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。
- ⇒ Mặt trời mọc phía đông và lặn phía tây.
- 2. 彼はとても泳ぎが上手だ。
- ⇒ Anh ấy boi lội rất giỏi.
- 3. 私は深夜までその問題について皆と討論した。
- ⇒ Tôi dã 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。
- ⇒ Uống thuốc này thì con dau 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			
0	72	2	20		
×	28	18 「の」の問題	2 その他の問題		
合計	100	40			

2 iaw /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 (Transfer Tree) 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(dil)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。

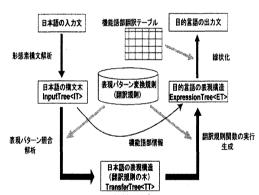


図2:パターン変換型翻訳システム jaw

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。



CModeV 機能後などの CModeJ 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 // 指示詞 CNoun *m designator CNoun *m pluralDesignator // 集合指示詞 // 所属者 **CNoun** *m possessor *m nounModifier // 名詞系修飾要素 **CNoun** *m quantifier **CNoun** // 数~、~ずつ *m_quantity **CQuantity** // 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 // 前置詞 **CString** m caseMarker

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	151
2		N2 α N1	10
3		N2N1	83
4		N1N2	8
5		特殊表現	8
6		固定表現	43
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 NII

「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 \underline{o} 本」 \rightarrow sách của tôi (本 \underline{o} 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、 Γ N2 α N1」と表現する。 α は、[所在]の場合 Γ o」、[性質]の場合 Γ co」となる。例: 「東京のおじ」 \rightarrow <u>ông bác ở Tokyo</u>

(おじ $\underline{\alpha}$ 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 = cua でもよい場合

「図書館への道」 → <u>dường đến thu viện</u> (道 へ 図書館)

(カラ、ヲ、ガ) がある。例えば、

「母<u>からの</u>手紙」→ <u>thu</u> <u>tù/ của me</u> (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」 → <u>tối thứ bảy</u> (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 qua (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"、"chiéc"、生物を示す"con"、丸くて小さい物を示す"hòn"、"cuc"、"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

χ 3 []	タイプ	」のベトナム語への翻訳 例	40	约冬	翻訳
		73	制約条件 N1 N2		開打 10 人
1	N2 của N1	[所属] 私の本 <u>sách của tôi</u> (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 <u>ông bác ở Tokyo</u> (おじ 在 東京)	地名/場所	人/動物(固体)/植物 (個体)	N2 ở N1
		[性質] 甘い香りの花 hoa có mùi thom (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 cóN1
		[格+の] 母からの手紙 thu từ/cùa me (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 dén 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	Λ	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]	抽象的関係/事/界/ 生命現象/具体物/	事/抽象物/具体物/間/ 抽象的関係	
4	N1N2	[数量] 1 羽のカラス 1 <u>con qua</u> (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 <u>thành phố</u> (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 dai học nam thứ 1 (学生 大学 年 第 1)			sinh viên N1 nam thứ [N2: 数]
		(数量比較) 去年の2倍 gấp 2 lần của nam 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 <u>分</u> 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,	

(5) 特殊表現

[分数]・[学年]・[数量比較」などの数量表現の翻訳は、(1)~(4)の場合とは違う形になる。ただし表3の数量比較の翻訳規則が当てはまるのは

「〜倍」の場合だけである。表3には整理できていないが、数量比較には、

a.「N2 của N1」となる場合 (タイプ1) 「去年の半分」 → 「<u>phân nửa</u> của <u>nam ngoái</u>」 半分 の 去年

b.「N2N1」となる場合

(タイプ3)

「ケーキの半分」→

「<u>phân nửa cái bánh</u>」 半分 ケーキ

もある。

(6) 固定表現

N1 が固定表現(「すべて」「多く」など)の場合と、N2 が固定表現(「うち」「ために」)の場合は、「の」を分離せずに、「<u>すべての</u>N2」「N1 <u>のうち</u>」のように統合に扱いそれぞれのベトナム語への翻訳処理を行う[5]。

表4: jaw /Vietnamese による翻訳規則の記述

4 「N1のN2」の日越翻訳

4.1 翻訳実験

3 で述べたように、「N1 の N2」の日越翻訳には、 さまざまの訳し分けが必要である。我々は jaw /Vietnamese の枠組みを用いて、訳し分けの実験 を行った。

jaw /Vietnamese では、次の3つのタイプの翻訳規則を記述することができる(表4)。

- ① 名詞 N2 をキーワードとして、補足名詞 N1 に条件をつける翻訳規則(Base タイプ)。
- ② 「の」をキーワードとして、N1、N2ともに条件をつける翻訳規則(Add-FW タイプ)。
- ③ 名詞 NI をキーワードとして、係り先名詞 N2 に条件をつける翻訳規則(Add-CW タイプ)。

翻訳規則中の意味制約条件にはNTTの語彙体系の意味属性を用いた[4]。

4.2 評価

「の」を含む 80 名詞句「N1 の N2」を抽出し、表 3 の翻訳規則を 4.1 に述べた方式で記述し、j aw / V i e t namese で翻訳実験を行った。結果は 67 個が正しく翻訳できた。表 1 の 40 個については、31 個が正しく翻訳できた。

ТҮРЕ	文節 番号	係り先 番号	自立語条件 (例)	機能語 条件	K₩	自立語条件 (例)	翻訳規則
-	1	2	人間 (社会集団)/人 間 (能力・性向)	0			·
Base			(農家)				
Dase	2	0			人。		người
Add-FW	1	2	時間 (土曜)		Ø		φ
	2	0				人/曆日 (夜)	
Add-CW	1	2			すべての		tất cả các
-	2	0				体言 (場合)	

5. おわりに

本論文では、日越機械翻訳における名詞句の処理、特に「N1 の N2」表現の意味分類と翻訳手法を提案した。精度を向上させるために、さらに分析を深め、次の段階では、「の」+連体形とタイプ3の表現の分類についてさらに分析することが課題である。

参考文献

[1]「パターン変換型機械翻訳システム j-aw につ

いて」今井啓允、謝軍、池田尚志、第1回 FIT 年 次大会 2002.9

[2]「日越語の構造比較と j-aw/Vietnamese による日越機械翻訳」Nguyen My Chau、今井啓允、酒井美和、池田尚志、第1回 FIT 年次大会 2002.9

- [3] 寺村秀夫(1991)『日本語のシンタクスと意味』 くろしお出版
- [4] NTT コミュニケーション科学基礎研究所(1997) 『日本語語彙大系 1 : 意味体系』岩波書店
- [5] グループ・ジャマシイ(1998)『日本語文型辞典』 くろしお出版