

ATNと格解析を融合した英文名詞句解析

飯田仁 小倉健太郎 野村浩郷

(日本電信電話公社 武蔵野電気通信研究所)

1. はじめに

機械翻訳システム **LUTE** (Language Understander, Translator & Editor) の英日翻訳版 (LUTE-EJ) における英語文の解析方式 (プロトタイプ) の名詞句解析の概要を報告する。

Winograd^[1] は英文パーサを “補強句構造文法システム”, “ATN-based システム” などに分類している。今日, 前者のパーサが種々作られているが, DIAGRAM, LSP^[2] などを除いて自前の文法作成が必要である。さらに具体的には, 名詞句中の動詞句による修飾句などの文法記述が準文 (quasi-sentence) 的であり, これを埋込文として再帰的に捉える点で工夫を要する。また後者のパーサでも同様なことが言える。これに対し, LUTE-EJ のパーサでは, 名詞句部分を ATN-based な解析が受持ち, その途中で出現する埋込表現を文レベルの解析をする格解析^[3] が受持つ。このような名詞句解析をすることにより, システムのモジュール化と, 文法の再帰的表現が行い易くなっている。

2. 名詞句の構造

名詞句の基本構造を次のように捉える。

$\langle NP \rangle ::= \langle NHD \rangle \{ \langle NP \rangle | \langle NOUN \rangle \} \langle NMP \rangle$

NP: 名詞句

NHD: 前方修飾句 “premodification”

NMP: 後方修飾句 “postmodification”

この構造は一見自然そうに見えるが, 左側より解析を進めるとき, NHD の解析に制限を加えることになる。この点については次節で述べる。しかし, この定式化により名詞句の主要名詞 (head of NP) と NHD, 並びに NMP との関係が捉え易いという利点を持つ。名詞句規則の規模については, 現在作成した ATNG のノード数が 81 個, リンク数が 108 個である。因に,

LUNAR システムに使われた規則^[4]では, 文全体の規則として, 70 ノード, そのうち名詞句に係わるノード数は 41 個である。

2.1. NHD の構造

$\langle NHD \rangle ::= \bigcup \{ \langle ADV_i \rangle \langle ADJ_i | \langle NOUN_i \rangle \}$

$\langle \text{現在分詞} \rangle | \langle \text{動名詞} \rangle | \langle \text{過去分詞} \rangle$

ADV_i: 副詞, ADJ_i: 形容詞

この NHD の構造では, 各種冠詞, 即ち不定冠詞・定冠詞の他に each, other などの形容詞的冠詞を便宜的に ADJ_i とみなしている。そして, ADV_i を ADJ_i の修飾語として ADJ_i という単位で考えれば, 動詞句を除いた NHD の構造を次の右枝分れ木として捉えられる。

$\langle \text{ADJ}_i (\text{ADJ}_{i-1} \dots (\text{ADJ}_1 \text{ NOUN}) \dots) \rangle$

つまり, 当名詞句解析では ADJ_i は “ADJ_{i-1} … NOUN” なる句を修飾するとみなす。これが前述の制限である。しかし, この制限は十分な意味処理導入により弱めることができ, そのための枠組が用意されている。

NHD として動詞句をとる場合は, 動詞が要求する構文的性質 (transitive, intransitive) により動詞の変化形が決まる。しかし, この場合は省略句を含む格解析を行うことにより, NHD の意味処理も同時に行うことになる。この点は次節の NMP でも同じである。

ここで, LSP における NHD に対応する修飾句 “LN (left adjuncts of noun)” をみてる。LSP では文法カテゴリーの詳細化により, ストリング列である LN のパターンを文法規則として次のように記述している。

$\langle LN \rangle ::= \langle TPOS \rangle \langle QPOS \rangle \langle APOS \rangle$

$\langle NSPOS \rangle \langle NPOS \rangle$

ここで各要素は空ストリングも許すので, 例えば定冠詞 the は次のように解析され

る。

```

L N
TPOS---QPOS---APOS---NSPOS---NPOS
L T R ; <LT>|<LT>・<冠詞>
L T ; 空ストリング|<副詞>・<冠詞>
T ; 冠詞
THE
```

