

Vises Vorasucha, Hozumi Tanaka, (東京工業大学 情報工学科)

Vises Vorasucha, Hozumi Tanaka, (Tokyo Inst. of Tech., Dep. of CS)

Head Grammar (HG) に基づいて、タイ語の解析を行う場合の問題とその解決法について述べる。タイ語独特な問題を解決するため、いくつかのタイ語特有の素性やルールなどを導入し、タイ語の HG 風、すなわち Feature-Unification をベースとした解析方法を提案した。

1. まえがき

1979 年ごろに Gazdar が発表した Generalized Phrase Structure Grammar (GPSG) [1] の拡張である Head Grammar (HG) [5] が Pollard によって 1985 年に提唱されて以来、その独特の素性やそれに伴う原則により、一般的かつ抽象的な枠組みで言語現象を扱うことができるようになった。HG は語順の任意性などにみられる日本語特有の言語現象を扱うためにさらに拡張され、日本語の句構造文法 (Japanese Phrase Structure Grammar: JPSG) [2][4] が ICOT を中心に研究されている。

一方、タイ語は基本的に語順が主語、動詞、目的語の順になっていて、英語とほぼ同じであるが、強調の意味で、主語と目的語が入れ替わることも多く、前置詞句は文の前からでも後ろからでもかかることが可能なので、文中の品詞の位置が英語と比べてはっきり決まらないという問題がある

[6]。そのため、CFG ベースの文法記述にすると文法規則数が莫大になると予想される。それを避けるために、本論文では HG の考え方を参考にして、タイ語文法を作成した。本稿で提案する HG 風の解析法を導入したことにより、簡潔なタイ語解析が可能となり、本解析法の有効性が示された。

2. タイ語の解析

実際の解析について説明する前に、解析の前提となっている素性、句構造規則および各規則について簡単に触れる。

統語範疇は素性の集合として表現される。素性には (1) 二値素性、(2) 多値素性および (3) 統語範疇を値とする素性がある。原則として、タイ語の文は主語、動詞、目的

語という語順をとるので、subcat の値は集合ではなく、リストとする。

句構造規則は 2 つある。

$$m \rightarrow h, c. \quad (1)$$

$$m \rightarrow c, h. \quad (2)$$

一般に、タイ語では、head が句の先頭 (句の最も左) にあるので、規則 (1) が基本である。規則 (2) は subcat の要素の数が 1 になった時などに使用される。

規則には Head 素性規則、Foot 素性規則、および Subcat 素性規則がある。

以上のような規則を利用して、タイ語解析を行ってみる。まず、典型的なタイ語の文の解析について述べる。次に、形容詞と副詞が名詞句や動詞句に含まれる場合の解析について述べる。最後に、前置詞句とタイ語の be 動詞の扱いについて述べる。以後、それぞれの解析例を示し、HG を使うことによって簡潔に解析できることを示す。

2. 1 基本文の解析

1 章で述べたようにタイ語は基本的に主語、動詞、目的語の順になっているので、文の解析は英語とまったく同じである。

図 1 は「*ǎu ǎn ɨɔɔ*」(「私はあなたを愛する」という意味) という文の解析例である。ここでは、分かりやすさのために、重要な素性のみ記述している。sem 素性は意味を表すための素性である。「愛する」の subcat の値によって統語範疇の結合ができる。

解析の過程で部分木ができると共に、その部分木に対する意味構造も同時に計算される。ここで、「愛する」の agent が human で、obj が thing でなければならないという意味的な制約も検査しているこ

