

## 日本語からアジア諸言語への機械翻訳の試み

今井啓允、謝軍、T.Samantha、  
酒井美和、高木優紀江、E.Nayana、M.Chau、卜朝暉、池田尚志  
岐阜大学工学部

## 1 はじめに

我々は、日本語からアジアの諸言語（現在のところ、中国語、シンハラ語、ベトナム語）への機械翻訳システム jaw(from Japanese to Asian and world languages) の開発を試みている。jaw は日本語文の解析、表現パターン変換辞書との照合による相手言語の表現構造への変換、表現構造からの相手言語の生成といういわゆるトランスファーの方式である。翻訳エンジンは同一で、変換規則、生成関数が相手言語ごとに異なっている。今のところ3つの段階（命題的内容の翻訳、用言後接機能語に関する翻訳処理、体言後接機能語に関する翻訳処理）での翻訳処理を考えている。未だ初期的なパイロットモデルの段階であるが、命題的内容の翻訳、用言後接機能語の翻訳処理方式について報告する。

## 2 機械翻訳システム jaw の概要

jaw は、日本語と目的言語の対応規則を日本語文の係り受け木構造のパターンとそれに対応する目的言語の表現構造の対（表現パターン変換辞書）という形で表現している。

まず日本語入力文を解析して日本語の係り受け構造 IT(Input Tree) を作る。次に IT と表現パターン変換辞書とを照合し、IT を日本語パターンと対応する翻訳規則の木構造に変換する（TT:Transfer Tree）。次に TT 中の翻訳規則に対応づけられた翻訳プログラム（dll として実現している）を実行することで目的言語の表現構造ネットワーク（ET:Expression Tree）を生成する（ET は、VC++ のオブジェクトのネットワークとして実現している）。最後に、各 ET に対応した生成関数（VC++ のメソッドとして実現している）を実行することで目的言語の文を生成する。（図 1）

jaw では、目標言語の表現構造のためのクラスを設計し、クラスメソッドである生成関数（線状化関数）を定めれば、個々の翻訳知識は表現パターン変換辞書中に記述されることになる。また、表現パターン変換辞書との照合による TT の作成、TT に対応づけられ

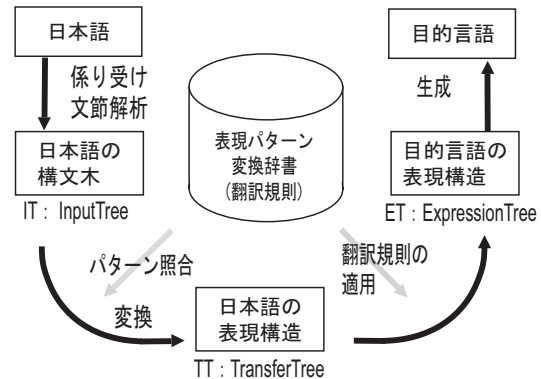


図 1: 機械翻訳システム jaw

た翻訳規則の実行による ET の生成、ET に対応づけられた生成関数の実行による目的言語の生成という方式は、翻訳知識データとその処理を完全に分離しており、任意の言語を目的言語として対象とすることができる。

## 3 日本語の表現パターン

日本語の入力文は本研究室で開発している文節・係り受け解析システム IBUKI を用いて解析し、日本語の構文木 (IT) を得る（ただし現在は文節解析システム ibukiB と接続しているのみで、係り受け解析システムとは未だ接続していない）。IT は文節をノードとする木である。各ノードは自立語と機能語に分割される。例文「彼と付き合ってみると面白い男だった。」を解析したときの IT を図 2 に示す。

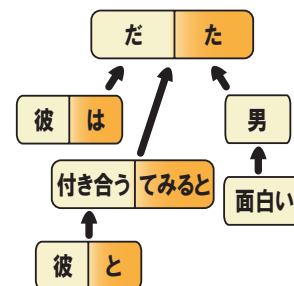


図 2: 日本語の構文木

機能語部分は文節構造解析によって、体言系文節の場合は副助詞前/格助詞/副助詞後/提題助詞に、用言系文節の場合は受身・使役等/時制等/判断等/接続等の要素に分割される。表 1 と表 2 にそれぞれ体言系文節の場合と用言系文節の場合の文節構造解析の例を示す。

表 1：体言系の解析

| 解析前   | 解析後 |      |     |      |      |
|-------|-----|------|-----|------|------|
|       | 自立語 | 副助詞前 | 格助詞 | 副助詞後 | 提題助詞 |
| 私は    | 私   |      |     |      | は    |
| 彼が    | 彼   |      | が   |      |      |
| 彼女だけの | 彼女  | だけ   |     | の    |      |
| 人から   | 人   |      | から  |      |      |
| 東京に   | 東京  |      | に   |      |      |

