

5C-5

日本語端末間対話解析のための句構造文法

吉本啓 小暮潔

ATR自動翻訳電話研究所

1. はじめに

日本語による端末間対話の翻訳を目的として、解析用句構造文法を記述した。Head-driven Phrase Structure Grammar (HPSG)^[1]を採用し、その日本語版である Japanese Phrase Structure Grammar (JPSG)^[2]の考えに拠っている。日本語の特に話し言葉において重要な述語部分の解析に重点を置いていく。これにより各述語構成要素間の相互承接関係が正しく指定できる。そのための手法は複文構造、およびテンス・アスペクトの解析に応用できる。現状は一文内の解析にとどまるが、文脈解析への入力となりうるよう考慮されている。

2. HPSG

2.1. ユニフィケーション・パーサー

Earleyのアルゴリズムにもとづくユニフィケーション・パーサーを開発した。特徴は、Wroblewskiのアルゴリズムを拡張して、ループをもつ素性構造のユニフィケーションを行う点にある^[3,4]。これにより、より詳細な文法記述が可能になった。

2.2. ユニフィケーション文法の利点

ユニフィケーションにもとづく語彙-統辞論的 (lexico-syntactic) なアプローチを採用したことにより、(1) 統辞論-意味論-運用論 (pragmatics) とした異なる部門に由来する情報の統合を可能にし、複雑な話し言葉の解析が容易になった。(2) 語彙-統辞論的手法はモジュラー性が高いので、新しい語、および各語への新しい情報の追加を容易にする。

2.3. HPSGの概略

HPSGは統辞規則としては句構造文法にもとづきながら、これを簡素化して代わりに個々の語の性質を(主としてSUBCAT素性に記入する)語義記述として詳細に表現する。

3. 文法の概略

3.1. 活用

学校文法に沿いながら、言語形式の差異に分類が対応するよう活用をとらえなおし、素性の組合せによって表現した。活用型(五段活用動詞、一段活用動詞等)はCTYPE(< conjugation type)の値として示される。活用形(未然形、連用形等)はCFORM(< conjugation form)の値として示される。この他、五段動詞の音便や不規則活用もすべて素性で扱う。

3.2. モダリティ素性

日本語の述語は、概略、直後に後接する要素により出現の正否・活用形が規定される。これを利用して、各語のSUBCAT中に、前接語の活用型・活用形・モダリティ素性を指定する。

モダリティ素性はHEAD素性であり、MODL(< modality)の値として示される。例えば、「ラレル」(直接受動)の語義記述は

```
[[HEAD [[POS V][CTYPE VOW][CFORM STEM][MODL [[DEAC PASS]]]]]
 [SUBCAT {[HEAD [[POS P][FORM が][GRF SUBJ]]]
  [SUBCAT {}]
  [SEM ?X]]
 [HEAD [[POS P][FORM に][GRF OBJ]]]
 [SUBCAT {}]
 [SEM ?Y]]
 [HEAD [[POS V][CTYPE (:OR VOW KURU ZURU)]
  [CFORM VONG]
  [MODL [[DEAC -][NEGT -][ASPC -][DONT -]
  [OPTT -][PAST -][EVID -][TENT -]
  [POLT -][INTN -][SFP-1 -]
  [SFP-2 -][SFP-3 -]]]]]
 [SUBCAT {[HEAD [[POS P][FORM が][GRF SUBJ]]]
  [SUBCAT {}]
  [SEM ?Y]]
 [HEAD [[POS P][GRF OBJ]]]
 [SUBCAT {}]
 [SEM ?X]]]
 [SEM ?SEM]]]
 [SEM [[RELN られる-NEUTRAL_PASSIVE]
 [OBJE ?SEM]]]
```

