

2F-5 マレーシア語日本語の機械翻訳システム

ノーシャイダン カディール、高田正之、小谷善行、西村恕彦
(東京農工大学工学部電子情報工学科情報工学講座)

1. はじめに

マレーシア語から日本語への機械翻訳システムを作成した。マレーシア語はマレー語族に属するが、格変化や時制がないなど、中国語に似た性質をもっている。システムは、基本的には構文トランスファ(変換)方式によるものである。つまり、入力言語を解析し、その構文構造を得る。得られた構文構造を変換規則を用いて目的言語の構文構造に変換し、そしてその構造に対応する文を生成する。しかし、従来の構文トランスファ方式では、入力言語を意味解析せずに、構文構造を変換し、また目的言語を意味合成せずに文を生成するのが大きな弱点である^[1]。これは、構文解析だけでは文の適格性が決められないからである。そこで、より正確な翻訳ができるようにするために本システムでは、格文法方式を利用してその文が意味しているところの概念を明確にする意味解析も行うようにした。

2. 辞書と知識ベース

辞書は、(1)マレーシア語の単語辞書、(2)マレーシア語日本語の単語変換辞書、(3)日本語の単語辞書の三種類がある。

(1)は、名詞、動詞、形容詞、副詞などの品詞の辞書である。また、名詞の場合は、その名詞の意味属性、動詞の場合は、その動詞の概念などが記述される。この辞書は、マレーシア語の文を構文解析と意味解析するとき使用される。

(2)は、マレーシア語の単語に対応する日本語の単語の辞書である。この辞書は、マレーシア語の構文構造を日本語の構文構造に変換するとき使用される。

(3)は、日本語の名詞の意味属性と動詞の概念辞書である。この辞書は、マレーシア語を日本語の単語に変換するとき、日本語の意味合成をするときに使用される。

知識ベースは、名詞の意味属性の関連を意味ネットワークで表現されたものである。この知識ベースは、マレーシア語の文を意味解析するとき、マレーシア語日本語の単語に変換するとき、日本語の意味合成をするときに使用される。

3. システムの構成

システムの構成を図1に示す。システムは、①構文解析・意味解析、②構文構造の変換・日本語意味合成、③日本語単語合成の三つの段階に分かれる。

①構文解析・意味解析は、文脈自由文法で記述されたマレーシア語の構文規則を用いて、入力文を構成する単語の品詞の並びが妥当な並びになっているかどうかをトップダウン解析法によって行う。それと同時に、構文解析の情報、辞書(1)の名詞の意味属性と動詞の概念の情報、知識ベースの情報から動詞を中心にして、その文の深層格を決める。

結果として、入力文の構文木及び、動詞を中心にした深層格のリストが作り出される。

②構文構造の変換・日本語意味合成は、①から得られた構文木を変換規則によって、日本語の構文木に変換しながら動詞、形容詞などの活用形を合成する。それと同時に、①から得られた深層格のリストの情報、辞書(3)の情報、知識ベースの情報から、対応する日本語の動詞を中心にして意味を合成しながら妥当な助詞を決める。

結果として、日本語の構文木が作り出される。

③日本語単語合成は、②から得られた日本語の構文木から日本語の単語を合成する。

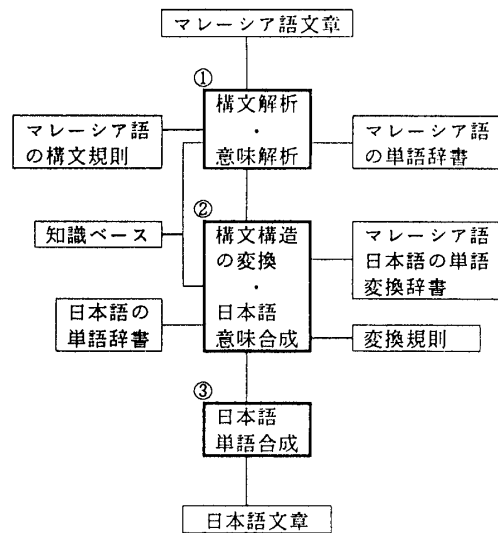


図1 システムの構成図

4. 翻訳過程の例

ここで、翻訳過程の例を、次の例文で示す。

入力文: saya pergi ke sekolah dengan kereta
(私) (行く) (へ) (学校) (で) (車)

この文を翻訳するために図2、3、4のような辞書と図5のような知識ベースが必要となる。

①は、この入力文を受け、マレーシア語の構文規則、図1の辞書と図5の知識ベースによって解析する。結果として図6のような構文木と深層格のリストが作り出される。

②は、①からの構文木を変換規則、図3と4の辞書によって日本語の構文木に変換しながら動詞の活用形を決める。それと同時に、①からの文の深層格のリスト、図4の日本語の動詞の概念辞書、図5の知識ベースを参照することによって妥当な助詞を付ける。結果として図7のような日本語の構文木が作り出される。

