

# 機械翻訳システム jaw における連体修飾構造の翻訳

## 一日中、日越を事例として

浅井良信 宇野修一 田中友樹 卜朝暉 Nguyen My Chau 池田尚志

岐阜大学工学部

### 1. はじめに

機械翻訳において、名詞を修飾する連体修飾構造は言語によって違いが大きく、翻訳処理の方法は重要である。

本報告では、我々の研究室で開発している、日本語から各国語への機械翻訳システム jaw における連体修飾構造の処理法について述べる。特に連体埋め込み構造の翻訳処理を中心に、また「A の B」の形の名詞-名詞修飾の翻訳処理について述べる。具体的な事例として、日本語から中国語およびベトナム語への翻訳について述べる。

### 2. 連体修飾構造

日本語では連体埋め込み文は、述語（動詞、形容詞、形容動詞）の連体形を用いた述語部 X と、修飾される名詞 N を並べた X N という形式で一貫している。N が形式名詞（「の」など）の場合も同様である。

中国語では、文全体の構造は SVO という英語と同様の構造で日本語とは異なっているが、連体埋め込みの構造は「N が実質名詞の場合：述語部 X と名詞 N を並べた X 的 N あるいは X N、N が形式名詞の場合：X 的 あるいは単に X」であって日本語とほとんど同じである。ただし、いずれの場合も「的」を使う場合と使わない場合の訳し分けが問題となる。

ベトナム語は中国語と同じ孤立語に属し文全体の形は同じく SVO という形をとるが、連体埋め込みについては中国語とは異なり、英語と同じように先行詞、関係詞を伴う関係節構造をとる場合と関係詞を伴わない単純修飾構造となる場合など、いくつかの表現形式がありその区別が問題となる。

「NのN」の形の名詞-名詞修飾に関してはいつれの言語も「NpN」のように二つの名詞を並べるという点で同様の形をとる。ただし中国語では $p=$ 「的/ $\Phi$ 」であって、「N的N」と表現する場合と直接に「N N」となる場合がありその区別が問題となる。ベトナム語

では英語と同様に場合に応じてさまざまな前置詞 $p$ が区別して用いられる（5個程度ある）。どのような場合にどの前置詞を当てればよいか問題となる。また $p=\Phi$ の場合もあるし、 $N_2N_1$ と順序が逆になる場合もある。

このように言語によってさまざまな形の連体修飾の構造がある。我々の機械翻訳システム jaw において、どのような仕組みによって連体修飾構造の翻訳処理を行っているかについて次節で述べる。

