

日英主体表現パターン対を用いた様相表現の日英翻訳

山田 武史

宮崎正弘

新潟大学大学院自然科学研究科

1 はじめに

日英翻訳システムにおいて、よりきめ細やかな日英翻訳を行うには、日本語文を、話者の感情・意思などの主観的表現を直接表現した主体表現と、対象を概念として捉えて表現した客体表現に分離し、客体表現を単文変換部を通して英語文として出力する必要がある。本稿では、このような考えに基づき、日本語単文から分離された主体表現をその属性ごとに主体表現属性パターン化し、これを変換された英文に融合することによって様々な主体表現を生成する方法を提案し、その有効性を示す。

2 主体表現の分類体系

言語表現は、以下のような主体表現と客体表現で構成されている。

主体表現：話者の感情・意思などの主観的表現を直接表現したもの。日本語では、助動詞や助詞など。

客体表現：対象を概念として捉えて表現したもの。日本語では名詞、動詞、形容詞、副詞など。

さらに、主体表現を以下の5つに大分類した。

判断：客体認識に対する話者の直接的な判断を表す。肯定・否定など。

態：主体となんらかの作用との関係を表す。使役・受身など

時制：話者が表現の焦点をどの時間においているかを表す。現在・過去・未来など。

相：動作や状況のあり方を表す。進行・完了など。

様相：客体認識世界に対する話者の主観を表す。推量・要求など。

なお、時枝・三浦文法では、態と相は客体表現に分類されているが、これらは話者に共通するごく少数の属

性のセットで表されることから、主体表現に含めて扱うこととした。

さらに各大分類を中分類・小分類・細分類からなる最大3階層に細かく分類し、ノード数約150の4階層からなる図1に示すような主体表現属性体系を構築した。

様相属性の階層の分類に関しては、同一単文内に同時に存在しない様相属性の排他性を考慮したグルーピングを行った。

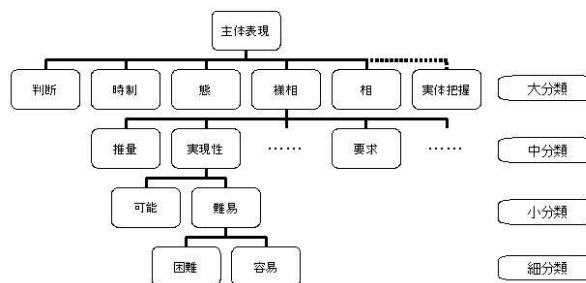


図 1: 主体表現属性体系（一部）

3 主体表現/客体表現分離融合方式

言語表現は主体表現と客体表現が混合された形態をとり、日本語と英語においても、言語族の違いによる認識の違いが反映されているため、表現の対応関係をとるのが困難である。そこで、言語の違いによる話者の見方の違いを取り出し、対象の共通性と言語による見方の違いを分けて処理するためには、まず、主体表現と客体表現を分離し、別々に処理することが有効である。

また日本語は、主体表現に助詞、助動詞などの単語が

用いられるのに対して、英語は助動詞や前置詞などの単語だけでなく語の屈折（言語変化）を伴って主体の直接的表出が行われることが多い。従って、日本語の主体表現の語と英語のそれとは直接的に対応しないことが多い。

一方、客体表現は、それぞれの言語の様々な固有な表現を持ち、その内容も実に多様である。一方、主体的表現は表層的表現は多様であるが、その内容は話者の主観的判断や感情に関する情報であり、極めて限定されている。従って、主体的表現はパターン化して、素性として扱うことが可能であり有効である。

そこで、日本語に表わされた話者の認識を解析するに当たって、話者の主観的感情や意志を分類し、与えられた日本語の主体的表現の部分が、どのような感情や意志を表わしているかを解析する。図2に主体表現解析例を示す。