SUBCAT中のVのCTYPEにより、「ラレル」に前接するVは一段・カ変・サ変のどれかの活用型でなければならないこと、CFORMによって未然形1でなければならないことを表す。また、MODLにより前接する語を限定する。例えば、「(テ)アル」のMODL素性は[MODL [[ASPC RESL]]]であり、これは「レル/ラレル」の[MODL [...[ASPC -]...]]とユニファイしないので、「テアル」は「レル/ラレル」に前接しないことを示す。

述語構成語間の承接の順序を反映して、各々の語のもつ、MODL中の素性間にはハイアラキーがある。

```
CAUS > DEAC > NEGT > ASPC > DONT > OPTT > PAST > EVID >
TENT > POLT > INTN > SFP-1 > SFP-2 > SFP-3
```

いま、このハイアラキーを $A_1 > A_2 > \dots > A_n$ とすると、テンプレート $[A_i -][A_{i+1} -] \dots [A_n -]$ に等しいと定義できる。これによって、「ラレル」のSUBCAT中のMODL素性は[MODL !DEAC-]と簡略化できる。

3.3 概略に対する例外

上に対し、(1)終助詞、(2)ラシイ、ソウダ、ヨウダ、ミタイダとタ、(3)ナイ、(4)動+マセンデシタ、は複雑な係り方を示す。特に(1)、(2)、(4)はSUBCAT素性だけでは扱えない。

(例文) a. 行きませんでした
b. *行きませんです
c. 行かなかったです
d. *行かなかったでした

「動詞+マス+ン+デス+タ」から「タ」を取り除くことはできないが、逆に、「動詞+ナイ+タ+デス」の後にさらに

「タ」を付け加えることはできない。このような現象を扱うために、補語に記入されるヘッドの素性に対する制限を示す COH (< constraint on head) 素性を使い、SUBCAT素性と逆に前接する要素に対しどのような要素が後続しうるかを書込む。

aの「デス」(「動詞+マセン」専用に後続する「デス」)は

```
[[HEAD [[POS V][CTYPE DESU][CFORM STEM]
  [MODL [[POLT-AUX +]]]]]
 [COH (:OR [[HEAD [[POS V][MODL [[PAST +]]]]]]
  [[HEAD [[POS V][MODL [[TENT +]]]]]]]]
 [SUBCAT {[[HEAD [[POS V][MODL [[MASU-NEGT +]]]]]]
  [SUBCAT {}]
  [SEM ?SEM]]}]
 [SEM ?SEM]]
```

COHは、パーザーのループの許容により使用可能になる。

4. 述語句取扱いの拡張

以上では動詞文を中心に見たが、形容詞・形容動詞・名詞等の述語構成語もMODL素性にもとづいて取り扱える。

4.1. 接続助詞

接続助詞、及びそれに相当する語もMODL素性によって用法が記述できる。「ナガラ」(継続)は次のように記述できる。

```
[[HEAD [[POS ADV][FORM ながら]
  [COH [[HEAD [[POS V]]]
  [SUBCAT {[[HEAD [[POS P][GRF SUBJ]]]
  [SUBCAT {}]
  [SEM ?X]]}]
  [SEM [[DURT ?S[[PARM ?Y]
  [RESTR [[RELN SIMULTANEOUS]
  [OBJE ?Y]
  [IDEN ?SV_SEM]]]]]]]]]]]
 [SUBCAT {[[HEAD [[POS V]
  [CFORM INFN]
  [MODL [[DEAC (:OR PASS SHON -)][NEGT -]
  [ASPC -][OPTT -][PAST -][EVID -]
  [TENT -][POLT -][INTN -][SFP-1 -]
  [SFP-2 -][SFP-3 -]]]]]]
  [SUBCAT {[[HEAD [[POS P][FORM が][GRF SUBJ]]]
  [SUBCAT {}]
  [SEM ?X]]}]
  [SEM ?SV_SEM]]}]
 [SEM ?S]]
```

