

# 日越機械翻訳システム jaw /Vietnamese における名詞句の処理

Nguyen My Chau 酒井美和 今井啓允 池田尚志  
岐阜大学工学研究科

## 1 はじめに

ベトナム語は、孤立語型の言語であり語形の変化はなく、主要な文法関係は語順で表現される。我々は日本語からアジアの諸言語への機械翻訳システム jaw を開発中である。ベトナム語についても開発を試みており初期的なパイロットモデルであるが jaw /Vietnamese の第1版を作成した[2]。図1にその翻訳例を示す。

- |   |
|---|
| 1. 太陽は東から出て西に沈む。<br>⇒ Mặt trời mọc phía đông và lặn phía tây.                             |
| 2. 彼はとても泳ぎが上手だ。<br>⇒ Anh ấy bơi lội rất giỏi.   |
| 3. 私は深夜までその問題について皆と討論した。<br>⇒ Tôi đã thảo luận về vấn đề kia với mọi người cho đến khuya. |
| 4. この薬を飲むと痛みがなくなる。<br>⇒ Uống thuốc này thì con đau biến mất.                              |

図1：翻訳例

図1のような比較的簡単な100文の翻訳を行ったところ、表1に示すように「N1のN2」の翻訳に問題が多いことが分かった。そこで、本論文では「N1のN2」のベトナム語訳について分析した結果を報告する。

表1：jaw/Vietnamese による100文の翻訳結果

翻訳結果	総文数	「の」を含む文	
		「の」の問題	その他の問題
○	72	20	
×	28	18	2
合計	100	40	

## 2 jaw /Vietnamese の概要

### 2.1 システムの概要

jaw における翻訳規則は、日本語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である。表現構造として VC++ のオブジェクトを利用している。

日本語入力文は解析されて構文 IT (Input-Tree) となり、翻訳規則のパターン辞書と照合されて対応する翻訳規則の木 TT (TransferTree) となる。

翻訳規則は目的言語の表現構造を作り出すプログラム (dll) として格納されており、TT の各翻訳規則関数を実行すると目的言語の表現構造の木 ET (ExpressionTree) が生成される。最後に ET の各オブジェクトのメソッドである線状化関数を実行することで目的言語の翻訳文が生成される (第1フェーズの翻訳)。

機能語部分の翻訳については、日本語の機能語部をその意味機能ごとにグルーピングし、それぞれの機能語グループごとに翻訳規則を定めている。この翻訳規則は機能語部の翻訳に必要な語句 (一般には複数の語句の集合になる) を ET 上に記述する。先に述べた生成関数は、これらの語句も含めて線状化処理を行う (第2フェーズの翻訳) [1]。

本論文では、第1フェーズの翻訳について、その中でも特に名詞句の翻訳を中心に述べる。

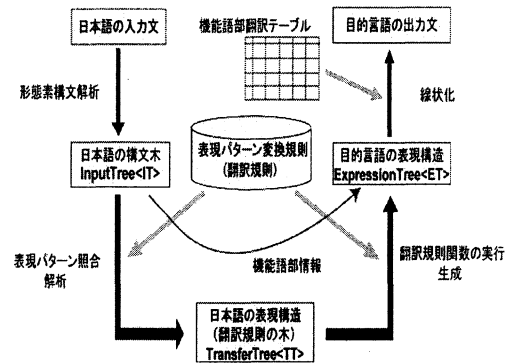


図2：パターン変換型翻訳システム jaw

### 2.2 名詞句の表現構造と線状化関数

ベトナム語の表現構造のために、次のようなクラスを設けた。

- |          |              |           |
|----------|--------------|-----------|
| CTObject |              |           |
| ├──      | CProposition | 動詞系       |
| ├──      | CNoun        | 名詞系       |
| ├──      | CQuantity    | 数量表現      |
| ├──      | CpConnection | 動詞系に関する連結 |
| ├──      | CnConnection | 名詞系に関する連結 |
| ├──      | CAverb       | 副詞表現      |
| CMoVe    | 機能後などの       |           |
| CMoJe    | 情報格納クラス      |           |

名詞系クラスには、以下のようなメンバーを持たせた。

CString	m_centerW	// 訳語の中心語
CString	m_classifier	// 類別詞
CNoun	*m_designator	// 指示詞
CNoun	*m_pluralDesignator	// 集合指示詞
CNoun	*m_possessor	// 所属者
CNoun	*m_nounModifier	// 名詞系修飾要素
CNoun	*m_quantifier	// 数～、～ずつ
CQuantity	*m_quantity	// 数量表現
CProposition	*m_adjective_1	// 形容詞(量の存在)
CProposition	*m_adjective_2	// 形容詞
CString	m_role	// case name
CString	m_caseMarker	// 前置詞

これらのメンバーは、線状化関数で次のような語順に並べられる。

Noun:=  
[quantity | quantifier | pluralDesignator]+unit  
+ classifier + N + adjective + designator  
+ nounModifier + possessor

### 3 「N1 の N2」の日越対応

日本語の「N1 の N2」という構造は、外見は簡単なだけに、その裏にある N1 と N2 の意味関係の多様さの中の規則性を探り出すことは難しい [3]。ベトナム語ではこの「N1 の N2」は、N1 と N2 の意味関係に依っていくつかの異なった形で表現される。我々は、日本語の 1102 例文から「の」を含む 369 文を抽出し、人手で翻訳してみた。表 2 に示すように以下の 10 のタイプに分類できることを見出した。

