

5C-6

自然言語処理のための英語文法

岩山 真, 徳永 健伸, Quek Chee Huei, 田中 穂積
東京工業大学

1 はじめに

自然言語処理システムを構築するためには統語解析が必要である。対象とする分野が広範囲になってくるとシステムが扱う文の種類も増えてくるために、統語解析のための文法規則も増加する。これまでも英語の大規模な文法規則を開発した研究はいくつかあるが[9,4,5]、本稿では、将来的には意味処理をおこなうことも考慮し、統語解析の段階でできるだけ、解析木の数を減らし、意味処理をおこない易くすることを目的に文法を作成した。また、外置を扱うための記述法を備えた文法記述方式を用いることによって文法の規則数の増加を抑えることができた。本稿では、著者らの作成した英語の文法規則の概要と名詞句と動詞句の規則についての説明、および外置を利用した文法規則の例とこれに関する今後の課題について述べる。なお、本文法は、東工大田中研究室で開発された自然言語解析システム LangLAB [10]上で開発した。

2 文法の概要

本稿で述べる英語の文法 EG20 は文法記述形式 XGS で記述されており、172 個の規則よりなる。ただし、これは文法規則の右辺の OR 項を展開しない数である。XGS は外置変形が扱えるように DCG を拡張した文法記述形式で、DCG と上位互換性を持っている。XGS の詳細については [2] を参照されたい。

文法の作成方針として以下のような指針を設けた。

- 主要な文のタイプとして、宣言文、命令文、Yes-No 疑問文、Wh 疑問文の 4 種類を考え、受理できる文の範囲の目標を Robinson の DIAGRAM [9]におく。
- 文脈自由文法の段階でできるだけ冗長な構文木や誤った構文木を排除するように規則を作成する。
- 各文法カテゴリに素性を付与し、文法規則の補強項で素性間のユニフィケーションをおこなうことによって統語的な検査をおこない、さらに不要な構文木を排除する。

文法カテゴリを認定するにあたっては、GPSG [6] や HPSG [8] のように情報を辞書に押し込め、文法規則を積極的に減らすのではなく、基本的な文法カテゴリは残し、それに素性を付加する方法を採った。これは、著者らが Prolog ベースの処理系を仮定しており、すべての統語的検査を素性間のユニフィケーションでおこなうと効率的に問題がある、という判断による。

3 名詞句

名詞句については [1] を参考にして、名詞句を次のように大きく分類した。

<i>all the very ugly big grey</i>	<i>buildings</i>	<i>in New York</i>
限定部	主要部	補部
兩位修飾語句	中心部	後位修飾語句

An English grammar for natural language processing
IWAYAMA Makoto, TOKUNAGA Takenobu,
Quek Chee Huei, TANAKA Hozumi
Tokyo Institute of Technology

まず、限定部に現れる要素には、1.名詞、2.形容詞、3.(序)数詞、4.冠詞類などがあるが、このうち、1.名詞は複合名詞として辞書に登録する、2.形容詞は名詞主要部の一部として解析する、という方法によって解決している。実際に名詞句の限定部として文法規則に現れるのは、3.(序)数詞と4.冠詞類である。これらは、さらに、冠詞、指示詞、数量詞、不定代名詞、数詞、序数詞、所有代名詞などに細分化され、相互の接続関係、順序が素性とカテゴリによって検査される。

補部に現れる要素は、1.名詞句、2.副詞、3.前置詞句、4.Wh 句、5.述部表現、6.節表現などがある。このうち、5.述語表現は、TO 不定詞句、ING 句、EN 句、形容詞句に細分化されている。

4 動詞句

4.1 動詞型の分類

動詞句については、ホーンビーの動詞型分類 [7] を基に分類した。ホーンビーの分類では、細分類も合わせて 53 個の動詞型が存在する。筆者らは、これらを意味構造の抽出、及び素性の導入という観点から 22 個の動詞型規則、3 個のメタ規則に再分類した。筆者らは、必要以上にカテゴリを細かく分けることは避け、素性の集合で細かくカテゴリを区別する方針を取っている。動詞句の場合も、意味的に同一であると判断できる型については、1つの規則に対応させ、実際の下位範疇化の制限は、素性のユニフィケーションでおこなう。

また、目的用法などの不定詞、副詞句、前置詞句の 3 つは、すべての動詞型に付加修飾可能なものとしてメタな規則で記述する。ホーンビーの分類では、これらが修飾する動詞型を別に定義している。ところが、たとえば、場所を示す副詞句の動詞への修飾の可能性などは、むしろ動詞の意味的な情報に依存すると考えた方がより柔軟な解釈が可能になると考えられる。そこで、EG20 では、統語解析の段階ではこれらの修飾句は、全ての動詞を修飾可能なものと解釈し、意味処理で修飾関係の正当性の検査やかかり受け関係の曖昧性の解消などをおこなうことにする。

4.2 助動詞

助動詞(can, must, would, should など)を含む文は、次のような規則で処理できる。

sdec --> np, aux, vp.

助動詞の承接にも様々なものがあり、それぞれについて、主語との対応関係の検査、動詞句との対応関係の検査をおこなう必要がある。また、文の型、不定詞句、動名詞句ごとに同じ記述を強いられる。したがって、すべての助動詞についてこのような記述をおこなうと、規則数が多くなってしまふ。そこで、EG20 では、動詞句の中で助動詞を処理し、助動詞情報は、素性として動詞句に加えることにした。

sdec --> np, vp.

vp --> modalp, tensed_v.

tensed_v --> havep, be, v_comp.

