

1N-6

LangLABを利用した
翻訳辞書自動作成の試み

高倉 伸(日本アイ・ピー・エム), 田中 穂積(東京工業大学)

1. はじめに

日本文と英文の対応する意味構造を比較することによって、その変換辞書を自動的に生成・更新するシステムを、LangLAB上で試作した。これにより、大規模辞書作成への足掛かりと成り得る成果が得られたので、システム構成および実際の変換辞書の生成例について報告する。

2. システムの構成と概要

本システムは、LangLAB上で構築され、英語・日本語の意味解析プログラムは別に用意されている。構成を図1に示す。

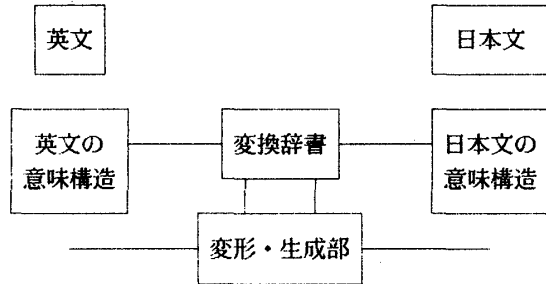


図1 システム構成図

2.1. 意味構造

LangLAB上には、DCKR [小山85] を用いた英語・日本語の意味解析プログラムが既に構築されている。

[奥村86, 高倉86] この意味構造は英語と日本語とで若干異なるが、基本的には、次のような構造をしている。

- [基本単語, (格1: フィラー1, 格2: フィラー2, ...),
- [フィラー1の意味構造],
- [フィラー2の意味構造],
- :
- :
-]

図2 意味構造の形式

2.2. 変換辞書

今回のシステムでは、英語から日本語への変換規則のみを考え、この変換規則の集りを変換辞書と呼ぶ。又、動詞の訳し分けに重点を置いたため、動詞以外の品詞は表面上の訳語の対応表の作成に留まっている。

2.3. 変形・生成部

英語と日本語とで、使用している格の種類が同じであることから、表現内容が同じならば、その意味構造も同じになる。そこで、基本単語同志、フィラー同志の対応を取ることは容易にできる。しかし、これだけでは表層の単語間の置き換えと変りがない。そこで、訳語の選択条件を導き出すことを考える。以下では、単文の意味構造の比較をもとに、比較のアルゴリズムおよび比較して得られる変換辞書について述べる。

2.3.1 比較アルゴリズム

英語から日本語への翻訳を前提に、以下のような手順で変換辞書を自動的に作成する。

- 1) それぞれの意味構造から(基本単語、格、フィラー)の3つからなる対を取り出し、英語、日本語とでそれぞれ esem, jjsem という述語で登録をする。
 - 2) 格が一致するフィラー同志の対応を取る。
 - 3) フィラーの対応がついた基本単語(動詞のみ)に対して訳語の選択条件を記述する。
- 上記の手順を具体例を使って説明する。

- I open the door. (2-1)
- 私がドアを開ける。 (2-2)
- 1) 上の2つの文から得られる esem, jjsem は図3のようになる。
- esem (open vp6ab, tense:gen). (2-3)
- esem (open vp6ab, object:door). (2-4)
- esem (open vp6ab, agent:I). (2-5)
- esem (door, det:the) (2-6)
- jjsem (開ける, agent:私). (2-7)
- jjsem (開ける, object:ドア). (2-8)

図3 意味構造の分解

2) ここで、格同志の対応が取れるのは agent と object だけである。そこで、それぞれのフィラーの対応および基本単語同志の対応を取ると図4のようになる。

- table (I, 私). (2-9)
- table (door, ドア). (2-10)
- table (open vp6ab, 開ける). (2-11)

図4 訳語の対応

- 3) (2-11)節に(2-3)-(2-5)の節を制約条件として付ける

An attempt at automatically generating dictionary for translation
Shin Takakura¹ Hozumi Tanaka² using LangLAB
1. I.B.M Japan 2. Tokyo Institute of Technology

ことができ、(2-12)を得る。

```
table ( open_vp6ab, 開ける ) :- (2-12)
    eesem ( open_vp6ab, tense:gen ),
    eesem ( open_vp6ab, object:door ),
    eesem ( open_vp6ab, agent:l ).
```

