

# 機械翻訳システム jaw と多言語への翻訳実験

浅井良信 宇野修一 卜朝暉 福本真哉 田中友樹 田中伸明 Nguyen My Chau Ma Ngin Khaw Cing

Samantha Thelijigoda 張穎 吉田鑑地 谷口真代 伊佐治和哉 松本忠博 池田尚志

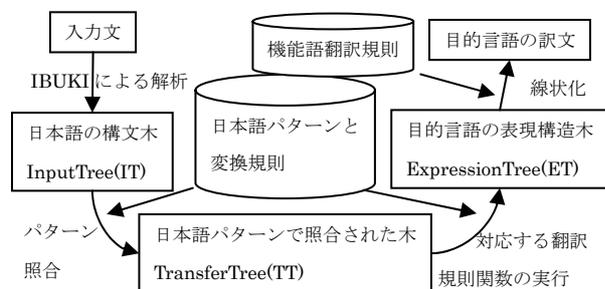
岐阜大学 工学部

## 1 はじめに

我々は、日本語からいろいろの言語への翻訳を行う機械翻訳エンジン jaw を開発している[1]。現在 jaw を用いていくつかの言語への機械翻訳システム (jaw/Chinese、jaw/Vietnamese、jaw/Myanmar、jaw/Sinhalese、jaw/SL) の構築を行っている。200 程度の例文を材料として翻訳実験をおこなっている段階である。本稿では、これらの翻訳実験におけるそれぞれの目的言語での対処事例・問題点などについていくつか報告する。なお手話への翻訳システム jaw/SL については[2]で詳しく述べた。

## 2 機械翻訳エンジン jaw の概要

図 1 に基本的な処理の流れを示す。



各目的言語ごとに行うことは、それぞれの言語の構造及び日本語との対応を分析して、目的言語の表現部品を格納する C++のクラスを設計すること、日本語からのパターン翻訳規則データベースと機能語の翻訳規則データベースを構築すること、表現構造から線形な文を作り出すための線状化関数(C++のクラスメソッドである)を定義することである。

## 3 対処事例・問題点など

### 3.1 jaw/Chinese における事例

[1]日中両言語での構文的ずれへの対処例

#### ・「目的語+他動詞」が自動詞に対応する事例

次の例文(1)は、ベトナム語、ミャンマー語では原文と同じ他動詞文として翻訳できるが、中国語、シンハラ語では「死を遂げる」を、「死亡する」を動詞とする自動詞文に翻訳し、原文の連体修飾の形容詞「悲劇的な」は連用修飾語に翻訳しなければならない。

(1) 彼は悲劇的な死を遂げた。

(1-C) 他 悲惨地 死亡了。

(1-V) Anh ấy đã kết thúc cái chết bi thảm.

彼 過去 遂げる 死 悲劇的

(1-M) သူ သည် ဝမ်းနည်းစရာကောင်းသော သေခြင်း နှင့် တွေ့ဆုံ သည်။

彼 は 悲劇的な 死ぬ事と 出会う 完成助辞

(1-S) ඔහු දුක්මුසුව මරණයට පත්විය

彼 悲劇的 死亡 した

このような構造のずれはいろいろの言語対でたくさん見られることと思われるが、Jaw/Chinese では次のように「死亡」に隠れ目的語(dummy-obj)を持たせることで対処した。

- ・ N が 死を 遂げる  
→死亡 (subj=N, dummy-obj=Φ)
- ・ 悲劇的な N  
→N(adj=悲惨)

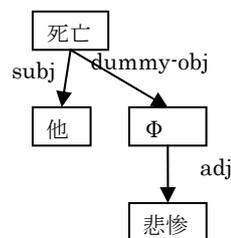


図 2: (1-C)の構造

線状化関数は、dummy-obj があってそれを修飾する連体語句があれば、その修飾語句を状語(動詞の前におく連用修飾語)として生成する。形容詞を連用修飾化する場合には形容詞の訳語+“地”とする。

#### ・他動詞を「目的語+他動詞」に分解翻訳する事例

次の例文(2)では、他動詞「発音する(VJ)」が「音を発する」のように「目的語(OC)+他動詞(VC)」の形に分解され翻訳される。(C2)では原文の他動詞(VJ)の目的語 OJ を介詞句として翻訳しているが、(C1)では OJ は目的語(OC)を修飾する連体修飾語として翻訳している。

- (2) 先生は生徒に各単語を正確に発音させた。  
 (2-C1) 老师 让 学生 正确地 发了 各单词的 音。  
 (2-C2) 老师 让 学生 对 各单词 正确地 发了 音。