③は、②からの日本語の構文木から日本語の単語を合成する。結果として次の訳文が出される。

訳文: 私は車で学校へ行く

5. おわりに

本システムは、入力文の構文木を目的文の構文木に変換するまえに格文法方式による意味解析を行うので「私は毎日石を食べている」のような文法的には正しいが、意味をなさない文をチェックすることができている。また、目的文を生成するときにも目的語の意味合成も行うため、多義語の問題や正しい助詞を付ける問題などはある程度解釈することができている。

マレーシア語には時制がないため、日本語のように動詞などは活用しない。そこで、過去や現在などの状態を表すために時間を示す副詞、または助動詞で述べる必要がある。また、動作の様態を表すときにも動詞の様態変化がないため、助動詞で表現しなければならない。本システムでは、過去形や現在形などのような入力文をうまく翻訳ができるように、時間を表す副詞や助動詞を調べるようにしている。また、動作の様態を表す入力文が翻訳できるようにもした。

本報告では、格文法を用いたトランスファ方式によるマレーシア語から日本語への機械翻訳システムについて述べた。本システムは、NEWS (SONY) のK-Prolog言語で作られている。

6. 参考文献

- [1] 牧野武則、電子情報通信学会編：機械翻訳 ニューメディア技術シリーズ オーム社、1989
- [2] 淵一博監修、古川浩一、溝口文雄共編：自然言語の基礎理論 知識情報処理シリーズ4、1986
- [3] 石綿敏雄、西村彦彦、田中穂積、菊地光昭、藤崎哲助、高橋延匡編：日本語情報処理 近代科学社、1986

辞書 (1)		
名詞	品詞	意味属性
saya	(hinsi, n, _)	saya (isa, humpron, _)
sekolah	(hinsi, n, _)	sekolah (isa, loc, _)
kereta	(hinsi, n, _)	kereta (isa, vehicle, _)
動詞	品詞	概念
pergi	(hinsi, iv, _)	pergi (agt, hum, sub)
		pergi (goa, loc, ke)
		pergi (ins, vehicle, dengan)
前置詞		
ke	(hinsi, prep, [loc, time, direct])	
dengan	(hinsi, prep, [hum, ins])	

図2 辞書 (1) マレーシア語の単語辞書

辞書 (2)	
名詞	
hnn	(n(saya), 名詞(私))
hnn	(n(sekolah), 名詞(学校))
hnn	(n(kereta), 名詞(車))
動詞	
hvb	(iv(pergi), 自動五段(行く))

図3 辞書 (2) マレーシア語日本語の単語変換辞書

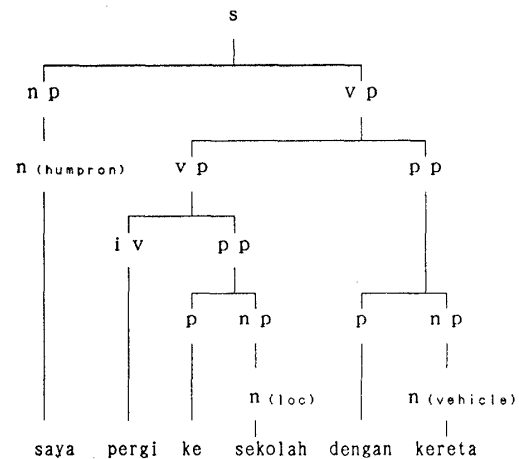
知識ベース	
animal	(isa, livingthing, _)
human	(isa, animal, _)
humpron	(isa, human, _)

図5 意味属性の知識ベース

辞書 (3)	
名詞	意味属性
私	(isa, humpron, _)
学校	(isa, loc, _)
車	(isa, vehicle, _)
動詞	概念
行く	(agt, human, は)
行く	(goal, loc, へ)
行く	(ins, vehicle, で)

図4 辞書 (3) 日本語の単語辞書

入力文の構文木



入力文の深層格のリスト

- [[pergi, agt, humpron, saya],
- [pergi, goal, loc, sekolah],
- [pergi, ins, vehicle, kereta]]

図6 入力文の構文木及び深層格のリスト

能動文

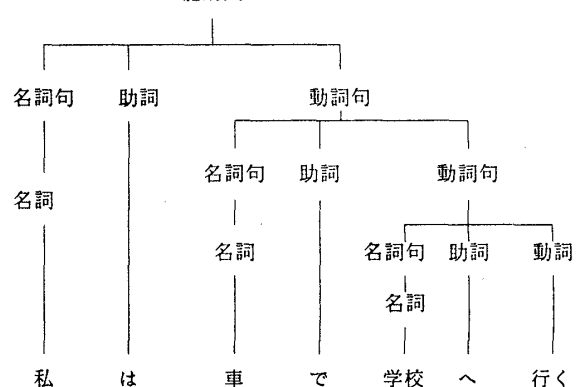


図7 日本語の構文木