### 3. jaw における連体修飾構造の処理方式

#### 3.1 機械翻訳システム jaw の概要

jaw は、日本語から多言語への翻訳を行うパターン変換型の翻訳エンジンである[5]。

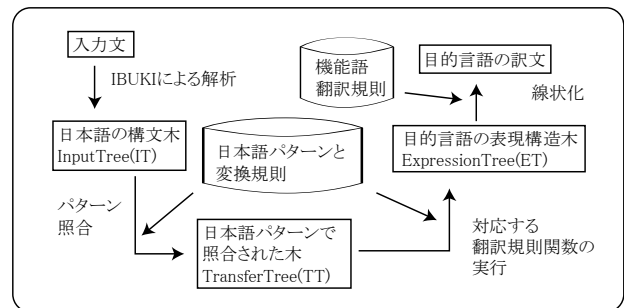


図1 翻訳処理の流れ

#### 3.2 jaw における連体埋め込み構造の処理

jaw における連体埋め込み構造の処理の流れを図2に示す。

図2の中でBの部分は連体埋め込み構造の場合だけの処理である。

X N という係り受け関係があるパターンに対する照合処理は、連体埋め込み構造でなくとも、一般に(1) N の BaseTypeRule の照合

例：〈人〉が〈場所〉を歩く

パターンのキーワード＝「歩く」

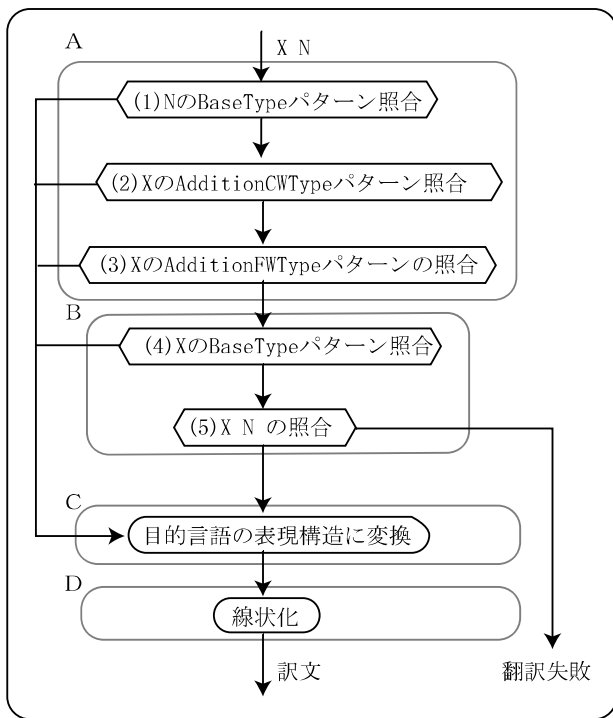


図2 翻訳処理の流れ

(2) X の AdditionCWTypeRule の照合

例：白く<具体>

パターンのキーワード=「白い」

(3) X の AdditionFWTypeRule の照合

例：<体言>の<体言>

パターンのキーワード=「の」

という順序でなされる。

連体埋め込み構造の場合は、X が述語で N が名詞であるが、さらに、

(4) X の BaseTypeRule の照合

(ただし、N も X に係ると考えて照合する)

例：<人>が<場所>を歩く

→<人>が歩く<場所> として照合するを行う。

連体埋め込み構造は一般に修飾部分と被修飾名詞の間に格関係がある「内の関係」と、格関係がない「外の関係」に分類できる。

言語によっては内の関係であるか外の関係であるかによって表現の仕方が異なる。日本語や中国語は異ならない例、ベトナム語は異なる例である。

現在のところ jaw では上記の照合処理のうち (1) (2) (3) で照合された場合を外の関係、(4) で照合された場合を内の関係と判断している。

jaw では、さらに(1) (2) (3)のパターン記述に基づかない外の関係の照合として、  
(5) <用言><体言>の照合  
を設けている。これはVNを照合するもので、defaultの外の関係の照合として機能する。

以上の照合処理は目的言語に依存しない翻訳エンジンとしての処理である。

目的言語に依存するのは線状化関数の部分である。図3が中国語の場合、図4がベトナム語の場合の連体埋め込み修飾されている名詞の線状化処理である。

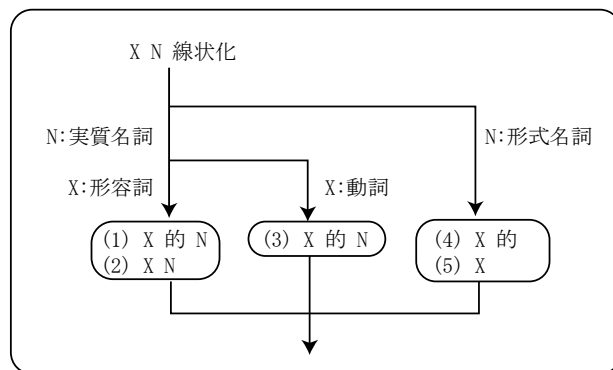


図3 線状化(中国語)

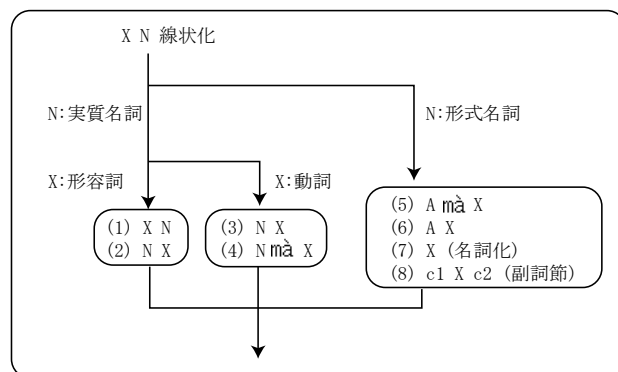


図4 線状化(ベトナム語)

中国語では、Nが実質名詞であり、Xが動詞である場合、常に”X的N”のように”的”を介在させる。Xが形容詞である場合、およびNが形式名詞である場合は訳し分けが必要である。表1は形式名詞「の」の翻訳規則の一部である[1][2]。