したがって、(2-1),(2-2)の入力文に対して、(2-9),(2-10),(2-12)という対応関係が導き出せた。

2.3.2 変換辞書の更新

変換辞書は、次のようなアルゴリズムで更新する。

- 1) ヘッドがマッチする節が変換辞書の中にあれば2)へ、なければ、その変換規則を変換辞書に加えて終了する。
- 2) マッチした節にボディがあれば3)へ、なければ、その変換規則を無視して終了する。(規則の一般化)
- 3) ボディの中の格を比較する。(フィラーの抽象化)

3-1)両方の節に存在する格

この格は必須格であるため、それぞれのフィラーの共通上位概念を探索し、それをその格のフィラーの記述とする。

ア)共通概念が階層構造をなす場合、一番下のものを上位概念とする。

イ)ルート(根)の異なる共通概念が複数個ある場合、それらをすべてORで並べる。

ウ)共通概念がない場合、そのまま2つをORで並べる。

3-2)片方の節のみに存在する格

この格は任意格であるため、trueをORで並べる。

このアルゴリズムを具体例を用いて説明する。今、先程得られた(2-9),(2-10),(2-12)が変換辞書中に存在する状態で、次の入力文を与える。

He opens the window. (2-13)

彼が窓を開ける。 (2-14)

この2文から得られる変換規則は図5のとおりである。

table (he, 彼). (2-15)

table (window, 窓). (2-16)

table (open_vp6ab, 開ける) :- (2-17)

```
    eesem ( open_vp6ab, tense:gen ),
    eesem ( open_vp6ab, object>window ),
    eesem ( open_vp6ab, agent:he ).
```

図5 変換規則例

1) (2-15),(2-16)は、どの変換規則ともマッチしないため、そのまま加える。

2) (2-17)は(2-12)とヘッドがマッチし、ボディがある。3-1)2つの節のボディ中で同じ格の記述を探す。agent格をみるとそのフィラーは、"he"と"l"で、そのどちらでも"open"を「開ける」と訳すことから"he"と"l"の共通上位概念を探索する。イ)"he"と"l"の上位概念として"human"と"pron"の2つが見つかる。同様に、object格のフィラーの抽象化から、(2-18)を得られる。

```
table ( open_vp6ab, 開ける ) :- (2-18)
    eesem ( open_vp6ab, tense:gen ),
```

```
    eesem ( open_vp6ab, object:thingOpen ),
    eesem ( open_vp6ab, agent:human );
    eesem ( open_vp6ab, agent:pron ).
```

3. おわりに

英語と日本語の意味構造から、その変換規則を自動作成することは、ある程度可能であるという結果を得た。しかし、この課題には多くの問題が含まれており、長年にわたる研究が必要である。以下に問題点をいくつか挙げてみる。

(1) 本システムの前提として、英語、日本語の意味解析が可能であることとその格構造が同じであることが必要である。しかし、実際には日本語と英語とで格構造が異なる場合がある。たとえば、"I take a walk."の訳文として「私は、散歩する。」を考える。このとき、「散歩する」を複合サ変動詞として解析してしまうと、「take」の終状態格の"walk"がういてしまう。そこで、「散歩する」を「散歩をする」というように助詞「を」を入れて意味解析をする必要があるなど、構造が異なる場合の処理をいくつか考えなければならない。

(2) 日本語と英語とで動詞の格は同じになっているが、他の品詞に関しては一致していない。そのため、両言語間で整合のとれた意味記述を行なう必要がある。

(3) 複文など文の構造が複雑になった時の考察ができていない。



上位概念が1つで、階層構造をなす場合 ルートの異なる上位概念が複数個存在する場合

図6 上位概念の区別

(4) 変換規則を更新するときに行なうフィラーの抽象化において、上位概念が階層構造をなしている場合と、ルート(根)の異なる共通上位概念が2つ以上ある場合(図6)とを効率良く判定し、この時どちらの上位概念を抽象化した値とするかの判定が難しい。

参考文献

[小山85] 小山,田中,"Definite Clause Knowledge Representation",Proceedings of Logic Programming Conference'85 pp95-106,1985
 [高倉84] 田中,高倉,今野,"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",Proceedings of Logic Programming Conference'84,12-2,1984
 [奥村86] 奥村,"日本語理解システムに関する基礎的研究",東京工業大学理工学研究科修士論文,1986
 [高倉86] 高倉,"英日機械翻訳システムにおける変換辞書の自動作成に関する基礎研究",東京工業大学理工学研究科修士論文,1986