

# 構文構造に基づく中日機械翻訳システム

## A Chinese-Japanese Machine Translation System Using Characteristics of Sentence Structures

2P-6

范 莉馨 任 福繼\* 宮永 喜一 柄内 香次  
Lixin Fan Fuji Ren Yoshikazu Miyanaga Koji Tochinai  
北海道大学工学部 C S K 技術開発本部\*

### 1. まえがき

一般に中国語文では複文の頻度が大きい。従って中日機械翻訳システムの性能は複文に対する翻訳能力に支配される。本稿では複文の構造的特徴に着目した中日機械翻訳手法について述べる。

中国語の複文には以下のような特徴がある<sup>1,2)</sup>。

a. 複文とは意味的關係がある二つまた二つ以上の単文が集まったものである。複文を構成する単文の間は意味的關係をもつ単語で繋がる。

b. 構文から見れば、一つの複文は必ず二つ以上の構文要素を持たなければならない。そして、各々の構文要素はまた互いに独立である。

c. 複文を構成する単文は「主述文」でなければ、「無主語文」または「独語文」である。

例文1: 因為我們是為人民服務的<sup>a</sup>, 所以我們如果有欠点和錯誤<sup>b</sup>, 就不怕別人批評指正<sup>c</sup> (私達は人民に仕奉するため<sup>a</sup>, もし欠点と誤りがあったら<sup>b</sup>, 他人の批判や指摘を恐れない<sup>c</sup>。)

上例では単文Aと単文B, Cの間には原因・結果の關係があり, 単文Bと単文Cの間には仮定關係がある。

複文は, 単文と単文間の意味的關係によって9種に, 或は細分すれば17種に分けることができる<sup>1,2)</sup>。また多くの文は多重複文からなる。当然, 複文を構成する単文が多ければ多いほど複文の構造はますます複雑になる。

中国語文は単語間の切れ目のない漢字「べた書き」形式で表記され, 単語を区分できるスペースがなく, 構文要素を提示できる平かなで書かれた助詞, 助動詞などの機能語もない。さらに動詞, 助動詞, 形容詞などの語尾変化するかわり活用現象もない。以上の原因により, 中国語複文に対して構文解析, 翻訳処理を行う際には一般に英, 日文より曖昧性が大きい。

そこで, 以下本稿では中国語複文のもつ前述のような性質と日本語文との対応關係を利用し, 意味的關係を持つ複文の構文解析をいくつか互いに独立な単文の構文解析に還元する手法を提案する<sup>3)</sup>。これにより複雑な構文解析に伴う曖昧性や誤りを避け, 翻訳性能を高めることができる。

本手法により構築する中日機械翻訳システムの概念図を図1に示す。

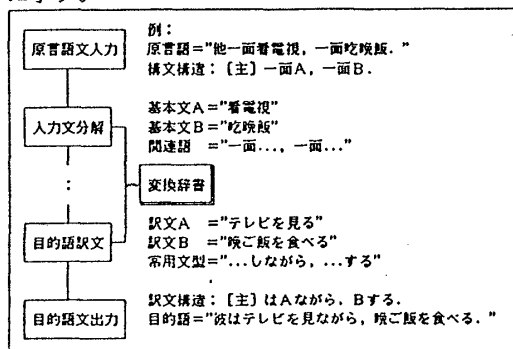


図1 中日機械翻訳システム概念図

まず入力された中国語文(複文)をいくつかの基本文(単文)と関連語に分ける。基本文に対して構文解析, 変換処理などを行い, 関連語に対しては解析を行わず変換処理を直接行う。そしてこれらを日本語の構文規則によって集め, 最終の訳文が得られる。

本システムはつぎに示す三つの部分からなる。

- ① 中国語文の解析及び関連語の抽出
- ② 関連語に対応する日本語表現における多義性, 多義性の解消
- ③ 基本文の処理

以下, この各々について述べる。

### 2. 中日機械翻訳における関連語

本中日機械翻訳システムでは単文を繋ぐための関連語は, その共起關係により一語性と多語性に分けられる。また, 対応する日本語常用文型<sup>4)</sup>の接続法により訳文のアスペクトとモダリティも自動的に満足される。

