

今井啓允、謝軍、池田尚志

岐阜大学工学部

1 はじめに

我々は、日本語を原言語としたパターン変換型機械翻訳システム j-aw (from Japanese into Asia & World) の開発を行っている。未だ第一ステップというべき段階であるが、現在のところ中国語、シンハラ語、ベトナム語を対象として開発を進めている。

本報告では、翻訳規則パターン辞書との照合処理、照合結果から翻訳規則を用いた表現構造の作成および目的言語文の生成を中心にその概要を報告する。

2 j-aw の概要

この機械翻訳システムでは、日本語と目的言語の対応規則を翻訳規則パターン辞書として日本語文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という形で表現している。

日本語入力文の係り受け構造 IT(InputTree) は、翻訳規則パターン辞書との照合により日本語パターンと対応する翻訳規則の木構造に変換される (TT:TransferTree)。次に TT 中の翻訳規則に対応づけられた翻訳プログラム (dll として実現している) を実行することで目的言語の表現構造木 (ET:ExpressionTree) が生成される (VC++ のオブジェクトとして実現している)。最後に、各 ET に対応した生成関数 (VC++ のメソッドとして実現している) を実行することで文が生成される。(図 1)

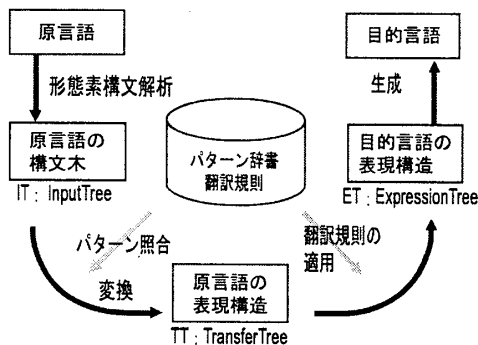


図 1: パターン変換型翻訳システム j-aw

このシステムは、翻訳知識の多くを翻訳規則パターンとして辞書に表現しているため知識の追加が容易である。また、パターン辞書との照合による TT の作成、TT に対応づけられた翻訳規則の実行による ET の生成、ET に対応づけられた生成関数の実行による目的言語の生成という方法は、翻訳知識データとその処理を完全に分離しており、任意の言語を目的言語として対象とすることができる。

3 パターン照合と翻訳規則の木

パターン照合に使用する辞書には、翻訳規則が対応づけられている日本語の表現パターンを記述したパターン辞書、単語の意味属性の包含関係を示した意味属性辞書、および機能語をグループ化した機能語条件辞書の 3 つがある。表現パターンには、BaseType、AdditionType の 2 種類がある。いずれも記述の中心となるキーワードを持つ。表 1 にこれらの例を示す。

表 1: 表現パターンの例

Type	文節番号	係り先番号	KW	自立語条件	機能語条件
Base	1	3		動物(熊)	「が」格
	2	3		場所(道)	「を」格
	3	0	歩く	Null	Null
Base	1	0	彼	Null	Null
AddCW	1	2	早く	Null	Null
	2	0		動作(走る)	Null
AddFW	1	2	の	人(私)	Null
	2	0		具体物(本)	Null
AddFW	1	2	ので	動作(出る)	Null
	2	0		動作(沈む)	Null

BaseType はキーワードが内容語 (自立語) であり、それをどのような文節が修飾するかを記述したパターンである。修飾する文節は特定の具体的な文節であったり、変数化されて意味属性で制約された文節であったりする。また、係りの表現においてどのような機能語 (格助詞相当語) が使われるかという制約条件 (機能語条件) が指定される。BaseType のパターンはいわゆる格構造ないし結合価パターンと同等である。意味属性辞書、機能語条件辞書が制約条件の記述に使われる。

AdditionType は、キーワードが内容語または機能語であり、それがどのような文節を修飾するかを記述したパターンである。AdditionType には Addition-CW と Addition-FW のふたつがある。CW 型は内容語がキーワードである場合、FW 型は機能語がキーワードである場合である。