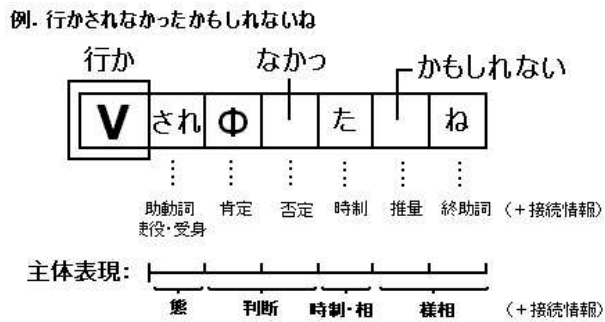


図 2: 日本語文の主体表現解析例

この過程で、もとの日本語は客体表現情報と主体表現情報に分離される。主体表現情報は目的言語の主体表現に訳出できる範囲の分解能と精度で分類すれば良く、このため、その他の情報は客体表現の構造を示す情報として保存し、パターン変換の枠組みを經由して目的言語に変換する。変換対象となる客体表現には、文要素の格を表示する格助詞および格助詞相当語が含まれることになる。この客体表現は変換後、英語の客体表現（英語素文）となる。そのあと、既に抽出されている話者の感情や意志が英語素文に対して組み込まれる。この組み込みでは、助動詞や前置詞の挿入の他、種々の語の屈折（変形）が行われる。このようにして、日本語中から分離された主体表現情報が英文生成の段階で英文に融合される。

図3に主体表現/客体表現分離融合方式による翻訳処理例を示す。

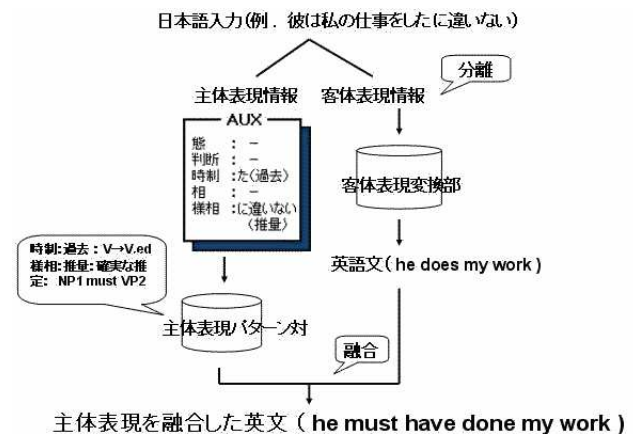


図 3: 主体表現/客体表現分離融合方式による翻訳例

4 英文生成における様相表現の生成

主体表現の属性のうち、本稿では様相表現を融合した英文生成を中心に述べる。

4.1 日英主体表現パターン対

様相表現生成用日英主体表現パターン対の構成を以下に示す。

["中分類属性", "小分類属性", "細分類属性", "英文パターン", "動詞情報", "パターン内動詞"]

- ・中分類：主体表現の分類体系の中分類にあたる属性
- ・小分類：主体表現の分類体系の小分類にあたる属性
- ・細分類：主体表現の分類体系の細分類にあたる属性
- ・英文パターン：主体表現の分類体系の葉にあたる属性に対応する英文パターン
- ・動詞情報：英文パターンによって VP 部に及ぼす動詞の活用情報（形態素調整用）
- ・パターン内動詞：英文パターン内の動詞（形態素調整用）

助動詞パターン 15、助動詞以外のパターン 47 の計 62 のパターン対が作成されている。作成したパターン対の例を以下に示す。

```
$modality_db =
[["実現性", "可能", "-", "NP be able VP",
  "VP.inf", "be"],
["評価", "-", "-", "NP be worth VP",
  "VP.gnd", "be"], ...]
$auxiliary_verb_db =
[["推量", "確実な推定", "-", "NP must VP", "-"],
["実現性", "可能", "-", "NP can VP", "-"], ...]
```

4.2 様相表現生成処理の流れ

処理の流れは、以下の方針をとり、1～6の順に行う。

方針：単文の型「NP AV MO VP」の各要素に合う適当な英文パターンを、入力される主体表現属性情報から生成し、これを型の各要素に代入することによって英文生成を行う。型の各要素の定義を以下に示す。

NP： noun phrase（名詞句）

VP： verb phrase（動詞句）

AV： auxiliary verb（助動詞）

MO： modality phase（様相句）

1. 主体情報、客体情報を入力
 - ・客体情報から品詞をキー、入力英文を値にもつハッシュを生成
 2. 様相表現パターン対と主体表現情報を照合。
 - ・結果から英文パターンと、動詞情報を抽出
 3. 複数の様相情報を一つにまとめる。（形態素調整含む）
 4. ハッシュに、キーを AV または MO とし、その値に生成した助動詞パターンまたは様相表現パターンをしたものをストアする。
 5. 動詞情報に従い、動詞の語尾変換を行い、VP 部を変換。（形態素調整）
 6. 単文の型：「 NP AV MO VP 」の各要素に、生成したハッシュの同名キーの値を代入し英文を生成する。
- 1～6にあたる処理の流れを、図4に示す。