(2-C1)のように翻訳するためには、発する(VC)の目的語(OC)を記述する際、その内部構造についても記述する必要がある。

- A が B を 発音する (2-C1)  
 →発(subj=A, obj=音(of=B(ofW="的")))  
 ofW は連体修飾構造で"的"を使うかどうかを指定する標識。ofは名詞への連体修飾する名詞を指す。
- A が B を 発音する (2-C2)  
 →発(subj=A, obj=音,  
 modi=B(cname="target", cmarker="对"))  
 modi は連用修飾要素、cname はその意味役割、cmarker はその際使われる介詞。

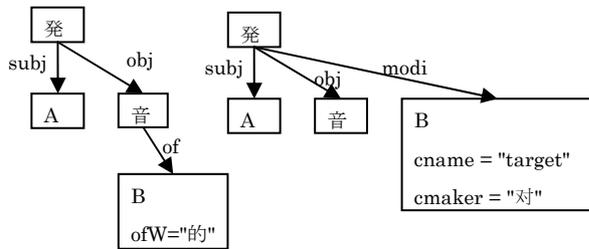


図 3：発音する(2-C1)      図 4：発音する(2-C2)

現在の jawEditor は、(2-C1)のように深い階層まで記述する機能は持っていない。現在は jawEditor が出力する翻訳プログラムを手手で修正して実験している。このような階層的に深い記述も出来るように jawEditor の機能を強化することは次の課題である。

[2] 「自由にする/になる」など

jaw の baseRule では、内容語だけでなく後接する機能語を含めたパターンに対して翻訳規則を設定することが出来る。例文(3)の「自由にする」は (3-C)のように「让…自由」という使役の多述語構造に翻訳される。

- (3) 政府は囚人たちを自由にした。  
 (3-C) 政府 让 犯人们 自由了。

このために jaw では形容動詞「自由だ」ではなく、機能語「にする」を含めた「自由にする」に対する翻訳規則を書いている。

- A が 自由だ  
 →自由(subj=A)
- A が B を 自由にする  
 →自由(modi=A(cname="causer"),

subj=B, type="使役", prep="让")

type は文型、prep は特殊文型で使う動詞や助動詞。

「自由になる」に対しても同様である。

- A が自由になる  
 →変(subj=A, comp=自由)

このように jaw では機能語を含めた個別の翻訳規則を書くことは可能である。しかし、「～する」や「～なる」は、「自由にする/なる」、「静かにする/なる」、「大人にする/なる」、「ゆっくりする/となる」、「ぴょんぴょんする」など、いろいろの語について頻繁に現れ、また対応する中国語での表現はさまざまな構造をとる。個別に規則を書くのは量的に大変であり、一般的に対応できる規則も必要である。日本語とそれぞれの言語での表現の対応について分析し、一般的な翻訳規則と例外的な個別の翻訳規則を整理する必要がある。

3.2 jaw/Vietnamese における事例(複合語)

複合語の表現は下の例に示すように言語毎にさまざまである。

(1) 3m20cm	J,C,V,M
m 3 cm 20	S
(2) 5月3日6時5分	J,C,S,M
日3月5 6時5分	V
(3) 5000円	J,C,V
円 5000	M,S
(4) 岐阜大学工学部	J,C,M,S
大学 岐阜 学部 工学	V
(5) スタンダード石油会社	J,C,M,S
会社 石油 スタンダード	V

(J:日本語、C:中国語、V:ベトナム語、M:ミャンマー語、S:シンハラ語)

上記の例では、日本語との対応は中国語は全く同じであり、ベトナム語が一番違いが大きいようである。

jaw では、複合語に関しては今のところ暫定的に、長さの表現、時の表現、その他一般の数量表現、その他一般の複合語という4つのクラスを設けて処理している(図5)。それぞれのクラスに対する線状化関数で翻訳表現を作り出すことが出来るので、jaw/Vietnamese の場合、(1),(2),(3)は正しく生成できるが、(4)と(5)の区別は現在出来ていない。一般の複合語については全逆順に線状化するようにしているので、(4)が正しく線状化できない。さらに広く複合語の語順の対応について整理し、必要な複合語クラスと線状化関数を整備することは今後の課題である。

