家族モデルを用いた文の分解に基づく日中機械翻訳システム†

任 福継†† 范 莉馨††
宮永喜一†† 栃内香次††

日本語文によく現れる文型，“けっして…ない”，“まるで…ようだ”などを中国では“慣用句型”と呼び，本論文では“常用文型”という語句を用いる．中国人は日本語を勉強する際，常用文型を非常に重視している．特に日本語文を翻訳する際，常用文型に対しては，その部分の詳しい文法分析を行わず直接中国語訳文を得るという方法を用いている．そして，このようにして得られた中国語訳文は，常用文型に対応して選択された補助語によりアスペクトとモダリティも満足させる．本論文ではこの方法に基づいて，文の分解による家族モデルを提案し，このモデルを用いる日中機械翻訳実験システムを述べる．家族モデルは父親モジュール，太郎モジュール，次郎モジュール，花子モジュールからなる．父親モジュールは翻訳全体の制御，とくに入力文を基本文と常用文型に分解する．太郎モジュールは，基本文の解析を行い，目的言語文法に従って語順の調整を行う．次郎モジュールは，常用文型の処理を行う．花子モジュールは補助語を推定する．最後に，父親モジュールによってすべての情報を総合して訳文を整理する．4つのモジュールが独立的に構築され，翻訳システムの改良などが容易に実現される．また，太郎，次郎，花子モジュールは相互間に情報交換がないので，並列処理ができる．実験により，本論文で提案した日中機械翻訳手法の有効性が確かめられた．

1. はじめに

現在，世界各地で機械翻訳の研究が盛んに行われ，いくつかの機械翻訳システムが実用化されている[1]．しかし，構文構造の複雑さ，表層形と意味の対応の複雑さ，原言語と目的語の表現方法の隔たりなどが原因で，正しい翻訳結果が得られないことが多い[3]．また，日本でも中国でも英語を主要な対象とした機械翻訳の研究開発は多いが[4)~9)]，日中両言語間の機械翻訳に関しては本格的研究が開始されたばかりであり，未開拓の部分がきわめて多い[13]．

一般に，日本語から中国語への翻訳において，自立語については容易に対応する中国語単語が得られる場合が多い．しかし，付属語については，中国語表層として独立かつ適切な単語表現がない場合や，付属語の各単語でなく全体として中国語表現に対応する場合などが多く，付属語の詳細な解析を行わなければ中国語訳文を得ることができない[16)~18)]．例えば，

例① 「原子はけっして物質の可分性の限界ではない」

例② 「金は王水にしか溶けない」

における“けっして…ない”，“しか…ない”がこのような付属語を含む表現である．中国ではこれらの語句を“慣用句型 (guanyong-juxing)”と呼んでいる[19),20)]．“慣用句型”は日本語における“慣用句”とは完全には一致しない．そこでわれわれは“慣用句型”の日本語訳として「常用文型」という語句を用いている[17]．

一般に中国では，日本語の学習に際してまず日本語のいくつかの格助詞を学び，その後は前述のような常用文型を中心に学習する．すなわち，これら常用文型については詳細な文法解析を行わず，直接中国語訳文を得ている．また，このようにして得られた中国語訳文は常用文型に対応して選択される補助詞によりアスペクトとモダリティも満足させる．

われわれはこのような中国における日本語学習の方法が日中機械翻訳に効果的であると考え，日本語文を基本文と常用文型に分解して翻訳する方式を提案する．翻訳システムは入力文を分解し，以後の翻訳を制御，管理する部分と，各部分の翻訳を行う部分からなる．この構造は家庭における日常の仕事の分担に類似しているとみなせるので，本論文ではこれを家族モデルという．本論文における家族モデルの構造を図1に示す．このモデルは父親モジュール，太郎モジュール，次郎モジュール，花子モジュールからなる．

基本文と常用文型の範疇については定説がないが，中国語側から見た日本語の理解，翻訳の便などを考えて選ばれている．例えば，上述の例文①では，基本文は“原子は物質の可分性の限界である”，常用文型は“けっして～ない”であり，例文②では，基本文は“金

† A Japanese-Chinese Machine Translation System Using a Family Model by FUJI REN, LIXIN FAN, YOSHIKAZU MIYANAGA and KOJI TOCHINAI (Faculty of Engineering, Hokkaido University).