<APOS>は名詞、または形容詞をとりNHD
そのものに対応するが、今の例でみる様
に主要名詞と冠詞を含むNHDとの関係の
記述に多数の文法カテゴリーが存在する。
これに対し、主要名詞が確定した時点では、
それを制限するNHDが名詞のもつ性質を
どう限定しているか、その名詞から
すぐに分かることが望ましい。次節で述
べる名詞の構造記述と、ATN-basedな解
析により、この点を実現している。

2.2. NMPの構造

<NMP>::=<前置詞句>|<過去分詞句>|
<現在分詞句>|<不定詞句>|<関係詞節>
|<基数><単位><形容詞>

NMPが動詞句・節より構成されるとき、
その構造はNMPの動詞に従って、主要名
詞を含んだ格構造として捉えられる。従
って、名詞句解析の段階でNMPの処理と
して文レベルの解析である格解析を行い、
NMPの解析を可能とする。後述するよう
に、この処理では主要名詞を保持し、
ATNの“HOLD操作”を拡張した振舞い
をするようにしている。“基数”を含む
NMPの構造は、例えば、

“John six feet tall”

の様な名詞の補語的な句である。これに
類似した構造として、

“John, his brother, -----”

の様な同格用法もあり、

“NOUN, <同格表現>,”

なる句もNMPと同様に扱う。

3. 名詞句解析^{[5],[6]}

英語文の名詞句解析はATN風にトップ
・ダウンで左から右へ行われる。名詞句
解析への入力には各単語の辞書情報のリス

トとcontextである。文解析の入力となっ
た単語列は、辞書引き、形態素解析を行
ない、数・性・人称などが既に与えられ
ている。すなわち、名詞句解析を行う時
は図1に示すようなものが既に得られて
いる。それをここでは辞書情報リストと
呼ぶことにする。contextは主語となるべ
き名詞句を切り出して来るのか、目的語
もしくは補語となるべき名詞句を切り出
して来るのかを示すものであり、代名詞
を処理する時に利用できる。^[7]

解析の途中では、句構造を作りながら、
各入力単語フレームの構成単位であるス
ロットに、それを充すべき内容が決定し
た時点で、その内容を埋め込む。途中で
生成される句構造は、単語フレームのス
ロットを埋める時に、それを参照するこ
とによって利用される。ネットワークは
LISP プログラムに埋め込まれた形で
実現されており、ネットワーク自体をデ
ータとしては持っていない。

解析の結果は可能な名詞句が総て得ら
れる。例えば、

“internal representations for data”

を解析すると二つの名詞句が切り出され
る。一つは“internal representations”であ
り、もう一つは“internal representations
for data”である。そしてそれぞれの結果
は、その名詞句の主要名詞を示すシンボ
ルと、その名詞句として組み込まれな

```

(SETQ INPUT '(INTERNAL REPRESENTATIONS FOR DATA))
                                     ↑ 元の単語列
(SETQ INPUT-INFORMATION (GETDIC INPUT))
                                     ↑ 辞書引き, 形態素解析
(((((*SELF.INTERNAL)(*SELF-GCAT.ADJ)(*NHD)
  (*MODIFIER)(*MODIFYING)(*COORD)(*INSTANCE)))
  ((((*SELF.REPRESENTATION)(*SELF-GCAT.NOUN)
    (*CATEGORY)(*NUMBER.PLURAL)(*NHD)
    (*NMP NIL NIL)(*MODIFYING)(*APPOSITION)
    (*PRE)(*DET)(*COORD)(*INSTANCE)))
    ((((*SELF.FOR)(*SELF-GCAT.PRE)(*INSTANCE)
      (*OBJECT)(*COORD)(*MODIFYING))
      ((*SELF.FOR)(*SELF-GCAT.SUBORD)(*INSTANCE)))
      ((((*SELF.DATA)(*SELF-GCAT.NOUN)(*CATEGORY)
        (*NUMBER)(*NHD)(*NMP NIL NIL)(*MODIFYING)(*DET)
        (*PRE)(*COORD)(*APPOSITION)(*INSTANCE))))
```