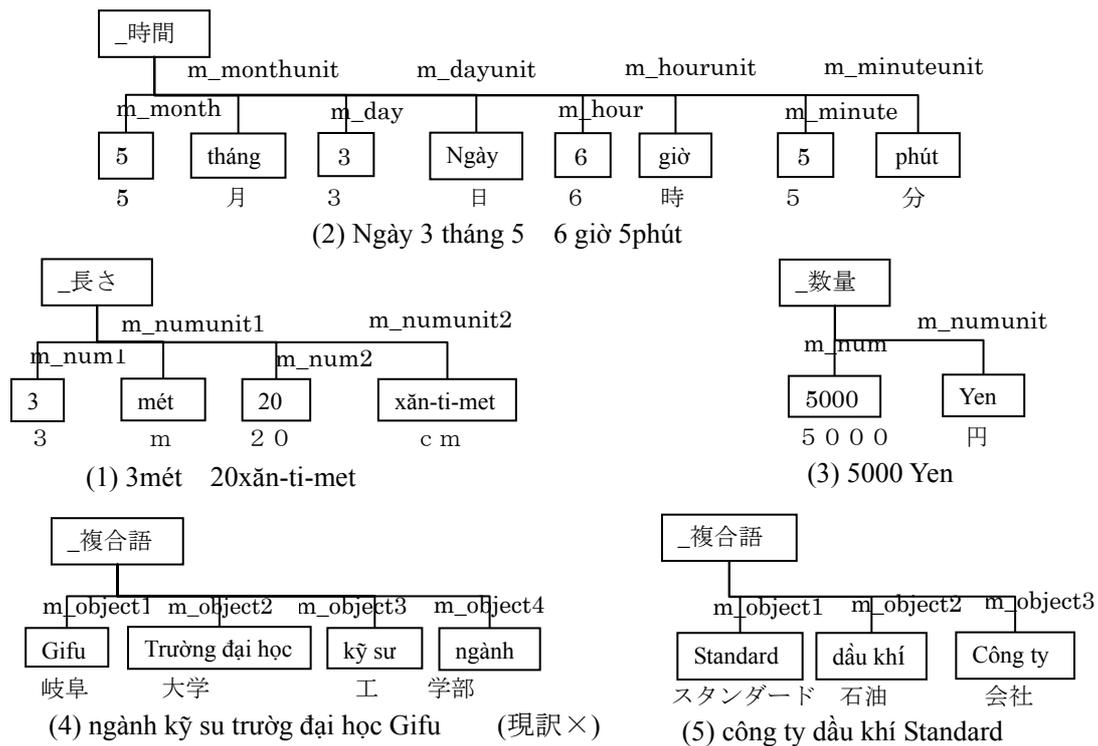


図 5：複合語のクラスと線状化

### 3.3 jaw/Myanmar における事例 (用言後接機能語)

jaw では用言後接機能語の翻訳規則は、日本語の用言後接機能語とそれに対応する目的言語での表現要素（一般には複数）の対応表として記述する。訳し分けが必要な場合はその条件も表で記述するようになっている。述語クラスの線状化関数が、機能語の語順規則にしたがって述語部構造を生成する。

ミャンマー語は膠着語であり、日本語の用言後接機能語部は基本的にミャンマー語でも用言に後接する部分の表現に翻訳される。しかし、ミャンマー語では同じ動詞が2度繰り返して表現されたり、補助動詞や接続助辞が使われたり複雑である。ミャンマー語の言語構造についてはまだ研究が浅く、特に述語部の構造、ヴォイスやモダリティなどに関する研究はほとんどない。

著者らはこれまでに日本語との対応関係を念頭においてミャンマー語の述語部の構造を分析してきた。その結果、ミャンマー語に現れる表現要素および語順を整理することが出来た。その一部を表 1 に示す。たとえば、日本語の用言後接機能語が「なければならないかもしれない」などのモダリティを表現する機能語を含む場合には、対応するミャンマー語の述語部にはモダリティ補助動詞とそれに関連する部分が現れる。

1	合せ動詞の名詞部	31-38	判断、副動詞など
2	アスペクト助辞	39	モダリティ接続助辞
3	4 に対する否定	40	41 に対する否定
4	主動詞	41	モダリティ補助動詞
5-11	使役、受身など	42	41 に対する進行助辞
12	4 に対する進行助辞	43	41 に対する過去助辞
13	4 に対する過去助辞	44	41 に対する複数助辞
14-28	許可、可能、複数、副詞呼応助辞など	45	41 に対する完成助辞
29	繫辞(copula)	46	間投助辞
30	4 に対する完成助辞		

表 1：述語部の表現要素と語順

機能語 (なければならない)	翻訳規則 (ミャンマー語)	
V+なければならない	$\theta + V + \alpha_2^{\circ} + \theta + \eta + \alpha_2^{\circ}$ 3+4+39+40+41+45	A
V+なければならないか	$\theta + V + \alpha_2^{\circ} + \theta + \eta + \alpha_2^{\circ} + \alpha_2^{\circ}$ 3+4+39+40+41+43+45	A
N+なければならない	$N + \alpha_2^{\circ} + \eta + \theta + \alpha_2^{\circ}$ 29 + 41+45	B