上の記述により、「ナガラ」は特定のMODL素性とユニファイする述語構成語にしか後接しない(たとえば、[mod1 [...[past +]...]]となる「タ」には後接しない)こと、「ナガラ」をヘッドとする従属節には主語は含まれないことが示されている。同様に、他の接続助詞がどのような従属節を許容するかもMODL素性の指定により記述できる。

4.2. テンスとアスペクト

「タ」「テイル」の機能の曖昧性を減らすため、サブカテゴリー化する述語構成語のASPE素性(< aspect)-MODL素性を利用する。関係節や従属節の中などの従属節に「タ」があらわれるときは主文に対する相対的なテンス・アスペクトをあらわすが、これはMODL素性とASPE素性との両方により規定する。

この他、形式名詞や否定のスコープ決定等、日本語統辞論の根幹に係わる諸現象にMODL素性を利用する。

5. 運用論

運用論に関する情報も、一文の解析で扱える範囲内で、ス

ピーチ-アクト^[3]や敬語-受給表現^[4]を中心として扱っている。

受給の補助動詞「(テ)モラウ」は下ののように記述される。話し手が「動詞+てもらう」の間接目的語よりも主語に感情移入しているという情報が、SEMとは別のPRAG(< pragmatics)に収められている。PRAGのRESTRSの値は各部分の和を取り、矛盾するもの同士は共存を許さないものとする。

```
[[HEAD [[POS V][CTYPE CONS-UV][SFCONS W][CTYPE STEM]
  [MODL [[DONT BEND]]]]]
 [SUBCAT {[[HEAD [[POS P][FORM が][GRF SUBJ]]]
  [SUBCAT {}]
  [SEM ?X]]
  [[HEAD [[POS P][FORM に][GRF OBJ2]]]
  [SUBCAT {}]
  [SEM ?Y]]
  [[HEAD [[POS ADV][FORM て]
  [MODL [[DEAC PASS][ASPC PROG][DONT BENO]
  [OPTT -][PAST -][EVID -][TENT -]
  [POLT -][INTN -][SFP-1 -][SFP-2 -]
  [SFP-3 -]]]]]]
  [SUBCAT {[[HEAD [[POS P][FORM が][GRF SUBJ]]]
  [SUBCAT {}]
  [SEM ?Y]]}]
  [SEM ?SEM]]}]
 [SEM [[RELN もらう-RECEIVE_FAVOR]
  [AGEN ?X]
  [RECP ?Y]
  [OBJE ?SEM]]]
 [PRAG [[SPEAKER ?SPEAKER]
  [HEARER ?HEARER]
  [RESTR [[RELN EMPATHY-DEGREE]
  [MORE ?X]
  [LESS ?Y]]]]]]]
```

6. 結び

MODL素性を付加する等の句構造文法の拡張によって、日本語文の統辞論・意味論・運用論の一部の基本を形式化した。結果は日英語通信文翻訳システムNADINE (NATURAL Dialogue INTERpretation Expert system)として実現した。運用論的側面、ゼロ代名詞などのテキスト言語学的側面への応用^[5]を現在進めている。

謝辞

本研究の機会を与えてくださったATR自動翻訳電話研究所の樽松明所長、ならびに、同言語処理研究室長の相沢輝明室長に感謝します。

引用文献

- [1] Pollard & Sag, 1987, *Information-based Syntax and Semantics*, vol. 1, CSLI Lecture Notes no. 13.
- [2] Gunji, 1987, *Japanese Phrase Structure Grammar*. Reidel.
- [3] Kogure, et al., 1988, "A Method of analyzing Japanese speech act types" 2nd Int'l Conf. on Theoretical and Methodological Issues in Machine Translation of Natural Languages.
- [4] Maeda, et al., 1988, "Parsing Japanese honorifics in unification-based grammar" The Annual Meeting of ACL.
- [5] Yoshimoto, 1988, "Identifying zero pronouns in Japanese dialogue" COLING'88, Budapest.