

E-47 日越語の構造比較と j-aw/Vietnamese による日越機械翻訳

Japanese- Vietnamese Structural Comparison and Machine Translation System by j-aw/Vietnamese

Nguyen My Chau[†] 今井啓允[†] 酒井美和[†] 池田尚志[†]
 גען מיי צאוו Yoshimasa Imai Miwa Sakai Takashi Ikeda

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。文字についてはフランス語の影響を受けており、29 文字のアルファベットと6つ声調表記で表現される[2]。

我々は j-aw[1]の枠組みを用いて、日本語からベトナム語への機械翻訳システムの構築を試みている。まだその第1段階であるが、簡単な日本語文の翻訳実験を行って見たので報告する。

2. ベトナム語の構造の概要

① 名詞句の構造

名詞句への修飾は主として後ろからなされるが、数量表現は前方に位置する(図1)。性、数、人称による変化はない。数は quantifier(量詞)で示される場合もある。しかし動詞がそれに呼応して変化することはない。特徴的な修飾要素として classifier(類別詞)がある[4]。類別詞は名詞の種類・形などを示すもので、無生物を示す"caì", "chiéc", 生物を示す"con", 丸くて小さい物を示す"hòn", "cục", "viên"など、よく使われるのは40個ほどである。

Noun:=	
negation+[quantity quantifier pluraldesignator]+classifier+N	
+adjective+designator +n_modifier +possessor	
quantity:=	[number adjective1]+unit // 数詞, 多い, メートル
quantifier:=	{"mấy", "mỗi"} // 数~, ~ずつ
pluraldesignator:=	{"các", "những"} // それぞれ, ~等
classifier:=	{"cái", "chiéc", "con"} // 無生物, 無生物, 動物
adjective:=	adverb+adjective2 // とても+きれい
designator:=	{"này", "kia", "ấy"} // この, その, あの
n_modifier:=	preposition+noun // 前置詞+名詞
possessor:=	"của"+noun_possessor // の+所有者

図1 名詞句の構造

名詞を修飾する埋め込み文は、図1で adjective の位置に来る。関係代名詞としてはいつも"mà"が使われる。

例1: 昨日読んだ本はとても面白かった。

Quyển sách mà tôi đã đọc hôm qua rất hay.

本 関係代名詞 私 ~た 読む 昨日 とても 面白い

① 単文の構造

日本語の語順は SOV である。日本語は状態記述的(“は”的)

[†]岐阜大学大学院工学研究科,

Faculty of Engineering, Gifu University

であるのに対して、ベトナム語は英語のような行動表現的な言語に近いと言える。ベトナム語の語順は基本的に SVO 型であり、主語、目的語の省略は許されない。

例2: 私は赤いリンゴを食べます。→ Tôi ăn trái táo đỏ.

S O V S V O

また主語のほかに主題を表現することがあり、TSVO 型の言語というべきかも知れない。主語、目的語などの基本的な要素以外の語順についてはまだ全般的な規則化は出来ていないが、概ね時が先頭に来て、目的語の後ろにその他の修飾要素が連なる。

例3: 象は鼻が長い。

Con voi cái vòi thì dài.

T S V

動詞に語形変化はない。時制副詞が動詞に前置される("đã"た, "vừa/ mới"したばかり, "đang"しているところ, "sắp"もうすぐ, sẽ 未来)。否定文は、否定詞 "không"が時制副詞と動詞の間に置かれる(時制副詞+ 否定詞+ 動詞)。

Tôi đã không đi=「私」+「~た」+否定+「行く」

= 私は行きませんでした。

(態) : được khen/ bị mắng 「~される」+「ほめる/叱る」

= ほめられる/叱られる

(法) : ăn đi 「食べる」+「~ください」= 食べてください

③ 複文の構造

主節を S1、従属節を S2 とすると、ベトナム語の複文は:

c1 S2 c2 S1

の形をとる。

c1, c2 は接続詞である。c2 は省略される場合もある。

例4: いろいろな心配があるので、寝られなかった。

Vì tôi có nhiều lo âu nên tôi không ngủ được.