とに注意されたい。

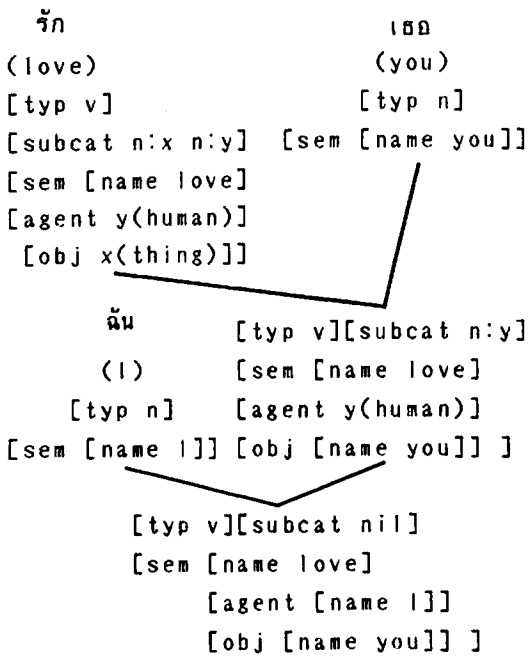


図1 基本文の解析例

2.2 形容詞と副詞

形容詞と副詞は[2]で使われている adjunct という素性を使う。素性 adjunct の値と同じ統語範疇がその adjunct 素性をもつ統語範疇と結合できる。adjunct によって結合された統語範疇の sem 素性は adjunct 素性を持たない統語範疇の sem に adjunct 素性をもつ方の sem の値が加えられるものとする。図2に adjunct 素性を使った例を示す。

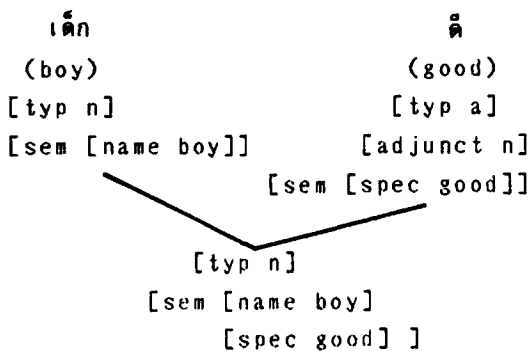


図2 Adjunct 素性を用いた解析例

2.3 名詞句の解析

adjunct を使うことによって、複数の形容詞が連続して名詞にかかっている場合

でも容易に解析できる。しかし、形容詞の種類によって、2個以上連続してはいけないものもある。(例えば、「หนังสือ ๓ เล่ม ๓ เล่ม ...」(本3冊3冊...)の「๓ เล่ม ๓ เล่ม...」)この時には、次のような方法で解決できる。

ここで、cou素性の役割に注目されたい。冊の adjunct 素性中にある cou 素性がマイナスであることは、この「冊」に対する統語範疇が cou 素性の値がマイナスとなっている名詞にしかかかることができないことを表している。一方、「冊」自身はプラスの cou 素性を持っており、Foot 素性規則によって、この素性は root の方に流される。一般的に、名詞の cou の値はデフォルトでマイナスとなっているので、最初の「冊」と名詞との unification はできるが、その時点で名詞句の cou の値がプラスと変わるので、「冊」という統語範疇はもうかかることができなくなる。

ここで、もう1つ注意したいのは adjunct の値が統語範疇のリストになっているということである。

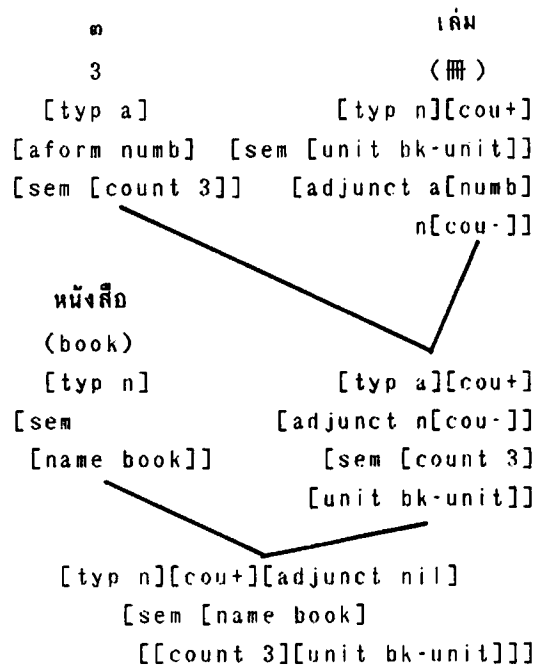


図3 形容詞の解析例

2.4 動詞の解析

タイ語では動詞の活用がないため、時制や態などを表すのにすべて副詞を用いる。特に、目的語の含まれている動詞句には副

詞が目的語の後ろからかかることになっていて、Long distance dependency の問題があるが、HG で扱くと、容易に処理できる。(図4)