ベトナム語では、Nが実質名詞の場合、Xが形容詞であれば、XNとなるかNXとなるかの訳し分けが必要であり、Xが動詞であれば、単純修飾構造(NX)か màを関係詞とする関係節構造(N mà X)かのいづれかとなり、その訳し分けが必要である。Nが形式名詞の場合は、

- (5) 先行詞 A、関係詞 mà をとる関係節構造となる
  - (6) 先行詞 A で関係詞をとらない単純修飾構造となる
  - (7) X のみの名詞節構造となる
  - (8) 接続詞 c1, c2 をとる副詞節構造となる
- のいずれかになる。これらの構造に関する選択規則と先行詞の決定規則が必要になる。表 2 は形式名詞「の」の翻訳規則の一部である[3][4]。

### 3.3 「A の B」の処理

名詞が名詞を修飾する表現は、日本語では「A の B」のように「の」でかかるか、あるいは「A からの B」のように「格助詞+の」で結ぶかであり単純であるが、言語によっては様々な使い分けをしており訳し分けが必要になる。

中国語の場合は、日本語と同様に比較的単純であるが、「A 的 B」となるか「A B」となるかの訳し分けが必要である。表 3 は訳し分け規則の一部である[1][2]。

ベトナム語の場合は、英語と同様に複雑であり、「B p A」、「B A」、「A B」のいずれかの形になり、前置詞 p は 5 個程あり、細かな使い分け規則が必要となる。表 4 は訳し分け規則の一部である[3][4]。

jaw ではこれらの訳し分け処理は図 2 の(1)(2)(3)のパターン照合処理を通じて実行される。すなわち(1) B をキーワードとする BaseTypeRule、(2) A をキーワードとする AdditionCWTypeRule、(3) 「の」をキーワードとする AdditionFWTypeRule で訳し分け処理がなされる。

表 1 中国語：形式名詞「の」の翻訳規則(一部)

番号	パターン構成	従節述語の属性	主節述語の属性	[の]の訳	例文
3	Adj のは(が/を)P	形容詞	用言	的	(1) 安いのがいい→便宜的的好。
5	V のを P	動詞	用言	Φ	(2) 先生が飲食店でコーヒーを飲むのを見かけた→看见老师在饮食点喝咖啡。
6	1 P1 のは(が)N だ	用言	N=具体名詞	的	(3) メガネをかけているのは私の姉だ→戴眼镜的是我的姐姐。
	2 P1 のは(が)N だ	用言	N=抽象名詞	Φ	(4) そのような人に出会ったのは不幸の始めだ→遇见那样的人是不幸的开始。
8	1 P1(V)のは Adj	目的語を伴わない他動詞	形容詞	的	(5) 私がさっき読んだのは一番新しい→我刚才读的最新版。

表 2 ベトナム語：形式名詞「の」の翻訳規則(一部)

区分	日本語の構造	「の」の訳語	翻訳例
関係節	① 強調構文「~V/Adj のは/が N です」.	先行詞 [người, vật, cái, con, nước, việc, điều, những gì]	花びんを壊した[の]はトムです. [Người] mà đã làm vỡ bình hoa là Tom.
	② 被修飾形式名詞「の」と連体埋め込み節が「内の関係」にあり、かつ連体埋め込み節の述語が動態動詞.		太郎が説明した[の]が分からなかった. (Tôi) đã không hiểu [điều] mà Taro đã giải thích
単純修飾	③ 被修飾形式名詞「の」と連体埋め込み節が「内の関係」にあり、かつ連体埋め込み節の述語は静態動詞あるいは形容詞、形容動詞.	cái	赤い[の]を見せてください. Hãy cho xem [cái] màu đỏ.
	④ 形式名詞「の」を用いた強調構文「N は~の が ~」.	日本文の主題名詞 N	りんごは赤い[の]がいい. [Táo] đỏ thì ngon.

表 3 中国語：「N1+の+N2」を「N1+N2」に訳す場合の判定条件(一部)

判定条件	翻訳規則	例文
①数量詞+の+名詞 (例外あり)	「N1+N2」	一匹の猫→一只猫 (三分の一→三分之一 3メートルのリボン→3米(长)的彩带)
②性別名詞+の+名詞	「N1+N2」	雄の兔→雄兔 女性の研究者→女性研究者
⑤時間名詞+の+時間名詞 (例外あり)	「N1+N2」	今日の午後→今天下午 (来年の今日→明年的今天)
⑦学科名詞+の+名詞	「N1+N2」	体育の授業(先生)→体育课(老师)

