助動詞が主辞であるほうが都合がいい。そこで、助動詞を「左側の動詞にくっついて、動詞句にするような動詞」と考えてみよう。

このとき“ない”の語彙項目は次のように定義する。

$$\left[ \begin{array}{l} \text{word} \\ \text{SYN: } \left[ \begin{array}{l} \text{HEAD: } \left[ \begin{array}{l} \text{動詞} \\ \text{FORM: 終止} \end{array} \right] \\ \text{COMPS: } \langle \left[ \begin{array}{l} \text{SYN: } \left[ \begin{array}{l} \text{HEAD: } \left[ \begin{array}{l} \text{verb} \\ \text{FORM: 未然} \end{array} \right] \\ \text{COMPS: ①} \end{array} \right] \mid \text{①} \rangle \\ \text{SEM: } \left[ \text{INDEX: ②} \right] \end{array} \right] \\ \text{SEM: } \left[ \begin{array}{l} \text{INDEX: ②} \\ \text{RESTR: } \langle \left[ \begin{array}{l} \text{PRED\_NAME: “否定”} \\ \text{ARG1: ②} \end{array} \right] \rangle \end{array} \right] \end{array} \right] \end{array} \right]$$

左側の動詞に対する制約は COMPS 素性に書かれている。“ない”の場合、左側の動詞は未然形でなければならなかったので、COMPS 素性の最初の要素の FORM 素性は未然となっている。また、左側の動詞句の補語が親である動詞句の補語になることは、①による構造共有で表されている。また、“ない”は動詞に意味を付け加えるので、意味構造では、左側の動詞の意味(②)に“否定”関係がかかっている、という情報が追加される。この語彙項目と、head-complement-schema を組み合わせると、“きつねが山に住まない”という文に対して第6回と似た形の構文構造が得られる。

なお、ここまでで示した文法規則では、“きつねが山に住まない”に対して、“きつねがやまに住ま”が COMPS となることも許されてしまう(句構造で言うと((きつねが山に住ま)ない)となる)。実際は、head-complement-schema において主辞側の FORM 素性が

終止形になっているかチェックするようにすれば、必ず「住まない」の方が先にくっつき、終止形の動詞句ができる。従って((きつねが山に)(住まない))の形しか許さないようにできる。

### 1.3.2 受け身

受け身の文“油揚げがとんびに取られる”の助動詞“れる”は、「左の動詞にくっついて動詞句となる」という点で、助動詞“ない”と良く似た振舞いをするが、いくつか相違点がある。

- “取ら”の補語は、通常の“取る”と補語の形・語順が違う
- “油揚げがとんびに取られる”の意味は、“とんびが油揚げを取る”と同じになっている。

これらの性質のため、受け身文を扱うためには“取ら”に新たな語彙項目を用意するか、“取ら”と“れる”がくっつくときに COMPS 素性を書き換える必要がある。どちらの場合にも語彙項目をうまく設計することで、“油揚げがとんびに取られる”と“とんびが油揚げを取る”で、解析結果の意味構造が同等になるようにすることができる。

### 1.3.3 格助詞句による名詞の修飾

1.2.1 節で、形容詞が前から名詞を修飾する構造を扱ったが、このような修飾には格助詞を使った物もある。例えば、名詞句にかかる格助詞句“きつねの油揚げ”では、格助詞句“きつねの”が主辞“油揚げ”にかかっている。形容詞を head-complement-schema で解析すると問題があったように、この場合もこの句を head-complement-schema で解析するわけにはいかない。

その際、意味構造としては、

```
[ PRED_NAME : “所有”  
  ARG1 : “きつね”  
  ARG2 : “油揚げ” ]
```

などの構造を、RESTR 素性に追加するようにするのが良いだろう。

### 1.3.4 さまざまな構文

受身のところでは、“油揚げがとんびに取られる”と“とんびが油揚げを取る”が同じ意味を表す、という例を見た。このように、表面上は違う表現が同じような意味を表している、という例は他にもある。

例えば、連体形の動詞・助動詞による修飾を伴う名詞句“きつねが買った油揚げ”では、“きつね”と“油揚げ”の間に

```
[ PRED_NAME : “買う”  
  AGENT : “きつね”  
  THEME : “油揚げ” ]
```

のような関係があると考えられる。

また、例えば“きつねがとんびに油揚げを取らせる”は、直接表面に現れていないが、

$$\left[ \begin{array}{l} \text{PRED\_NAME : “取る”} \\ \text{AGENT : “とんび”} \\ \text{THEME : “油揚げ”} \end{array} \right]$$

という意味内容に言及している。

このようなさまざまな構文についても、HPSGを使って意味的に整合性のある解析ができることが知られている。また日本語以外にも英語ドイツ語など多くの言語が今回扱ったHPSGに似た枠組みで解析されている。