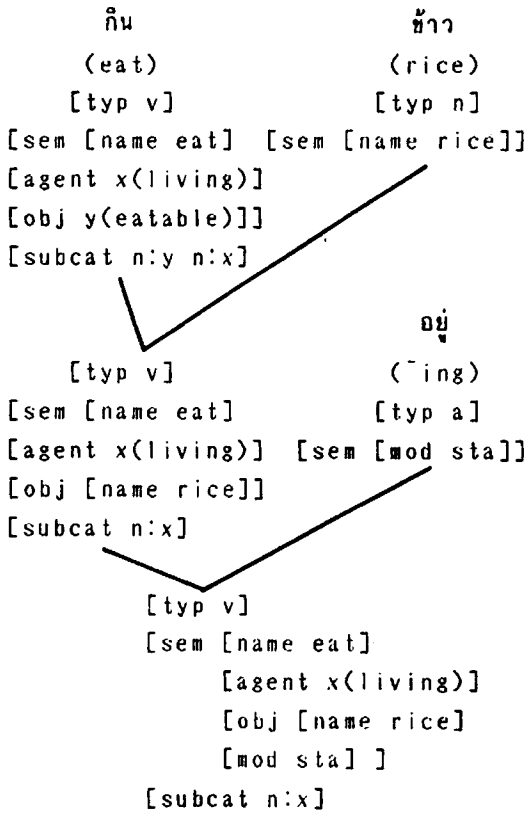


図4 動詞句の解析

タイ語では、否定のための副詞など、副詞によって、前からかかるものもある。これらの副詞に二値素性 pre という素性を付加する。この素性のある統語範疇はほかの統語範疇にかかるときに前からでもかかることが許されるとする。(図5)

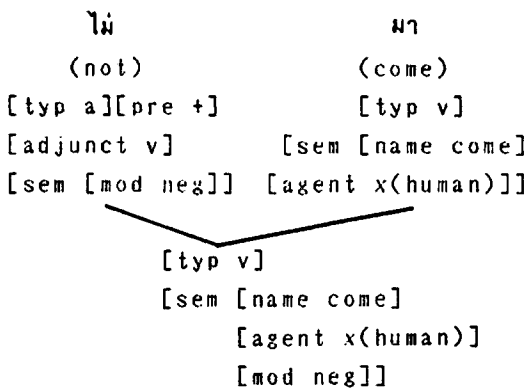


図5 pre 素性の使用例

2.5 Be 動詞の扱い

タイ語では英語の be 動詞にあたる動詞を省略することが多い。したがって、名詞句だけで文の役割を果たすことがある。しかし、全ての名詞句が文の役割を果たせるとは限らないので、文の役割を果たすことができるものだけに pred という二値素性を割り当てることにする。pred 素性のある統語範疇は subcat 素性を持つことが許され、それにより文の役割を果たすことができるようにする。

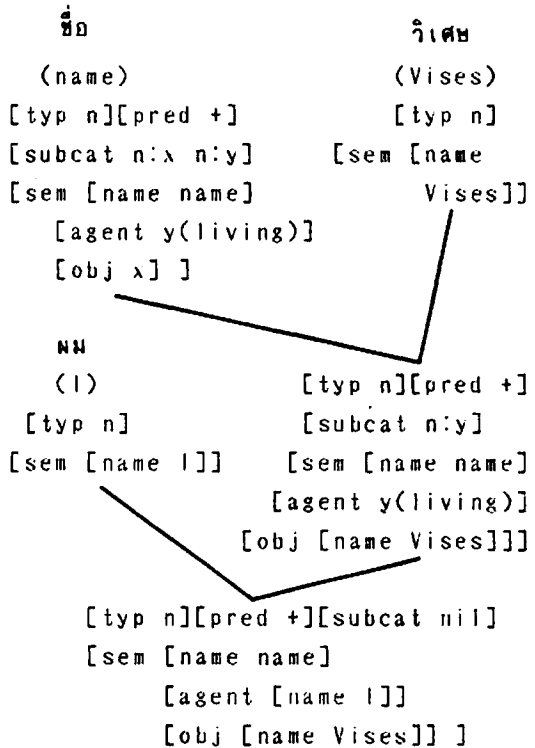


図6 Pred 素性の使用例

2.6 前置詞句の扱い

前にも述べたように、タイ語の前置詞句は文の前からでも後ろからでもかかることができる。そのため pre 素性を利用するが、pre 素性を二値素性から多値素性にかえる。そして、Complement が前からか、後ろからか、それとも双方からかかってくるかによって、pre 素性に +, -, 0 の3つの値を持たせる。(図7)