表 2：用言系の解析

| 解析前       | 解析後  |     |      |     |     |
|-----------|------|-----|------|-----|-----|
|           | 自立語  | 使役等 | 時制等  | 判断等 | 接続等 |
| 歩いた       | 歩く   |     | た    |     |     |
| 食べさせた     | 食べる  | させる | た    |     |     |
| 付き合ってみると  | 付き合う |     | てみる  |     | と   |
| 怒らないでくれと  | 怒る   |     | ない   | でくれ | と   |
| いらねえかったらう | いる   | られる | ない/た | らう  |     |

これらの機能語部情報は、ET 中のメンバー ModeJ の値として書き込んでおく。機能語部分の翻訳方法については 5 節で述べる。

パターン照合に使用する辞書には、表現パターン変換辞書のほかに、単語の意味属性の包含関係を示した意味属性辞書、および機能語をグループ化した機能語条件辞書の 3 つがある。意味属性辞書と機能語条件辞書は表現パターンが持つ自立語および機能語の条件に IT が適合するかどうかを判断するための辞書である。

表現パターンは記述の中心となる語（キーワード）を持つ。キーワードの種類やキーワードを持つ文節に係る文節などの条件から、表現パターンは BaseType、AdditionType の 2 種類に分けられる。表 3 にこれらの例を示す。

BaseType はキーワードが内容語（自立語）であり、それをどのような文節が修飾するかを記述したパターンである。修飾する文節は特定の具体的な文節であったり、変数化されて意味属性で制約された文節であったりする。また、係りの表現においてどのような機能語（格助詞相当語）が使われるかという制約条件（機能語条件）が指定される。BaseType のパターンは、

いわゆる格構造ないし結合価パターンと同等である。意味属性辞書、機能語条件辞書が制約条件の記述に使われる。

使役や受け身の機能語がある場合には BaseType のパターンを適切に書き換える処理を行う。また BaseType の引数が入力文で省略されている場合は、TT 上の省略された場所に省略のマークを置いておく。

表 3：表現パターンの例

| Type  | 文節番号 | 係り先番号 | KW   | 自立語条件 | 機能語条件 |
|-------|------|-------|------|-------|-------|
| Base  | 1    | 3     |      | 人(彼)  | 「が」   |
|       | 2    | 3     |      | 具体(男) | Null  |
|       | 3    | 0     | だ    | Null  | Null  |
| Base  | 1    | 3     |      | 人(私)  | 「が」   |
|       | 2    | 3     |      | 人(彼)  | 「と」   |
|       | 3    | 0     | 付き合う | Null  | Null  |
| Base  | 1    | 0     | 彼    | Null  | Null  |
| Base  | 1    | 0     | 男    | Null  | Null  |
| AddCW | 1    | 2     | 面白い  | Null  | Null  |
|       | 2    | 0     |      | 具体(男) | Null  |
| AddFW | 1    | 2     | の    | 人(私)  | Null  |
|       | 2    | 0     |      | 具体(本) | Null  |
| AddFW | 1    | 2     | てみると | 動作    | Null  |
|       | 2    | 0     |      | 動作    | Null  |

AdditionType は、キーワードが内容語または機能語であり、それがどのような文節を修飾するかを記述したパターンである。AdditionType には Addition-CW と Addition-FW のふたつがある。CW 型は内容語がキーワードである場合、FW 型は機能語がキーワードである場合である。

これらの表現パターンには次節で述べる翻訳規則が対応付けられており、IT と表現パターンとの照合の結果として、対応する翻訳規則の木 (TT) が作られる。「彼と付き合ってみると面白い男だった。」を照合したときの TT を図 3 に示す。

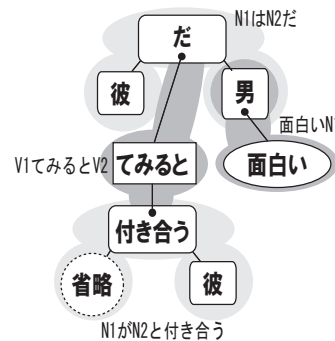


図 3: TT

## 4 表現構造と翻訳規則

表現構造 (ET) は目的言語の文の一種の意味表現である。我々はそのために VC++ のクラスを利用した。名詞、動詞などのクラスを定義し、中心語とそれを修飾する表現など、表現のための部品をクラスメンバーとして定義しておく。また、そのインスタンスが与えられたときに、これらの部品を線形につなぎ合わせて目的言語の言語表現として作り上げるための関数 (生成関数あるいは線状化関数と呼んでいる) と、クラスのメンバー関数 (メソッド) として定義しておく。これらのクラスの設計 (どんなクラスを設けるか、どんなメンバー変数を設けるか、どんなメソッドを与えるか) は、目的言語毎にその文法構造に応じて設計することになる。クラスとそのメンバの例を表 4 に示す。