図1 辞書引き、形態素解析後の
入力単語列の辞書情報リスト

った部分の辞書情報リストとして出力される。一つの名詞句の解析結果は主要名詞のフレームのスロットに埋め込まれているポインタをたどることによって利用される。一つの名詞句の各単語ごとに、それに対応するシンボルを生成し、そのシンボルの値としてその単語の一つの意味に対応するフレームを持つ。

また、名詞句解析で無駄な多岐を生じないようにするため、delimitationに関する知識や言語学的知識を利用した発見的手法を導入している。

3.1. 名詞句解析プログラム

本名詞句解析用プログラムは、大きく三つに分類することができる。すなわち

- (1) 名詞句解析の主要部分を担当するプログラム
- (2) 動詞句解析を呼び出すプログラム
- (3) 等位接続詞に関連するプログラムに分類することができる。

3.1.1. 名詞句解析の主要部分を担当するプログラム

名詞句解析の主要部分を担当するプログラムは次の5つのプログラムである。

- (a) NP: 名詞句解析の実質的なトップ関数。普通の名詞を処理するSUBNPを呼び出す部分と、動名詞、名詞節、代名詞など特殊な名詞を処理する部分から成る。
- (b) SUBNP: 動名詞、名詞節、代名詞などでない普通の名詞句を処理する。
- (c) NHD: NHD部分を処理する。
- (d) NMP: NMP部分を処理する。
- (e) PREPH: 後から名詞を修飾する前置詞句を処理する。

3.1.2. 動詞句解析を呼び出すプログラム

動詞句・節に対応するシンボル（例えば過去分詞句に対してはVPP）をcontextとして持って、さらに動詞句・節などにより後から修飾される名詞（関係代名詞の場合は先行詞）をグローバルな変数として引き渡して、格解析による動詞句解

析プログラムを起動する。これにより、ATNにおけるHOLD操作に対応する処理を行っている。そして動詞句解析の結果を受けて、名詞と動詞句との修飾関係を各単語フレームのスロットに埋め込む。

3.1.3. 等位接続詞に関連するプログラム

等位接続詞に関するプログラムは、等位接続詞で結び付けられているものが、名詞句、名詞句のNHD、名詞句のNMPであるかによって3通りある。

解析の結果、等位接続詞がどのようなものを結び付けているのかを示すために、*obj-cat スロットというものを等位接続詞のフレームに設けている。

- (a) 名詞句の場合: NP
- (b) 名詞句のNHDの場合: NHD
- (c) 名詞句のNMPの場合: NMP

を*obj-cat スロットに埋め込む。文と文を結び付ける等位接続詞の場合にはNILが埋め込まれている。

また等位接続詞が何と何を結び付けているかは、等位接続詞フレームの*object スロットにシンボルのリストとして示す。

例 (*OBJECT ADJ*1 ADJ*2)

これは等位接続詞がADJ*1とADJ*2を結び付けるものであることを示す。

3.2. 名詞句解析で抽出される構造

単語フレームのスロットには、辞書引きの過程で埋めるスロットと、解析の過程で埋めるスロットがある。

辞書引きの過程で埋めるスロットとしては、

- *SELF : 原形（上位概念）
- *SELF-GCAT : 品詞
- *CATEGORY : 意味カテゴリ
- *NUMBER : 数
- *CASE-CONJ : 文法的な格

などを挙げるができる。これらは解析の過程で利用される。図1では、*self と *self-gcat が総て埋まっており、名詞の“representations”の*number が plural になっている。“data”の*number が plural