#### (A) 一語性関連語

[定義]: 複文のただ一個所で出現し, 一回しか出現しない関連語を一語性関連語と言う。

例文2: 他太用功, 所以得了腦病。

例文3: 由於漢字結構複雜, 書写不方便。

上例中の「所以」, 「由於」は一語性関連語であり, その訳文関数<sup>4)</sup>を(1), (2)式に示す。

A, 所以B → A, それでB. .... (1)

由於A, B → Aので, B. .... (2)

上例の訳文は次のとおりである:

訳文2: 彼は勉強をしすぎ, それで腦の病気になる。

訳文3: 漢字は結構が複雑なので, 書くのは便利でない。

#### (B) 多語性関連語

[定義]: 複文中の異なる個所に出現し, 或は二回以上出現している関連語を多語性関連語と言う。

例文4: 這只船不僅能在水上航行, 而且能在雪地航行。

例文5: 他們有時打棒球, 有時打麻将。

上例中の「不僅~, 而且~」, 「有時~, 有時~」は多語性関連語であり, その訳文関数を(3), (4)式に示す。

[主]不僅A, 而且B→[主]Aだけでなく, B.

..... (3)

[主]有時A, 有時B→[主]時にはA, 時にはB.

..... (4)

上例の訳文は次のとおりである:

訳文4: この船は水上を航行できるだけでなく, 雪原を航行できる。

訳文5: 彼は時には野球をし, 時にはマージャンをする。

#### (C) 常用文型により決定される訳文の時態

例文6: 他剛從中国回来。

例文7: 既然他說很忙, 我們就找別人好了。

例文6での「剛~」は一語性関連語であり, 「既然~, 就~」は多語性関連語である。これらの中国語文には時態をはっきり表す助動詞「着, 了, 過」などがないが, これ

らに対応する訳文関数(5),(6)式には特有の日本語接続詞を含むので、その接続関係により動詞の語尾変化を行うと、時態を満足できる日本語訳文を生成できる。

- [主] 剛～ → ～たばかりだ。 …… (5)  
 既然A, [主]就～ → Aたからには[主]は～ …… (6)

(5),(6)式を見て分かるようにその接続方法は「動詞+た」という形である。これらに従って対応する動詞の語尾変化を行うと、つぎに示す訳文が得られる。なお、上述の例文はすべて過去形である。

- 訳文6: 彼は中国から帰ったばかりだ。  
 訳文7: 彼が忙しいと言ったからにはは我々は他の人を頼みましょう。

3. 関連語の多義性および解消

関連語は多義性と多訳性を持っているので、訳文の生成に際して解消しなければならない。本手法では多義性と多訳性を厳密に分ける必要がないので、両方をともに多義性と呼ぶ。我々は大量の例文の分析に基づき、関連語の多義性がその前後共起語の意味属性により解消できるという結論を得た。

例えば、中国語の関連語「才」にはつぎに示す四つの日本語の意味がある。

- ① " やっと "
- ② " ばかり "
- ③ " わずかに "
- ④ " こそ...んだ "

- 例文8: 試了幾回, 才成功。  
 訳文8: 何回か試みて, やっと成功した。  
 例文9: 他才从中国回来的。  
 訳文9: 彼は中国から帰ったばかりだ。  
 例文10: 這個小孩子才六歲。  
 訳文10: この子はわずかに六才だ。  
 例文11: 這才是好榜樣0尼。  
 訳文11: これこそいい手本なんだ。

例文8中の「才」は一定の時間または時間間隔を経て、やっと得た結果を表し、その前共起語は時間また回数を表す単語である。従ってその訳語は「 やっと 」である。

例文9中の「才」はある動作が発生したばかりであることを強調するために用いられ、文中の動詞の後には助動詞「了」また「的」がつく。従って意味は「 ...たばかり 」となる。