表 4: クラスとメンバの例

| クラス                    | メンバ            | 説明            |
|------------------------|----------------|---------------|
| CProposition<br>動詞・形容詞 | m_centerW      | クラスを中心語 (訳語)  |
|                        | m_subject      | 主格 (CNoun)    |
|                        | m_object       | 目的格 (CNoun)   |
|                        | m_nounModifier | その他の格 (CNoun) |
|                        | m_pConnection  | 動詞との接続        |
| CNoun<br>名詞            | m_centerW      | クラスを中心語 (訳語)  |
|                        | m_casemarker   | 前置詞           |
|                        | m_role         | 格の役割          |
| CpConnection<br>接続詞等   | m_connection   | 接続詞等の訳語       |
|                        | m_pSubordinate | 接続先の動詞        |
| CModeJ<br>日本語の機能語情報    | m_elem1        | 使役・受身等        |
|                        | m_elem2        | 時制等           |
|                        | m_elem3        | 判断等           |
|                        | m_elem4        | 接続等           |

日本語パターンに対応する目的言語の表現として記述された翻訳規則は、TT から ET を作るプログラムに変換される。これはコンパイルされて翻訳規則のライブラリ (dll) となる。

TT の各要素には、それぞれ翻訳規則関数 (dll) が割り当てられている。これを根ノードからある種の再帰的アルゴリズムで実行することによって、目的言語の表現構造である ET が生成される。

図 4 に入力文「彼と付き合ってみると面白い男だった。」に対する 3 つ言語での ET の例を示す。

## 5 機能語部分の翻訳方法

機能語部分は日本語の文法構造の中核をなす重要な部分であり、日本語からの機械翻訳において、それをどう処理するかは重要なポイントとなる。我々は個々の小単位の機能語ごとに翻訳処理を考えるのではな

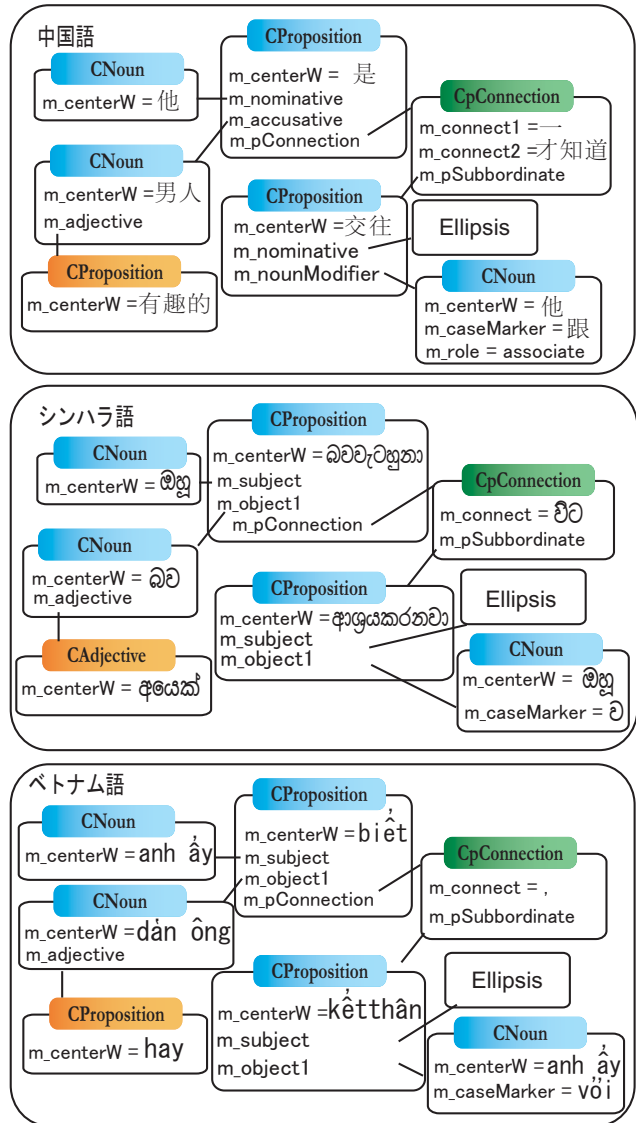


図 4: ET

く、3 節で述べた文節構造の要素ごとに翻訳処理に対応させる方針をとった。これは、我々が定めた文節構造の各要素の異なり数は、対処可能な有限の範囲に収まるということを仮定したものである。機能語部全体をひとつのまとまりとして翻訳処理対象とすることができれば、さらに対処は容易であろうが、それでは異なり数が膨大になり現実的ではない。しかし、4 つの要素に分割して考えれば対処可能であろう、との仮定である。小単位の機能語で考えれば、異なり数はもちろん最小であるが、対処の内容が複雑になり過ぎると考えた。