†† 北海道大学工学部

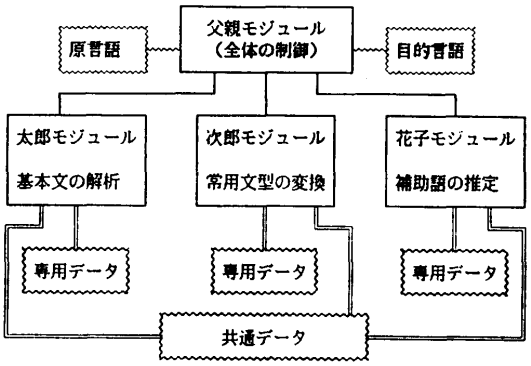


図 1 家族モデルの基本構成
Fig. 1 The structure of the family model.

は王水に溶ける", 常用文型は "しか～ない" である.

家族モデルによる翻訳方式では, まず, 父親モジュールによって日本語文を基本文と常用文型に分解する. そして, 太郎モジュールによって基本文の解析を行い中国語文法に従って語順を変換する. それと同時に, 次郎モジュールでは常用文型の変換, 花子モジュールでは小詞 (本論文で, 日本語格助詞に対応する中国語の補助語を指す) の推定を行う. 最後に, 父親モジュールによってすべての情報を総合して訳文を整理する. そして, 4つのモジュールが独立に構築され, 翻訳システムの改良, 新たな規則の追加・変更などを容易に実現する. これは段階的なシステムの構築・改良という方法論が取れることを意味している. また, 家族モデルでは, 太郎モジュール, 次郎モジュール, 花子モジュールはほとんど同時に並列的に処理できるので, 家族モデルは並列モデルということができる.

本論文ではこの家族モデルを用いた日中機械翻訳システムについて述べる. なお, 現在のシステムでは, 品詞自体の多義性を考慮していない. また処理の対象文は単文に限定されている.

以下, 第2章で常用文型と入力文の分解について述べる. 第3章で家族モデルによる日中機械翻訳について述べる. 第4章で実験結果および考察について述べる. 第5章では今後の課題について述べる.

2. 常用文型と文の分解

2.1 常用文型

日本語の常用文型を中国語に翻訳する際, 常用文型に対応する訳文関数によって直接中国語訳文を生成することができる[17]. 以下に常用文型の2例を示す.

例③ シリコンの生ゴムはまるで透明な飴のようだ. この例で "まるで～ [の] ようだ" が常用文型であり, その訳文関数は(1)式で与えられる.

AまるでBようだ → A宛如B一様 (1)

訳文: 有機硅生橡膠宛如透明的糖飴一様.

例④ 彼女は車を運転することができるに違いない. 例④は英日機械翻訳システム[23]でモダリティの複数ある例として述べられているものであり, 本システムにおける多重常用文型に対応する. この例で, "ことができる" と "に違いない" が常用文型であり, それぞれの訳文関数は(2)式と(3)式で与えられる.

…ことができる → [主] 能 (2)

…に違いない → [主] 肯定 (3)

訳文: 她 肯定 能 駕駛 汽車.

なお, 常用文型はその機能によって用言修飾要素と体言修飾要素の2種に分けられ, また, その構成形態から一語性と多語性とに分けられる. ここで, 一語性常用文型は語句が一個所に集中するものであり, 多語性常用文型は分離した複数の語句からなるものである. さらに, 呼応語の有無によっても分類される. ここで呼応語とは例①の "ない" のように常用文型の主部をなす付属語 (ここでは "けっして") と必ず組み合わされて出現する句型をいう. われわれは, 文献 19) に記載されている 451 個の常用文型を検討し, **表1**に示すような結果を得た.

2.2 文の分解

家族モデルでは, 日本語文を基本文と常用文型に分けて処理している. 文の分解は常用文型の抽出と常用文型を抽出した後の文の修正の2段階からなる.

常用文型を抽出するために, 付属語テーブルを設けている. その構造を**表2**に示す. ここで, キーワード

表 1 常用文型の類別および頻度
Table 1 The classifications and frequencies of guanyong-juxing.

付属語	呼応語	多義性	出現数	頻度 (%)		常用文型の例
1 個	有		92	20.4	94.5	しか～ない
	無		334	74.0		なければならない
		有	208	46.1	94.5	うちに
		無	218	48.3		けっして～ない
2個以上			25	5.5		～おろか～でさえ

表 2 付属語テーブルの例
Table 2 The attachment words table.