例文10中の「才」は数量の少ないことを表すもので、その後共起語は数量を表す単語である。従ってその意味は「 わずかに 」となる。

例文11中の「才」には実際の意味はなく、ただ語気を強めて他人の注意を引き出すために用いられており、文末には語気助詞「0尼」がつく。従ってその意味は「 [主]こそ...んだ 」となる。

以上のように、前後共起語の特徴を利用することにより、関連語の多義性を解消できる。

4. 文の分解と訳文の合成

本システムでは文の分解過程は関連語の抽出及び抽出後の原文を修正する二つの部分からなる。図1中の例文を用いてその流れを説明する。

予め作成されている関連語の情報(表1)により入力文からキーワード1(一面)を検索する。表1から分かるようにその「構成条件」は多語性なので、キーワード2(一

面)を検索する。その結果、もしキーワード2が存在しなければ、キーワード1の「一面」はただ一つの単語として処理する。本例では再び「一面」が見出され、多語性の条件が満たされる。

表1 関連語情報一覧

キーワード	構成条件	日文表現	影響要素	多義性
一面	多語性 (一面)	ながら	2006 (v1連用)	-
才	一語性	-	-	1009
:	:	:	:	:

関連語抽出後に修正を行うと、A、B二つの単文が得られる。この二つの単文を基本文の処理部に入力して翻訳処理を行う。同時に常用文型の接続法により影響された要素v1(見る/看)に対して語尾変化を行う。最後に上述の結果を主部に返送し、日本語文の構文規則によって配列し、訳文が得られる。

5. 基本文の処理過程

本中日機械翻訳システムでは基本文の翻訳処理は①中国語解析, ②中日変換, ③日本語生成という三つの部分からなる。まず最長距離マッチング法により中国語文に解析を行い、その直後に中間処理を行って曖昧性を解消し、ついで文型照合法により中間表現を求め、かつ日本語の中間表現に変換して最終の訳文を生成する。詳細の説明は文献[3, 5, 6]にゆずる。

6. おわりに

本論文では中日両言語間の構文構造上の特徴を利用した中日機械翻訳の一手法を述べた。本手法の要点は、構文解析に際して曖昧性の大きい複文を、互いに独立な単文と関連語に分解し、原言語文の構造によって関連語を分類して、解析の曖昧性をできるだけ早い段階で解消し、全システムの正訳率を高めるというものである。

今後、現在構築中の本手法に基づく翻訳システムにより、実文献を用いて実験を行い、本手法の有効性を検証し、翻訳性能をさらに高める予定である。

参考文献:

- (1) 陳国梁: "現代漢語語法教程", 中国西安市・西安交大出版社(1986.05)。
- (2) 劉月華, 吳文娛他: "实用現代漢語語法", 中国外語教学与研究出版社(1986.10)。
- (3) 范莉馨, 任福繼, 宮永喜一, 柄内香次: "中日機械翻訳における構文構造の特徴に着目した訳文の生成", 電子情報通信学会言語理解とコミュニケーション研究会論文NLC91-30, PRU91-65, pp25-32(1991.10)。
- (4) 任福繼, 范莉馨, 宮永喜一, 柄内香次: "家族モデルを用いた文の分解に基づく日中機械翻訳システム", 情報処理学会論文誌第32巻第10号(1991.10)。
- (5) 范莉馨, 任福繼, 宮永喜一, 柄内香次: "中国語浅層解析の実験システムについて" 電子情報通信学会春季全大論文集D-91, pp.6-91(1990.3)。
- (6) 范莉馨, 任福繼, 宮永喜一, 柄内香次: "中国語文中の複合語の生成について", 電子情報通信学会言語理解とコミュニケーション研究会論文NCL90-1, A190-29, pp.1-8(1990.5)。