表4 ベトナム語：「N<sub>1</sub>のN<sub>2</sub>」の翻訳(一部)

	翻訳	タイプ	制約条件		例
			N <sub>1</sub>	N <sub>2</sub>	
①	N <sub>2</sub> của N <sub>1</sub>	所有	主体/地名/場所/自然/具体物/抽象物(精神)	人間/場所/具体物/自然物/抽象/自然現象/人間活動	私の本 sách của tôi (本の私)
②	N <sub>2</sub> ở N <sub>1</sub>	所在	地名/地域/公共施設/仕事場	人/動物(固体)/植物(個体)	東京のおじ ông bác ở Tokyo (おじ在東京)
③	N <sub>2</sub> N <sub>1</sub>	人に関する属性	男女/老若/感覚/感情/意向/...	人	女の人 người phụ nữ (人女)
④	N <sub>1</sub> N <sub>2</sub>	数量	数量名詞	ベトナム語の類別詞を取る名詞	1羽のカラス 1 con quạ (1羽カラス)

#### 4. 翻訳実験

##### 4.1 jaw/Chinese での翻訳実験例

連体修飾構造を含む約 100 例文について jaw/Chinese で翻訳実験を行った。例文数が少ないがこの範囲内では、90%以上の正訳率を得ている。

- (1) 去年の10月修学旅行で京都に行きました。  
→ 去年十月因为修学旅行去了京都。
- (2) 昨晚、奇妙な夢を見た。  
→ 昨晚做了奇怪的梦。
- (3) 彼は体育の先生だ。  
→ 他是体育老师。
- (4) ご飯を食べたのは彼だ。  
→ 吃了饭的是他。
- (5) 彼に会ったのは不幸のはじめだ。  
→ 遇见了他是不幸的开始。
- (6) 私は紅茶はうんと熱いのが好きです。  
→ 我红茶喜欢非常热的。
- (7) ネコが私のベッドで眠っているのを見つけた。  
→ 看见了猫在我的床在睡觉。

##### 4.2 jaw/Vietnamese での翻訳実験例

連体埋め込み文について約 20 文、「A の B」について約 270 文、jaw/Vietnamese で翻訳実験を行った。それぞれ、70%近い正訳率であった。

- (1) あの子供を轆いた自動車  
→ chiếc xe hơi mà đã đụng đứa bé kia
- (2) 太郎が学校を休んだ原因  
→ nguyên nhân Taro đã nghỉ trường học
- (3) 花瓶を壊したのはトムです。  
→ người mà đã làm vỡ bình hoa là Tom
- (4) 私が住んでいるのは日本だ。  
→ nơi tôi đang sống là nước Nhật
- (6) 太郎に殴られた子供が入院しています

→ đứa con mà đã [bị/được] Taro đánh đang nằm viện

(7) この本は読む価値がある

→ sách này có giá trị đọc

#### 5. おわりに

機械翻訳システム jaw における連体修飾構造の翻訳処理法について述べた。具体的な対象として日中、日越における連体修飾構造の翻訳規則について述べ、また、小規模であるが翻訳実験結果について述べた。

翻訳規則としては「A の B の C」のような多層連体修飾の構造、造語法の違いによって複合語となる場合の分析、関係詞のための内と外の関係の判定(ベトナム語の場合)など、さらに例文を広く集めて規則を整備し実験・評価して行くことが必要である。翻訳処理の手法としては概ね見通しが得られたと考えているが、翻訳実験例を増やしなが、さらに問題の発見と処理法の整備を進めたい。

#### 参考文献：

- [1] ト朝暉, 他, 日中機械翻訳における連体修飾構造の翻訳について, FIT, vol.4, pp.73-76, 2005
- [2] Zhaohui BU, et al., Translation of an Adnominal Construction in Japanese-Chinese Machine Translation -Japanese “No” and Chinese “De”-. Proceedings of IEEE NLP-KE 2005, pp410-415, Oct. 2005
- [3] Nguyen My Chau, 他, 日本語-ベトナム語機械翻訳における連体修飾構造の翻訳, 自然言語処理 Vol.12. No.3, pp.145-182, 2005
- [4] Nguyen My Chau, 他, 日本語-ベトナム語機械翻訳における「N<sub>1</sub> の N<sub>2</sub>」の処理, 自然言語処理 Vol.13. No.2, 2006
- [5] 宇野, 他, 日本語から多言語への機械翻訳エンジン jaw, 言語処理学会第 11 回年次大会, pp538-541, 2005