表 2 : 「の」を含む 369 文中「の」の分類

分類		369 文中「の」の出現回数
1	「N1 の N2」	N1 của N2
2		N2 α N1
3		N2N1
4		N1N2
5		特殊表現
6		固定表現
7	形式名詞	65
8	の+連体形	22
9	終助詞「の」	25
10	慣用句	6
合計		421

本論文では、このうち「N1 の N2」のベトナム語への翻訳の 6 つのタイプについて考察する (表 3)。

#### (1) 「N2 của N1」

「N2 của N1」は、ベトナム語の所有表現である。「của」はベトナム語の所有詞である。名詞 N1 は所有者であり、名詞 N2 は所有されるヒト・モノ・コトである。日本語の「N1 の N2」に対してベトナム語の所有表現は「N2 của N1」と順序が逆になる。

例：「私の本」 → sách của tôi  
(本 の 私)

#### (2) 「N2 α N1」

「所在」や「性質」や「格関係」などを表す場合、ベトナム語では、「N2 α N1」と表現する。αは、「所在」の場合「ở」、[性質]の場合「có」となる。

例：「東京のおじ」 → ông bác ở Tokyo  
(おじ 在 東京)

「格関係」の場合は、αは格を表現するベトナム語の前置詞である場合と、α=của でもよい場合 (カラ、ヲ、ガ) がある。例えば、

「図書館への道」 → đường đến thư viện  
(道 へ 図書館)

「母からの手紙」 → thư từ của mẹ  
(手紙 から/の 母)

#### (3) 「N2N1」

「人に関する属性」、「物に関する属性」、「事に関する属性」、「時間的属性」の場合は、前置詞を使わずに「N2N1」と表現する。ベトナム語には固定表現の場合も含めてこの型が多い。

例：「土曜の夜」 → tối thứ bảy  
(夜 土曜)

#### (4) 「N1N2」

N1 が N2 の数量を表現する場合には、ベトナム語では、例えば

「1羽のカラス」 → 1 con quạ  
(1 羽 カラス)

のように「N1N2」と表現する。この例で「con」は類別詞 (classifier) である。類別詞はベトナム語に特徴的な修飾要素で、名詞の種類や形などを示すものである。無生物を示す“cái”、“chiếc”、生物を示す“con”、丸くて小さい物を示す“hòn”、“cục”、“viên”などがある [2]。従って、日本語の数量表現「N1 の N2」に対する翻訳は{ [N1:数] [N2:類別詞] N2 } である。ただし、「2つの都市」のような例では「都市」は「種類」や「形」などが決められないから、類別詞はとらない。

表3：「N1のN2」のベトナム語への翻訳

タイプ	例	制約条件		翻訳
		N1	N2	
1 N2 của N1	[所属] 私の本 sách của tôi (本 の 私)	主体/地名/場所/自然	人間/場所/具体物/自然物/抽象	N2 của N1
2 N2 α N1	[所在] 東京のおじ ông bác ở Tokyo (おじ 在 東京)	地名/場所	人/動物,(固体)/植物(个体)	N2 ở N1
	[性質] 甘い香りの花 hoa có mùi thơm (花 ある 甘い香り)	刺激/影響/感情	具体物/様相/報告/訴え/人/機関/団体/党派/宗教	N2 có N1
	[格+の] 母からの手紙 thư từ của mẹ (手紙からの 母)	[格]= から		N2[格] N1 または N2 của N1
	[格+の] 東京までの切符 vé đến Tokyo (切符 まで 東京)	[格]=まで,へ,と,にとって,について,として,		N2 [格] N1
	[隠れた格+の] Aさんの絵 (bức tranh của A)	[格]= を,が		N2 của N1 (曖昧)
3 N2N1	[人に関する属性] 農家の人 người nông gia (人 農家)	男女/老若/人間(社会集団)/人間(能力・性向)/産業/感覚/感情/意向/健康・不全/年齢/国名/交際/適合・不適合/場	人	N2 N1
	[時間的属性] 土曜の夜 tối thứ bảy (夜 土曜)	時間/自然現象/人間/交際	人/時間/具体物/人間活動/気象・天象/出来事	
	[物に関する属性]	具体物/学問・学科/場所/性質/度量衡	動作/支配/取引/業/具体物/抽象物(精神)	
	[事に関する属性]	抽象的關係/事/界/生命現象/具体物/	事/抽象物/具体物/間/抽象的關係	
4 N1N2	[数量] 1羽のカラス 1 con quạ (1 羽 カラス)	数量名詞	ベトナム語の類別詞を取る名詞	[N1: 数][N2: 類別詞]N2
	2つの都市 2 thành phố (2 都市)	数 + 助数詞	ベトナム語の類別詞を取らない名詞	[N1:数] N2
5 特殊表現	[学年] 大学の1年生 sinh viên đại học năm thứ 1 (学生 大学 年 第 1)			sinh viên N1 năm thứ [N2: 数]
	[数量比較] 去年の2倍 gấp 2 lần của năm ngoái ( 2 倍 の 去年)		「～倍」のみ	gấp [N2: 数] lần của N1
	[分数] 3分の4 3 phần 4 (3 分 4)			N2 phần [N1= 数]
6 固定表現	[固定表現+N2]	N1= すべて,かぎり,全く,ほど,多く,たくさん,いくつ,何,		「の」を分離しないで扱う
	[N1+固定表現]		N2= まま,間,とき,節,うち,ところ,方,ような,ために,おかげで,	

