E-014

日本語-ミャンマー語機械翻訳システム jaw／Myanmar における述語部構造の翻訳について

The Translation of Predicate Structure For Japanese-Myanmar Machine Translation System Jaw/Myanmar

マニンコウシン†、福本真哉†、池田尚志†

Ma Ngin Khaw Cing、Fukumoto　Shinya、Ikeda　Takashi

1. はじめに

我々は日本語からアジア諸言語への翻訳をおこなう機械翻訳エンジン jaw を開発している。jaw／Myanmar は日本語からミャンマー語への機械翻訳システムである。

ミャンマー語の言語構造についてはまだ研究が浅く、これまで十分に分析されていない。本稿では、述語部の構造を中心にミャンマー語の言語構造に関する我々の分析を述べ、さらに機械翻訳システム jaw／Myanmar における述語部の翻訳法について述べる。

2. ミャンマー語と日本語

ミャンマー語は日本語と同様の膠着語である。基本的に主語+目的語+述語の語順をとる。述語の後に様々な語が膠着してテンスやモダリティ等を表現する。このように日本語と非常に似た言語構造を持つが、異なる面も多い。

(1) 述語（動詞. 形容詞）には活用がない。

(2) 否定の機能語が動詞の前にくる。

(3)時制によって異なる機能語が使われる。

(4)モダリティを表現するための補助動詞がある。

(5)述語部の最後には必ず完成助辞が置かれる。

など、日本語と異なる多くの特徴を持っており、日本語と一対一に単純に対応させることは出来ない。名詞部も日本語と同様に基本的な文法関係を表現するため格助詞体系を持っているが、日本語の格助詞と単純に一対一に対応するわけではない。

3. ミャンマー語の述語部の構造

3.1 単独述語と複合述語

ミャンマー語の述語(動詞、形容詞)は単独の単語で動詞となる場合の外に、名詞+動詞、動詞+動詞のようにいくつかの単語が合わさって一つの動詞として使われる場合も多い。

■　単独述語

(1)　食べる。

စား	သည်။
食べる	完成助辞

(2)　花　は　美しい。

ပန်း	လှ	သည်။
花	きれい	完成助辞

(3)　美しい　花。

လှ	သော	ပန်း
きれい	連体接続助辞	花

■　複合述語

名詞と動詞による合わせ動詞

(4)　映画　は　面白い。

ရုပ်ရှင်	က	စိတ်ဝင်စားစရာ	ကောင်း	သည်။
映画	が	面白いこと	良い	完成助辞

3.2　述語部の構造

ここでは簡単のために単独述語の場合の構造について述べる。ミャンマー語の述語部の構造には「なければならない／かもしれない／してはいけない」などのモダリティを表現する場合の構造(複合構造)とそれ以外の単純構造の二つがある。

3.2.1　単純構造

動詞に膠着する助辞(機能語)を語順の観点から18 個のグループに分けた（図 1）。

アスペクト助辞＋否定“မ”+動詞+助辞部 1+
完成助辞+助辞部 2
助辞部 1：使役+許可+可能+判断 1 +希望+
義務+進行+過去、経験等+複数+
判断 2+判断 3+副詞呼応助辞+丁寧
助辞部 2：判断 4 +間投助辞

図 1　述語部の構造(単純構造)

† 岐阜大学工学部

- 否定の助辞は動詞の直前に置かれる(5)。
- 複数の助辞“ကြ”は主語が生物で複数である場合に呼応して使用される(6)。
- 「はじめる、だす、かける、続く」のアスペクト助辞は動詞の前に置かれる(7)。
- 副詞呼応助辞は「まだ、もう、べき」などの副詞の表現がある場合に呼応して使用される(8)。
- 完成助辞はミャンマー語に特徴的な助辞であり、文を完成する役割を果たす必須の助辞である。完成助辞にはテンス、否定、疑問などに応じて決まるものがある。

(5)　赤ん坊　は　御飯　を　食べ　ない。

ကလေး　က　ထမင်း　ကို　မ　စား　ဘူး။
赤ん坊　が　御飯　を　ない　食べる　完成助辞(否定)

(6)子供たち　は　学校　へ　行く。

ကလေး　များ　သည်　ကျောင်း　သို့　သွား　ကြ　သည်။
子供　たち　は　学校　へ　行く　複数助辞　完成助辞

(7)　電車　が　動き　始めた。

ရထား　က　စ　ထွက်　သည်။
電車　が　始める　出る　完成助辞

(8)　まだ　雨　が　降っている。

ယခုထက်ထိ　မိုး　ရွာ　နေ　သေး　သည်။
いまだに　雨　降る　ている　まだ　完成助辞

3.2.2　複合構造

「なければならない／かもしれない／てはいけない」などのモダリティを表現する場合にはモダリティ補助動詞とモダリティ接続助辞が使われる(図 2)。それぞれ数十個ほどある。(図 2)の“過去の助辞”は主動詞ではなく、モダリティ補助動詞につく過去を表現するものである(11)。