でなく *nil* になっているのは “data” を集合名詞とみなし、単数扱いしているからである。

解析の過程で埋めるスロットは、

*INSTANCE : その名詞句を代表する単語。

通常は名詞句の主要名詞(被修飾名詞)但し等位接続詞が名詞句を結び付ける場合は等位接続詞自身。

*NHD : 前からの形容詞、名詞による被修飾。

“ADJ*1・ADJ*2・NOUN*1”

のようなものは、ADJ*1の*nhdには *nil* を入れて前には修飾部がないことを示し、ADJ*2の*nhdには ADJ*1を入れて、“ADJ*2・NOUN*1”全体はADJ*1により修飾されていることを示す。NOUN*1の*nhdには ADJ*2が入る。

*DET : 冠詞

*MODIFYING : 修飾。自分自身が何を修飾しているかを示す。

*OBJECT : 冠詞の場合、冠詞によって限定される名詞。前置詞の場合、前置詞の目的語。等位接続詞の場合、等位接続詞が結び付けるもの。

*MODIFIER : 形容詞を修飾する副詞。

*PRE : 主要名詞のフレームの*preに名詞句を支配する前置詞を埋め込む。

*NMP : 後からの被修飾。値は修飾の種類と修飾部の中心語のシンボルとのリストである。

例 (VPP VERB*1)

*APPOSITION: 同格の名詞句

```
NP*1 ((INTERNAL . ADJ*1) (REPRESENTATION . NOUN*1))
ADJ*1 ( .... (*MODIFYING . NOUN*1) (*INSTANCE . NOUN*1))
NOUN*1 ( .... (*NHD . ADJ*1) (*INSTANCE . NOUN*1))
NP*3 ((INTERNAL . ADJ*3) (REPRESENTATION . NOUN*2)
      (FOR . PRE*1) (DATA . NOUN*3))
ADJ*3 ( .... (*MODIFYING . NOUN*2) (*INSTANCE . NOUN*2))
NOUN*2 ( .... (*NHD . ADJ*3) (*NMP PREPH PRE*1) (*INSTANCE . NOUN*2))
PRE*1 ( .... (*INSTANCE . NOUN*2) (*OBJECT . NOUN*3) (*MODIFYING . NOUN*2))
NOUN*3 ( .... (*PRE . PRE*1) (*INSTANCE . NOUN*3))
```

図2 “internal representations for data” の解析結果

*OBJ-CAT : 等位接続詞の結び付けるものの種類。(3.1.3.参照)

フレームの中に現れるスロットは品詞ごとに異なる。例えば名詞、形容詞、前置詞は図1に示したようなスロットを持つ。*instanceは総ての品詞にある。また等位接続詞は全ての品詞を結び付け得るので、上で挙げた総てのスロットを持つ。

3.3. 名詞句構造記述

ここでは、

“internal representations for data” の解析結果を例にとって、名詞句構造が具体的にどう表現されているかを示す(図2)。図3にNP*3の解析結果を見易く表現するために矢印を使った形式で示す。

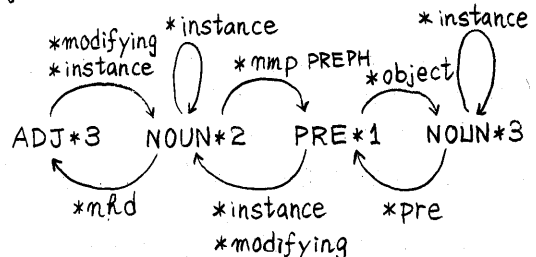


図3 NP*3の解析結果の矢印を使った表現

例えば、NOUN*2の*instanceスロットには自分自身が入っており、それがその名詞句の主要名詞になっていると解釈する。またNOUN*2の*mmpはPRE*1によって代表される前置詞句により後から修飾されていることを意味する。

3.4. 多岐性

名詞句の解析には種々の多岐性がある。そのいくつかを次に挙げる。

(1) 名詞句が連続して続く場合の多岐性
例 “名詞1・名詞2・名詞3”

この場合は二つの多岐性が生じる。一つは、連続したものが一つの名詞となるかいくつかの名詞句に分かれるかという多岐性である。連続したものがいくつかの名詞に分かれる場合は、