日本語の入力文には、修飾・被修飾の関係の表現の他に、使役、受け身、時制、判断様相、取り立て、などに関する表現もある。使役・受け身などについては、BaseType のパターンを適切に書き換えることで対応できる。その他の表現については、ET 中に pMode, nMode などしかるべきメンバーを設け、そこにそれらの情報を書き込んでおいて、生成関数で対応することとしているが、现阶段の生成関数は、まだそれらへの対応はしていない。

BaseType の引数が入力文で省略されている場合は、Ellipsis という仮のパターンを対応させている。このパターンは ET においても仮のオブジェクト Ellipsis を作り出す。例として「熊がゆっくり道を歩いていた」を照合したときの TT を図 2 に示す。

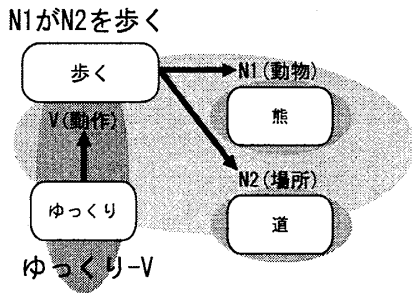


図 2: TT

4 表現構造

表現構造 (ET) は目的言語の文の一種の意味表現である。名詞、動詞などのクラスが定義されており、それを修飾する要素や前節で述べた pMode, nMode などがある。また、そのインスタンスが与えられたときに、それを目的言語の言語表現として作り上げるための関数 (生成関数あるいは線状化関数と呼んでいる) が、クラスのメンバー関数 (メソッド) として定義してある。これらのクラスの設計 (どんなクラスを設けるか、どんなメンバー変数を設けるか、どんなメソッドを与えるか) は、目的言語毎に設計することになる。

5 翻訳規則の木の実行による表現構造の作成

日本語パターンに対応する目的言語の表現として記述された翻訳規則は、TT から ET を作るプログラムに変換される。これはコンパイルされて翻訳規則のライブラリ (dll) となる。

TT の各要素には、それぞれ翻訳規則関数 (dll) が割り当てられている。これを根ノードからある再帰的アルゴリズムで実行することによって、目的言語の表現構造である ET が生成される。

図 3 に入力文「熊が道をゆっくり歩いていた。」に対する ET の例を示す。

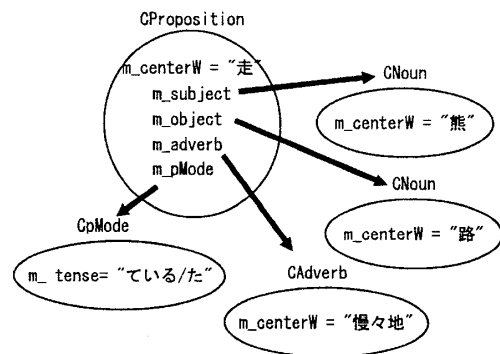


図 3: ET

6 表現構造からの文の生成

生成された ET の各ノードは、メンバ関数として生成関数 (線状化関数) を持っている。生成関数は、表現構造をたどって語順調整をしながら目的言語の翻訳文を生成する。線状化関数は各目的言語の文法規則を反映した目的言語ごとに必要な関数である。中国語の場合には、語順が「SVO」型であり、図 3 の ET の線状化関数を実行すると、翻訳文「熊慢々地走路」が生成される。

7 おわりに

パターン変換型翻訳システム j-aw の概要について述べた。現状は、未だ骨格的な文にしか対応していない。次の課題は、時制、判断様相、取り立てなどの表現の生成に取り組むことである。

参考文献

- [1] パターン変換型機械翻訳システムにおける日本語パターン照合処理に関する研究、河原、岐阜大学工学部応用情報学専攻修士論文 2001 年
- [2] パターン変換型翻訳システムの研究-照合処理に関する研究、今井、岐阜大学工学部応用情報学専攻卒業論文 2002 年