ので 私 ある 色々 心配 だから 私 否定 寝る

3. j-aw による日越翻訳規則の記述

現在のところ、動詞、名詞については図2,3のようなメンバー変数によってベトナム語の表現構造を記述している。時制やモダリティなどは pMode に記述するがその翻訳処理はまだ行っていない。

CQuantity	*m_quantity;
CNoun	*m_quantifier;
CNoun	*m_pluralDesignator;
CString	m_classifier;
CProposition	*m_adjective_1;
CProposition	*m_adjective_2;

CString	m_Designator;
CNoun	*m_nounModifier;
CNoun	*m_possessor;
CAdverb	*m_adverb;
CString	role_name;
CString	case_Marker;

図2 名詞クラスのメンバー変数

CString	m_pType;
CProposition	*m_complementary;
CNoun	*m_topic;
CNoun	*m_subject;
CNoun	*m_object;
CQuantity	*m_quantity;
CNoun	*m_time;
CNoun	*m_location;
CAdverb	*m_adverb;
CProposition	*m_propositionalAdverb
CpConnection	*m_pConnection;
CNoun	*m_nounModifier;
CString	m_pMode;
CString	m_connection;
CString	m_cnctPos;

図3 動詞クラスのメンバー変数

それぞれのクラスに定義されている linearize 関数を実行すると、語順が調整され翻訳文が出力される。孤立語型であるために、語形変化などの処理は必要ない。以下、規則記述の例を示す。

① BaseType の規則の例：

彼はとても泳ぎが上手だ。

→ Anh ấy bơi lội rất giỏi.

「彼」 pid 2000056、Type : Base

Class	memberName	memberClass	Value
CNoun	m_centerW	CString	anh ấy

「泳ぎ」 pid 2082777、Type : Base

Class	memberName	memberClass	Value
CNoun	m_centerW	CString	bơi lội

「上手だ」 pid 1006122、Type : Base

Class	mName	mClass	Value	roleName
CProposition	m_centerW	CString	giỏi	
CProposition	m_subject	CNoun	2	
CProposition	m_topic	CNoun	1	

② AddType-CW の規則の例

彼はとても泳ぎが上手だ。

「とても」 pid 4002、Type : Add-CW

Class	memberName	memberClass	Value
CProposition	m_adverb	CAdverb	
CAdverb	m_centerW	CString	rất

③ AddType-FW の例：私は耳で話を聞く。

→ Tôi nghe sự việc bằng tai.

「で」 pid 1007、Type : Add-FW

Class	memberName	memberClass	Value
CProposition	m_nModifier	CNoun	
CProposition	m_role	CString	method
CProposition	m_caseMarker	CString	bằng

④ AddType-CW の例：太陽は東から出て西に沈む。

→ Mặt trời mọc phương đông và lặn phương tây.

「～て」 pid 1003、Type : Add-CW

Class	memberName	memberClass	Value
CProposition	m_pConnection	CpConnection	
CpConnection	m_pSubordinate	CProposition	1
CpConnection	m_connect2	CString	và

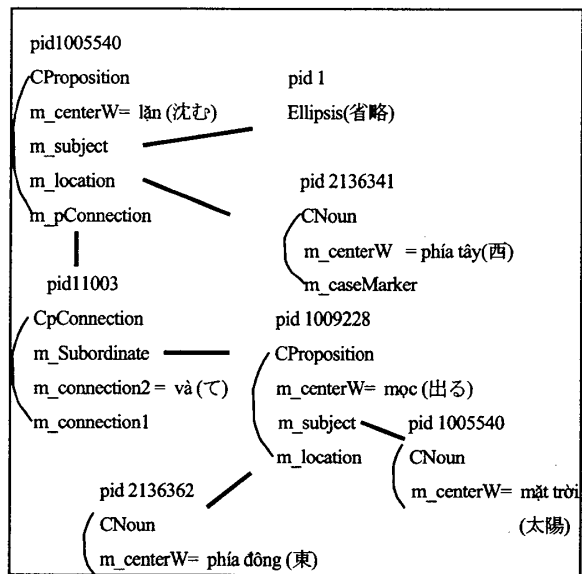


図4 「太陽は東から出て西に沈む」の表現構造

4. おわりに

日越機械翻訳のためのベトナム語の文法の概要と、簡単な日本語例文に対する j-aw/Vietnamese による翻訳例を示した。まだ第1段階であるが、次の段階では、否定、テンス・アスペクト、モダリティなどの翻訳について組み込んでいく予定である。

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] 宇根祥夫(1995)『ベトナム語早分かり』三修社出版
 [3] Đại học Tổng Hợp Hà Nội - Khoa Tiếng Việt (1980) *Giáo trình cơ sở Tiếng Việt thực hành Tập 1*
 [4] Diệp Quang Ban và Hoàng Văn Thung (2001) *Ngữ pháp tiếng Việt, tập 1&2*. Nhà Xuất bản Giáo Dục