表 2：機能語「なければならない」の訳語規則

翻訳規則の事例として「なければならない」に対してミャンマー語の表現を分析した結果、表 2 に示す翻訳規則が得られた。次の A,B は「なければならない」に対

応するミャンマー語の次の二つの異なる表現である。

(A) 日本語と同じく二重否定を用いる表現

3+4+39+40+41+45

(1) 私は東京へ 行か なければならない。

ကျွန်ုပ် ၏ တိုကျို ကို မ သွား လို့ မ ရ ဘူး။

私は 東京 へ 3 4 39 40 41 45

(2) 私は東京へ 行か なければならなかった。

ကျွန်ုပ် ၏ တိုကျို ကို မ သွား လို့ မ ရ ခဲ့ ဘူး။

私は 東京 へ 3 4 39 40 41 43 45

(B) 英語の“must”のように補助動詞を用いる表現

29+41+45

(3) 東京は 美しい 街 でなければならない。

တိုကျို ၏ လှပသော မြို့ ဖြစ် ရ မည်။

東京 は 美しい 街 29 41 45

「なければならない」に関しては、同じく膠着語であるシンハラ語でも類似の訳し分けが必要のようである。機能語に関する翻訳規則が jaw の方式でうまく処理できるかどうか、いづれにしてもそれぞれの言語規則の分析・整備がまず必要である。

### 3.4 jaw/Sinhalese における事例(格助詞)

シンハラ語は膠着語の系統であり、日本語と同じく格助詞の体系を持つ。しかし、日本語との対応は 1 対 1 ではなく、単純に対応させることは出来ない。表 3 に示す例は、「を」のシンハラ語への対応であるが、このように多くの「を」に対する格助詞がある。しかも名詞の単数/複数、定/不定、によって異なる。動詞の baseRule を用いれば、表 3 の 1~6 を区別することはできるが、

名詞の単/複、定/不定の区別は一般には困難を伴う。機械翻訳用の何らかの規則を作り上げていくことは大きな課題である。

## 4 おわりに

機械翻訳エンジン jaw によって、5 つの言語（中国語を除けばこれまでにほとんど取り扱われてこなかった言語）への翻訳システムを構築することができた。もちろん未だパイロットシステムの段階であるが、展望は得られたと考えている。今後は、まず第 1 に言語規則の分析・整備を進めること、そしてそれらの規則をシステム上に実現できるように、必要となるシステムの機能強化・整備を続けていく予定である。

## 参考文献

- [1] 日本語から多言語への翻訳エンジン jaw: 宇野修一、福本真哉、松本忠博、池田尚志(2005, 言語処理学会第 11 回年次大会発表論文)
- [2] 日本語-手話機械翻訳システム(jaw/SL)構築の試みと翻訳実験: 谷口真代、吉田鑑地、田中伸明、伊佐治和哉、松本忠博、池田尚志(2005, 言語処理学会第 11 回年次大会発表論文)
- [3] 日中機械翻訳システム jaw/Chinese における取立て表現の翻訳処理: 田中伸明、卜朝暉、浅井良信、宇野真哉、伊佐治和哉、松本忠博、池田尚志(2005, 言語処理学会第 11 回年次大会発表論文)
- [4] パターン変換型機械翻訳エンジン jaw に関する研究: 今井啓允(2004, 岐阜大学工学研究科修士論文)

	日本語	シンハラ語	格		
			単数・定	単数・不定	複数
1	本を愛する	<i>pothvalata adarakaranawa</i> පොත්වලට ආදරයකරනවා	<i>ata</i> අට	<i>akata</i> අකට	<i>valata</i> වලට
2	子供を連れて行く	<i>lamayava ekka yanawa</i> ලමයාව එක්ක යනවා	<i>ava</i> අව	<i>ekva</i> එක්ව	<i>anwa</i> අක්ව
3	橋を渡る	<i>palamen godawenawa</i> පාලමෙන් එගොඩවෙනවා	<i>en</i> එන්	<i>akin</i> අකින්	<i>valin</i> වලින්
4	ご飯を食べる	<i>bathØ kanawa</i> බත්Ø කනවා	<i>Ø</i>	<i>ak</i> අක්	<i>a</i> අ
5	犬を飼う	<i>ballek hadanawa</i> බල්ලෙක් හදනවා	<i>Ø</i>	<i>ek</i> එක්	<i>wo</i> ඕ
6	空を飛ぶ	<i>ahasehi piambanawa</i> අහසෙහි පියඹනවා	<i>ehi</i> එහි	<i>aka</i> අක	<i>wala</i> වල

表 3: 日本語の格助詞「を」とシンハラ語の格助詞の対応