否定“မ”+動詞+助辞部 1+引用接続副助辞+副動詞+モダリティ接続助辞+否定+モダリティ補助動詞+過去の助辞＋完成助辞+　助辞部 2

図 2　述語部の構造(複合構造)

(9)書か　なければ　なら　ない。

မ　ရေး　လို့　မ　ရ　ဘူး။
ない　書く　モダリティ接続助辞　ない　できる　完成助辞(否定)

(10)　食べる　とは　かぎらない。

စားမည်　လို့တော့　ပြော　လို့　မ　ရ　ဘူး။
食べる　とは　話す　モダリティ接続助辞　ない　できる　完成助辞（否定）

(11)　食べる　とは　かぎらなかった。

စားမည်　လို့တော့　ပြော　လို့　မ　ရ　ခဲ့　ဘူး။
食べる　とは　話す　モダリティ接続助辞　ない　できる　過去の助辞　完成助辞（否定）

4.　jaw／Myanmar による述語部の翻訳法

4.1　jaw システムの概要

jaw は命題部分と機能語部分の二段階に分けて翻訳を行う。命題部分の翻訳規則は、日本語の係り受け木構造のパターンとそれに対応するミャンマー語の表現構造である。用言後接機能語部分の翻訳は命題部分の翻訳規則とは別に機能語翻訳テーブルを介して行う。

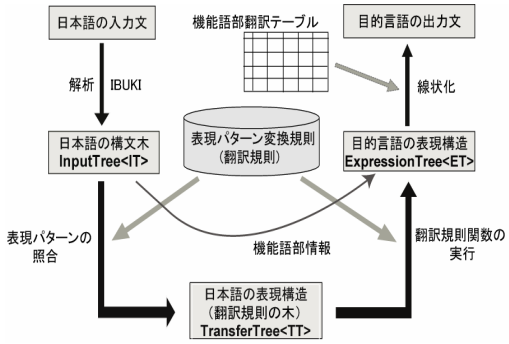

図 3　システム概要

4.2 述語部の翻訳規則

日本語の用言後接機能語を我々は使役等、時制等、判断等、接続の四つの要素に分類している。本稿では判断等に関する用言後接機能語の翻訳について述べる。

判断等に関する用言後接機能語のうち、新聞記事中で出現頻度が高かった表現[4 高木、他]を分析した。その結果、一対一で翻訳できるものは 38 語のみであった(表 1)。一対多の場合として「だろう」「らしい」についての翻訳規則を表 2 に示す。また、「かもしれない」に対応するミャンマー語は図 2 とも異なる特殊な構造をとる(表 3)。主動詞 V

や補助動詞 HV が繰替えし表現され、また、主動詞が未来か非未来か、肯定か否定か、「たい」がつくかなどによって対応する助辞表現が異なってくる。

さらに、主語が一人称であるか否か、表 3 のどの翻訳規則の区分であるかによって対応する完成助辞も異なってくる(表 4)。

機能語	否定	使役	許可	可能	判断 1	希望	義務	進行	過去、経験	複数	判断 2	判断 3	副詞呼応
てみたい					ကြည့်	ချင်							
べき													သင့်
なければならないかった	မ												

機能語	引用接続副助辞	副動詞	モダリティ接続助辞	否定	モダリティ補助動詞	過去の助辞	完成助辞	判断 4	間投助辞
ましょう								ကြစို့	
とはかぎらない	လို့တော့	ပြော	လို့	မ	ရ				
ではないか								မဟုတ်လား	
なければならなかった			လို့	မ	ရ	ခဲ့			

表 1　用言後接機能語の翻訳規則例(一対一の場合)