また、このように規則を分解して記述することによって、必要な統語的検査を対応した規則中でおこなうことができるので、

“might have been”などのような複雑な助動詞の組み合わせの扱いも容易になる。

5 外置を用いた文法規則

筆者らは、英文法を記述するにあたって、DCGの拡張であるXGSという形式を用いている。XGS形式では“./”（スラッシュ）の後のカテゴリをスラッシュカテゴリと呼んでいる。スラッシュを用いて文法規則を記述すると、スラッシュの前のカテゴリの下位範疇としてスラッシュカテゴリが欠けていることを解析中にパーザが自動的に発見し対応付ける。本節では、XGS形式を用いた文法記述の例、および、今後の課題について述べる。

5.1 受動文

スラッシュカテゴリを用いて記述した受動文の規則を以下に示す。

```
sd(Vp_A,_) -->
  np1(Np1_A,_) , vp_pass(Vp_A,_) ./ np1(Np1_A,_) .
```

この規則は受動文の表層上の主語(np1)が受動化された動詞句(vp_pass)の下位範疇として欠けているという記述であるから、受動態の規則は、この規則だけで十分である。

5.2 疑問文

疑問文の中から今回は、Yes-No疑問文、Wh疑問文の代表的な規則について例をあげる。

```
sq(Sd_A,_) -->
  bep(Bep_A,_) , sd(Sd_A,_) ./ bep(Bep_A,_) .
swhq(Sq_A,_) -->
  whnp(Wn_A,_) ,
  sq(Sq_A,_) ./ np([[cse|obj]|Wn_A,_) .
```

Yes-No疑問文(sq)では、be動詞(bep)が文頭に移動していることを示している。また、Wh疑問文(swhq)では、疑問詞(whnp)に対応した名詞句(np)が、目的語として疑問文(sq)の下で欠けていることを示している。

5.3 関係節

関係節については、[2]では次の規則によって、先行詞(np1)が、関係節(srel)のなかで欠けていると記述している。

```
np --> np1 , <srel ./ np1> .
srel --> <relpro> , sdec .
```

しかし、このような記述では、関係詞の意味が考慮されておらず、関係詞は統語的な指標としてのみ働いているにすぎない。筆者らは、関係詞(relpro)と先行詞(np1)との統語的、意味的対応をとることによって様々な関係節の統一的な処理を実現している。たとえば、

• 関係詞の格

```
[np a girl [ncomp.t [np whose eyes] [srel t are brown]]]
```

この例では、先行詞(girl)とwhoseの対応をとり、先行詞の格を所有格(girl's)に変更し、関係詞を含むnp(girl's eyes)が関係節の中で欠けていると考える。

• 先導の規約(piped piping convention)

先導の規約とは、節頭に取り出される構成要素が単に関係節中の先行詞と同一の名詞句だけでなく、その名詞句を支

配している上位の名詞句をも移動させることが出来るという現象である [3]。

```
[np1 books [ncomp.t [np2 the covers of which] [srel I designed t]]]
```

この例では、先行詞(books)とwhichの対応をとり、さらにその関係詞を支配している上位の名詞句(np2)が関係節の中で欠けていると考える。

この他にも、関係節で欠けるカテゴリがnpでない場合がある。また、意味構造の抽出を考えると、関係詞と先行詞との意味的なチェックも必要である(whoとwhichの違いなど)。

5.4 今後の課題

痕跡サーチの応用としてその他に、長距離依存関係、並列句におけるギャッピングの現象、省略の現象などがある。EG20でもこれらの現象に部分的には対処しているが、すべてが説明できるわけではない。XGSの痕跡サーチ機能は非常に強力なため、これまでのように、すべての現象を同じ痕跡移動メカニズムで扱おうとすると、非文まで受理してしまうという問題も生じる。今後は、様々な例外的現象も考慮し、更に精密な痕跡サーチ機能を実現していくことが必要であろう。

6 おわりに

本稿では、比較的広範囲な英文を解析できる文法について、その概要を述べた。5節でも述べたように、今後は外置のメカニズムを用いて様々な英語の言語現象を扱えるように文法規則を精密化するとともに、意味処理を付加し、自然言語の応用システムを開発する予定である。

参考文献

- [1] 井上和子, 山田洋, 河野武, 成田一. 名詞. 現代の英文法, 研究社, 1985.
- [2] 今野聡, 田中穂積. 左外置を考慮したボトムアップ構文解析. コンピュータソフトウェア, 3(2):115-125, 1986.
- [3] 大塚高信, 中島文雄. 新英語学辞典. 研究社, 1982.
- [4] 田中穂積, 高倉伸, 今野聡. ボトムアップ構文解析システム BUP 上での英語文法開発と BUP の評価. , *Logic Programming Conference '84*, 12-2, ICOT, 1984.
- [5] T. Briscoe, C. Grover, B. Boguraev, and J. Carroll. A formalism and environment for the development of a large grammar of English. In *IJCAI-87*, pages 703-708, 1987.
- [6] G. Gazdar and A. F. Pullum. *Generalized Phrase Structure Grammar: A Theoretical Synopsis*. Indiana University Linguistics Club, 1982.
- [7] A. S. Hornby. *Guide to Patterns and Usage in English, 2nd Ed.* Oxford University Press, 1975.
- [8] C. Pollard. *Generalized Phrase Structure Grammars, Head Grammars, and Natural Language*. PhD thesis, Stanford University, 1984.
- [9] J. Robinson. DIAGRAM : A grammar for dialogues. *Communications of the ACM*, 25(1):27-47, 1982.
- [10] T. Tokunaga, M. Iwayama, T. Kamiwaki, and H. Tanaka. LangLAB : A natural language analysis system. In *COLING '88*, 1988.