(a) 目的語を二つ持つ動詞の二つの目的語に相当する場合

(b) 前置詞句が主語の前に出ている場合

(c) 関係代名詞が省略してある場合などが考えられる。二つめは、修飾の問題である。すなわち右枝分かれの修飾と左枝分かれの修飾の問題が生じる。本解析では総て右枝分かれ修飾として解析する。左枝分かれ修飾は熟語として単語と同様な扱いをする。

(2) 等位接続詞による多岐性

等位接続詞による多岐性としては次の二種類の多岐性を挙げることができる。

(a) 修飾に関する多岐性

(b) 等位接続詞の前で名詞句として切れるかどうかという多岐性

(a) の例としては “形容詞・名詞1・等位接続詞・名詞2” の形をしたものを挙げられる。これは形容詞が名詞1だけを修飾する場合と、形容詞が名詞1と名詞2を修飾する場合がある。このような多岐性は、形容詞以外の前からの名詞修飾(限定)や後からの名詞修飾についても生じる。

(3) 前置詞句の係り先の多岐性

(4) 関係代名詞節による多岐性

(3)(4)については、総ての可能な名詞句を抽出する。

3.5. 解析に導入した heuristics

(1) 名詞連続に関する heuristics

名詞が連続する場合の多岐性を解消するために、名詞の複数形を利用している。具体的には、mfd の解析では名詞を修飾する複数名詞を認めない。さらに、主要名詞としては、次に名詞が来ないか、その名詞が複数である場合しか認めない。

この手法が巧く働かない可能性がある場合として 3.4.(1) の (a) ~ (c) を挙げられる。

(2) 冠詞に関する heuristics

冠詞は一つの名詞につき一回しか現れず、かつ最初の mfd として現れるので、二回目以降の mfd には現れない。本解析ではこのことを利用し、効率を上げている。

(3) context を使った heuristics

名詞句解析が呼ばれる時の context を利用することによって効率の良い解析を行うことができる。これは主に代名詞に対して利用できる。context として主語となる名詞句を要求する時は、当然目的格の代名詞は排除される。逆に目的語や補語となる名詞句を要求する時は、主格のものは排除される。3.4.(2)(b) で挙げた等位接続詞による多岐性も context の利用により部分的に解消できる。例えば、“名詞1・等位接続詞・they ~” の形の時に、context として目的語や補語となる名詞句を要求する時は、they が主格なので、名詞1と they は等位接続詞で結ばれているのではないことが分かり、その結果名詞1だけが名詞句として切り出される。

3.6. 同格

名詞句解析では何種類かの同格を処理している。一つは、

“名詞句1, 名詞句2,”

の形の同格表現を処理する。もう一つは、

“名詞1・固有名詞1”

の形の同格表現を処理する。同格になっていることは、それぞれ名詞句1の主要名詞、名詞1の *apposition スロットに、名詞句2の主要名詞、固有名詞1を入れることにより示す。

4. 解析例

動詞句による修飾を処理した解析結果を図4に、等位接続詞を処理した解析結果を図5に示す。

動詞句を処理する部分では、noun*119 の *nmp が埋められる。図5では、and の *obj-cat に埋め込まれている NHD は、

and が名詞句の nhd 部分を結び付けていることを意味している。

```
(n-par (memory locations by names
        called variables))
```

```
((NOUN*112 ((((*SELF . BY) .... )
(NOUN*114 ((((*SELF . CALL) .... )
(NOUN*118 NIL)))
```