機能語表現「だろう、らしい」				翻訳規則
だろう	推量	V+だろう+と思った。		Φ
		V(た)　+　だろう	過去	V+သည်+ လို့ထင်သည်။
		V(る)　+　だろう	現在または習慣	
			未来	V+မည်+ လို့ထင်သည်။
		V+ない+だろう		မ　+V+ ဘူး+ လို့ထင်သည်။
		形容動詞+だろうじゃない		သည်+မဟုတ်လား။
	確認	V(た)　+　だろう	過去	V+သည်+ မဟုတ်လား။
		V(る)　+　だろう	現在または習慣	
			未来	V+မည်+ မဟုတ်လား။
		V+ない+だろう		မ　+V+ ဘူး+　 မဟုတ်လား။
	疑問	V(る)+だろうか	現在または習慣	V+ သလား+　 မသိ
			未来	V+ မလား+　 မသိ
		V+た+だろうか		V+ သလား+　 မသိ
		WH 疑問詞+...　+ V(る)+だろうか	現在または習慣	V+ သလဲ+　 မသိ
			未来	V+ မလဲ+　 မသိ
		WH 疑問詞+...　+　V(た)+だろうか		V+ သလဲ+　 မသိ
		V+ない+だろうか		မ　+V+ ဘူးလား +　 မသိ
		どうだろう、なだろう		ဘယ်လိုလဲမသိ、ဘာ လဲမသိ
らしい		V（る）+らしい	現在または習慣	V+သည့် ပုံရသည်။
			未来	V+မည့် ပုံရသည်။
		V(た)　+らしい		V+သည့် ပုံရသည်။

表 2　一対多になる翻訳規則例　「だろう、らしい」

「K＝かもしれない」に関する表現			ミャンマー語への翻訳規則 「HV＝ဖြစ်」	
肯定	未来	V+る+K	V+ လျှင် + V	①
		V+たい+K	V ချင်+လျှင်+V ချင်	②
	非未来	V+た+K	V ခဲ့+ လျှင် + V ခဲ့	③
			V+ခဲ့+တာ+HV+လျှင်+HV	
		V+たかった+K	Vချင်ခဲ့+လျှင်+ V ချင်ခဲ့	④
		V+ている+K	V+ နေတာ+HV+လျှင်+HV	
		V+ていた+K	V+နေခဲ့တာ+HV+လျှင်+HV	
否定	未来	V+ない+K	V+ချင်မှ+V	⑤
		V+たくない＋K	မ+V+ချင်တာ+HVလျှင်HV	⑥
	非未来	V+なかった +K	V ခဲ့+ချင်မှ+V ခဲ့	⑦
		V+たくなかったK	မ+Vချင်ခဲ့တာ+HVလျှင်HV	⑧
		V+ていない+K	မ+V+နေတာ+HVလျှင်HV	
		V+ていなかった+K	မ+V+နေခဲ့တာ+HVလျှင်HV	

表 3「かもしれない」に関する翻訳規則

主語区別	翻訳規則区分	完成助辞
「私」	①②⑤	完成助辞(未来) မည်
	③④⑥⑦⑧	完成助辞(未来) လိမ့်မည်
「私」以外	完成助辞(未来) လိမ့်မည်	

表 4 「かもしれない」に対する完成助辞

機能語	表層の形式で訳し分け		デフォルト固定訳		正解率
	正解	誤	正解	誤	
だろう	12	1	22	5	85%
らしい	32	0	5	3	92.5%
かもしれない	36	1	3	0	97.5%
そうだ(伝聞)	29	2	8	1	92.5%
にちがいない	23	0	15	2	95%
合計	132	4	53	11	92.5%

表 5 翻訳規則の評価

6. 翻訳規則作成と評価

判断等に関する機能語で一対多の対応をするもののうち出現頻度が高いもの 5 語を選んで翻訳規則を作成した。表 2 の例にみるように表層の統語形式で訳し分けの規則が書ける場合と「習慣を表すのか未来を表すのか」のように意味解析をしないと訳し分けできない場合がある。表層の統語形式で訳し分けできない場合には統計的に推察して固定訳を与える翻訳規則とした。5 機能語の各 40 文づつを新聞記事から任意に抽出し、翻訳規則を評価した(表 5)。表層の規則を用いて翻訳できたものは 68%、デフォルトの固定訳の翻訳規則を用いたものは 32%であった。正訳率はそれぞれ 97%、82.8%、で全体としては 92.5%であった。

７．おわりに

ミャンマー語の述語部の構造を分析し、語順を決め、日本語の用言後接機能語(判断等)に関する翻訳規則を作成した。文脈や意味による判断が必要な場合に対してはデフォルトの固定訳を与えた。全体としては 85%の正解率を得る事ができた。意味、文脈を必要とする場合の固定訳でない翻訳規則については今後の課題である。また、一対一になる場合の翻訳規則と一対多の場合の一部のついては jaw／Myanmar に実装したが、未実装のものも残っており、今後の課題である。

参考文献

［1］森山卓郎、仁田義雄、工藤浩『モダリティ』岩波書店（2000）

［2］大野徹『日本語ビルマ語辞典』大学書林（1995）

［3］加藤昌彦『エクスプレス　ビルマ語』白水社（1998）

［4］高木優紀江、林誠悟、池田尚志『大規模コーパスにおける文パターンの分布調査』FIT 情報科学技術フォーラム講演論文集(2002)