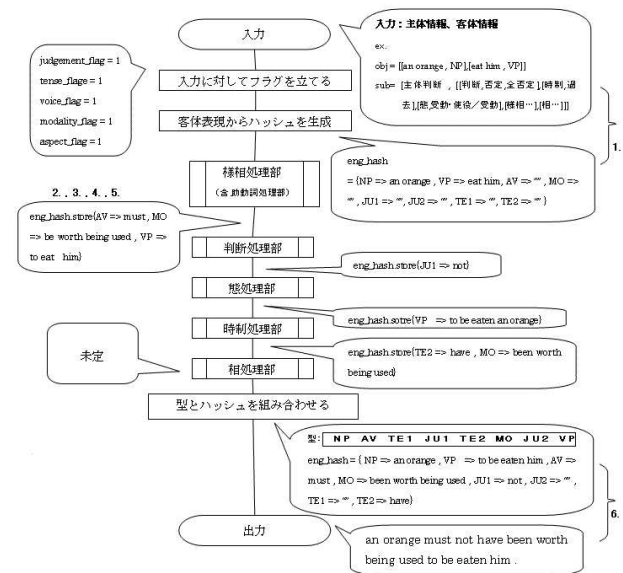


図 4: 英文生成プログラム全体図

4.3 様相表現生成処理の例

○客体翻訳（入力1）

```
["he", "NP"], ["eat an orange", "VP"]]
```

○主体表現情報（入力2）

```
["主体判断", ["判断", "否定", "全否定"],  
["時制", "過去"], ["態", "受動・使役／受動"],  
["様相", ["推量", "確実な推定"], ["評価"],  
["実現結果", "習慣", "習慣（過去）"]]], ["相"]]]
```

1. 客体翻訳からハッシュを作る

```
eng_hash={"NP"=>"he", "VP"=>"eat an orange"}
```

◎様相表現処理部

2-1. 助動詞を捕まえる

```
modality=[["推量", "確実な推定"], ["評価"],  
["実現結果", "習慣", "習慣（過去）"]]
```

↓助動詞 DB

```
av=["NP must VP", "-"]
```

2-2. 助動詞以外の様相表現を捕まえる

```
modality =  
["評価"], ["実現結果", "習慣", "習慣（過去）"]]
```

↓様相 DB

```
["NP be worth VP", "VP.gnd", "be"],  
["NP be used VP", "VP.inf", "be"]]
```

3. 複数の様相情報を一つにまとめる

```
modality_pattern  
= ["NP be worth VP", "NP be used VP"]
```

```
v_info = ["VP.gnd", "VP.inf"]
```

```
v_in_modality = ["be", "be"]
```

```
"being"
```

```
mo=["NP be worth being used VP", "VP.inf", "be"]
```

4. これらを eng_hash に新たにストアする

```
eng_hash={"NP"=>"he", "VP"=>"eat an orange",  
"AV"=>"must", "MO"=>"be worth being used"}
```

5. VP 部を mo[1] の動詞情報に合う形に変換する

```
mo[1]="VP.inf"
```

```
eng_hash["VP"]=>"to eat an orange"
```

◎生成出力処理部

6. 型:「NP AV MO VP」の各要素に、ハッシュの同名キーの値を代入し英文を生成する

```
"NP AV MO VP"
```

+

```
eng_hash={"NP"=>"he", "VP"=>"to eat an orange",  
"AV"=>"must", "MO"=>"be worth being used"}
```

↓

```
he must be worth being used to eat an orange
```

5 おわりに

本稿では、分離された主体表現を日英主体表現パターン化し、これを変換された英文に融合することによって様々な主体表現を生成することを可能とし、その有効性を示した。

また、一連の処理は、ルールベースで行っているため、データの追加・修正が可能である。

今後の課題には以下のようなものがある。

- ・ NP を複数取る形の文や慣用的な表現などの例外処理に対応していない。
- ・ 出力英文の実在性が確認されていないので英文の実在性を検証する必要がある。

現在、日英主体表現パターン対 DB の充実と、それぞれの属性ごとの優先順位、属性の排他関係を考慮したアルゴリズムの議論及び、様相表現以外の判断・態・時制処理部の実装を進めている。

なお、本稿で対象外とした日本語の係助詞・副助詞など客体表現の構成要素となり、実態の捉え方を表す。主体表現（図1の主体表現属性体系において大分類「実体把握」として拡張予定）については、今後検討する必要がある。

参考文献

1. 池原悟, 宮崎正弘, 白井諭, 林良彦: 言語における話者の認識と多段翻訳方式, 情報処理学会論文誌, Vol.28, No.12, pp.1269-1279 (1987)
2. 三浦つとむ: 日本語とはどういう言語か, 講談社学術文庫, (1976)