三つの名詞句が切り出される。

```
noun*117 ((*SELF . MEMORY) ....
(*MODIFYING . NOUN*118)
(*INSTANCE . NOUN*118))
```

```
noun*118 ((*SELF . LOCATION) ....
(*NHD . NOUN*117) (*NMP PREPH PRE*21)
(*INSTANCE . NOUN*118))
```

```
pre*21 ((*SELF . BY) ....
(*INSTANCE . NOUN*118)
(*OBJECT . NOUN*119)
(*MODIFYING . NOUN*118))
```

```
noun*119 ((*SELF . NAME) ....
(*NMP VPP VERB*26) (*PRE . PRE*21)
(*INSTANCE . NOUN*119))
```

```
(VERB*26 (*SELF . CALL) (AGENT NIL SUBJ)
(OBJECT NOUN*119 OBJ)
(IDENTIFIER NOUN*120 COMPL))
```

```
noun*120 ((*SELF . VARIABLE) ....
(*INSTANCE . NOUN*120))
```

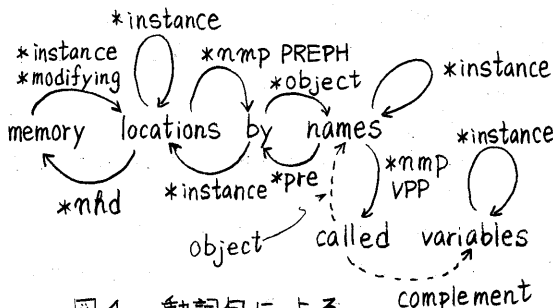


図4 動詞句による修飾を処理した解析結果

```
(n-par (input and output formats))
```

```
((NOUN*1 ((((*SELF . AND) .... )))
(NOUN*4 NIL))
```

```
noun*2 ((*SELF . INPUT) ....
(*INSTANCE . NOUN*4) (*COORD . COORD*1))
```

```
coord*1 ((*SELF . AND) ....
(*INSTANCE . NOUN*4) (*MODIFYING . NOUN*4)
(*OBJECT NOUN*2 NOUN*3)
(*OBJ-CAT . NHD)
```

```
noun*3 ((*SELF . OUTPUT) ....
(*INSTANCE . NOUN*4) (*COORD . COORD*1))
```

```
noun*4 ((*SELF . FORMAT) ....
(*NHD . COORD*1) (*INSTANCE . NOUN*4))
```

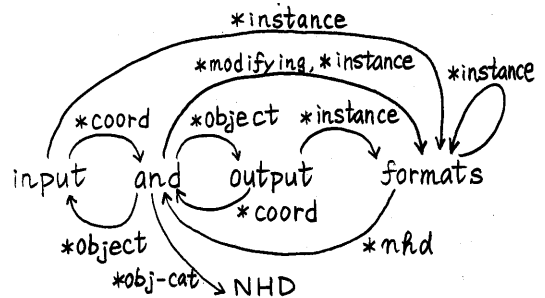


図5 等位接続詞を処理した解析結果

5. まとめ

名詞句の基本構造と文法規則の設定を行い、ATN-basedな解析と格解析を融合して名詞修飾の埋込表現を扱えるようにした。格解析では意味処理を行っているものの、名詞連続などのNHDでは漢字熟語と同種の問題があり、その意味処理は難しい。意味処理に備えた名詞句構造記述を示したが、その処理については別の機会に報告する。

謝辞 日頃御指導戴く基礎部畔柳部長、橋本統括役、第一研究室山下室長に感謝する。また、テキスト分析等を進めてくれた三上雅子社員に深謝する。

文献

1. Winograd, T., "Language as a Cognitive process, Vol. 1", Addison, 1982
2. Sager, N., "Natural Language Inf. Process.", Addison, 1981
3. 飯田・野村, "英文構造の分析と格フレームの構成", 信学技報 Vol.82, No.23, AL82-10, 1982
4. Bates, M., "The theory and practice of ATNG", in Lecture Notes in C.P., Springer, 1978
5. 飯田・野村, "格フレームを中心とした英文の解析モデル", 情処学会, 自然言語処理研究会資料27, 1981
6. 飯田・野村, "機械翻訳に向けた英語の名詞句処理", 第23回情処学会全国大会予稿, 1981
7. 小倉・飯田・野村, "英日翻訳における英文解析の照応関係処理", 情処学会, 自然言語処理研究会資料28, 1981