現在のところ、用言系文節の機能語部分に対応する

翻訳規則についての対処を行っており、機能語部要素ごとに機能語の翻訳規則を定め、各要素の翻訳規則を組み合わせることで機能語の翻訳を実現している。表5, 6, 7は、中国語の場合の各要素の翻訳規則の簡単化した例である。

表5: 「使役・受身等」の翻訳規則

| 機能語     | 文型  | 文型前置詞 | 副詞種類 | 副詞  |
|---------|-----|-------|------|-----|
| られる     | 受身文 | 被     |      |     |
| させる     | 使役式 | 使/让   |      |     |
| させる/られる |     |       | 評注性  | 不得不 |
| がちだ     |     |       | 頻度   | 常常  |
| られる/やすい | 受身式 | 被     | 頻度   | 容易  |

表6: 「時制等」の翻訳規則

| 機能語  | 関数    | 副動詞 | 副詞種類 | 副詞 | aspect | 趨向 |
|------|-------|-----|------|----|--------|----|
| た    | ta    |     |      |    | 了      |    |
| ている  | teiru |     | 時間   | 在  |        |    |
| てくる  |       |     |      |    |        | 来  |
| でもある |       | 是   | 重複   | 也  |        |    |

表7: 「判断等」の翻訳規則

| 機能語     | 助動詞種類 | 助動詞 | 副詞種類 | 副詞 |
|---------|-------|-----|------|----|
| らしい     |       |     | 評注性  | 好像 |
| かもしれ/ない | 可能性   | 可能  |      |    |
| しか/ない   |       |     | 範囲   | 只有 |

各要素の翻訳規則は目的言語側の機能語情報としてメンバー Mode に格納され、生成関数で適切な処理を受ける。

## 6 表現構造からの文の生成

生成された ET の各ノードは、メンバ関数として生成関数(線状化関数)を持っている。生成関数は目的言語の文法構造を反映した目的言語ごとに特有の関数である。中国語とベトナム語は SVO 型であり、シンハラ語は SOV 型である。これを基本として「場所」や「空間」などの修飾関係の並びを考慮しながら語順調整を行っている。

中国語の場合には、述語の修飾要素として表8に示す意味格をクラスメンバーとして定義しており、生成関数では表9に示す文型毎に語順を定めている。

表8: 意味格のクラスメンバ

| 主体 (nominative)  | 場所 (location)   | 範囲 (scope)     |
|------------------|-----------------|----------------|
| 客体 (accusative)  | 原因 (reason)     | 始点 (start)     |
| 与格 (beneficiary) | 道具 (instrument) | 方向 (direction) |
| 時間 (time)        | 根拠 (basis)      | 方式 (manner)    |
|                  |                 | 数量 (quantity)  |

表9: 中国語の文型

| 単述語                           | 多述語         |
|-------------------------------|-------------|
| 基本文、目的語前置文、把字文<br>受身文、存現文、使役式 | 多述語文<br>補語文 |

## 7 中国語、シンハラ語、ベトナム語への翻訳例

jaw による簡単な日本語文の翻訳例を以下に示す。

例文: 彼と付き合ってみると面白い男だった

中国語

试着跟他一交往, 才知道他是有趣的男人。

シンハラ語

ඔහුට ආශ්‍රය කරමින් ඔහු විනෝදකාමී අයෙක් බව වැටහුනා

ベトナム語

Thử kết thân với anh ấy, đã biết anh ấy là người đàn ông hay.

例文: 彼はいつもタバコを吸っている。

中国語

他总是在吸香烟。

シンハラ語

ඔහු නිතරම සීගරුව බොනවා

ベトナム語

Anh ấy lúc nào đang cung hút thuốc lá.

例文: 試合は引き分けに終わっらしい

中国語

比赛好像以平局结束了。

シンハラ語

කරගය දිනුම්පැරදුමකින් තොරව අවසන්වුනා වගෙ

ベトナム語

Trận đấu có vẻ đã kết thúc.

## 8 おわりに

パターン変換型翻訳システム jaw の概要について述べた。第2段階までの方式を設計し実験しているところである。第3段階(体言系文節の機能語である取立て詞の扱い)が次の段階の課題である。

## 参考文献

- [1] パターン変換型機械翻訳システムにおける日本語パターン照合処理に関する研究、河原、岐阜大学工学部応用情報学専攻修士論文 2001 年
- [2] パターン変換型翻訳システムの研究-照合処理に関する研究-、今井、岐阜大学工学部応用情報学卒業論文 2002 年
- [3] パターン変換型翻訳システム jaw について-、今井、謝、池田 FIT (情報科学技術フォーラム) 2002,E-44