3. パーサの作成

以上のような原理で簡単なパーサの作成を試みた。作成したパーサは Prolog で書

```

ตอนกลางวัน          คนแน่นมาก
(at noon)          (people very crowd)
[typ p][pre 0]      [typ v]
[adjunct v]         [sem [name crowd]
[sem [mod tim]       [agent [name
[name noon]]         people]]
                        [spec very]]
                        /
[typ v]
[sem [name crowd]
[agent [name people]]
[spec very]
[[mod tim][name noon]]]

```

図7 前置詞句の解析例

かれ， 以下のような特徴がある。

- i) 素性は2つの要素からなるリストで表す。統語範疇は素性のリストリストとする。しかし，統語範疇は素性の集合なので，リスト中の素性の順序は任意とする。
- ii) 解析は BUP 方式で行う。

図8にその解析例を示す。

อัน ไม้ หนัก ใจ เขา.

Processing Time = 0.200001 sec.

No. 1

:-sentence

:-n

:-c

:-อัน

:-h

:-n

:-h

:-n

:-h

:-ไม้

:-o

:-หนักใจ

:-c

:-ใจ

[[typ,v],[sem,[self,ไม้],[agent,อัน],[objd,หนักใจ],

[obji,ใจ]]]

図8 パーサによって得られた解析例

4. おわりに

本研究ではタイ語を Feature-Unification に基づく方法で解析することを試みた。本方法ではタイ語独特の語順の問題を解決するために，独自の素性などを導入し，その結果タイ語の解析が従来に較べ扱い易くなった。

タイ語の語順は日本語ほど緩くないので，

subcat の値をリストにした。しかし，例外（前置詞句，副詞など）も多いので，本論文では，complement が head にかかるときに前から後ろからかを決めるのに pre という素性を利用することにした。これは complement と head の関係を，素性を使って明白に記述することによって，解析の時に生じる complement と head の関係のあいまい性を減少することに大いに役に立つと思われる。

現在，まだ考慮していないタイ語の現象がまだたくさん残されている。例えば，Be 動詞に関しては，名詞と形容詞とからでも文になることができる。したがって，文中に現れてくる名詞と形容詞の並びが名詞句として扱うか，それとも文として扱うか決定的に判断する手段がまだ得られていない。このような現象などをどのようにして我々の枠組みで扱うかが今後の課題である。

参考文献

[1] Gazdar, G., Klein, E., Pullum, G. and Sag, I.: Generalized Phrase Structure Grammar, Oxford, Basil Blackwell, 1985.

[2] Gunji, T.: Japanese Phrase Structure Grammar, Dordrecht, D. Reidel Publishing Company, 1987

[3] Matsumoto, Y., Tanaka, H., Hirakawa, H., Miyoshi, H. and Yasukawa, H.: BUP: A bottom-UP Parser Embedded in Prolog, New Generation Computing, Vol. 1, No. 2 (1983), pp. 145-158.

[4] Miyoshi, H., Gunji, T., Sirai, H., Hasida, K. and Harada, Y.: A Phrase Structure Grammar for Japanese, Computer Software, Japan Society for Software Science and Technology, Vol. 3, No. 4 (1986), pp. 39-45.

[5] Pollard, C.: Generalized Phrase Structure Grammars, Head Grammars, and Natural Language, Ph.D. diss., Stanford Univ., 1984

[6] Vorasucha, V. and Tanaka, H.: Thai Syntax Analysis Using GPSG, Proceedings of the Second Conference of Japan Society for Software and Technology, 1985, pp. 45-48.