キーワード	常用文型の構成条件	変形文型	中国語表現	影響されるコード	訳文関数類別	MC
しか	用言否定形	用言なし	只	3009（否定→肯定）	③	—
まるで	体言＋のようだ 用言連体形＋ようだ	略	宛如…一様	—	④	—
うちに	—	—	—	—	—	100
……						

は常用文型の一部であり*，文の分解および常用文型の判断に使用される情報である．MC は，常用文型の修飾条件に対応して複数の中国語表現がある場合に使用する訳語条件テーブルへのポインタである．また，訳文関数類別欄には，対応する訳文関数の種類が記録されている[18]．

文の分解の手順として次のようにまとめられる．

（1） 付属語テーブルによる常用文型構造の確認

（2） 常用文型の抽出

（3） 常用文型が抽出された後の文の修正

以下，**図 2** の例を用いて文の分解の概要を述べる．

なお，常用文型の構成条件，訳文関数などはあらかじめ付属語テーブル（表 2 参照）に登録されているものとする．

入力文からまずキーワードを探索する．図 2 の文 1 の場合常用文型“あまり…ない”に対応して，“あまり”がキーワードとして登録されている．もし，入力文の中にキーワードがなければ，入力文の全体は基本文である．キーワードが発現された後，さらに常用文型の構造を確認する．例えば，文の中にキーワード“あまり”があるが，用言の否定形態がなければ，その文は常用文型ではない．本例では，形容詞“大きい”の否定形態“大きくない”が存在するので，常用文型“あまり…ない”であることが確認される．ついで，常用文型“あまり…ない”および常用文型に関連する要素“大きい”を抽出し，太郎モジュールの入力情報とする．最後に，常用文型を抽出した後の文の修正を行う．本例文では“あまり”の位置には修正の必要がないが，“ない”を抽出したところに形容詞の原形を復元しなければならない．復元の手段としては，文献 22) と同じく語幹・変化語尾テーブルを利用して行っている．修正は主要に次の二種がある．

①用言原形の復元

* キーワードは，一語性常用文型では常用文型それ自身であり，多語性常用文型ではその一部である．

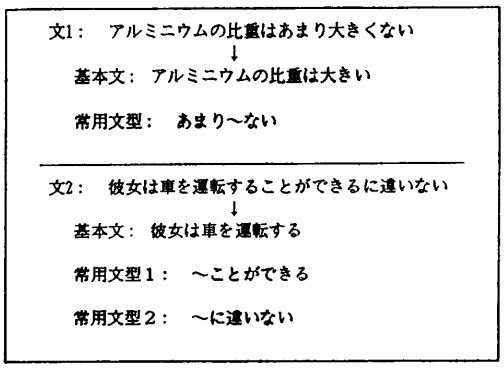


図 2 日本語文の分割の例
Fig. 2 Examples of partition of Japanese sentence.

②格助詞の付加

文 1 は用言原形の復元の例である．格助詞の付加の例は 3.5 節で述べる．

次に，常用文型が多重になっている場合に文の分解について述べる．

日本語文において多重常用文型は交差がないことがわかっているので，スタックを用いて文の分解と多重常用文型の処理を行っている[17),18)]．

図 2 の文 2 では，まず常用文型“ことができる”が先にスタックに入り，“に違いない”は後にスタックに入る．それゆえ，スタックの操作規則により，“に違いない”が先に，“ことができる”は後に処理される．もし，修正が必要であれば，1 つの常用文型の場合と同様にその修正を行う．

3. 家族モデルによる日中機械翻訳

3.1 太郎モジュールおよび基本文の解析

言語学では，中国語は代表的な“孤立語”であり，日本語は“膠着語”であり，日本語と中国語は異系統の言語であると言われている[14)]．日本語は述語文節を中心とし，他の文節の役割をその文節の格助詞で指定することによって文全体の意味を表現する．一方，中国語では語と語との関係は語順によって示される．

太郎モジュールでは，日本語の格構造と中国語の語順との対応関係を利用して，コード方式[13]を用いて，基本文の翻訳を行っている．日本語文の解析手法については既に多数の論文が発表されているが[1),2),10),22)]，本システムで用いる解析手法もそれらと大きく異なるものではなく，格文法を利用して，文節端の助詞を中心に文の基本成分を決定している．

一方，中国語文側では，意味文法[15]を利用して中国語文の基本成分の語順を決定している．**図3**に中国語文の基本成分と構造を示す．

なお，コード方式は変換と解析の融合方式[22]と類似した方式であり，コードは次郎モジュールにおける処理との結合しやすさを考慮して定められている．以下に太郎モジュールにおける基本文翻訳の概要を述べる．

まず，翻訳しようとする日本語文を文の意味的な基本単位（これをコード元素という）に分解し，これを格助詞と活用形を中心に解析し，構文および意味属性を表す記号を付加した2つ組（これをコードという）を生成する．なお，コード元素すなわち文の意味的な基本単位としては文節を採用している[25]．以上の処理により文はコードの並びで表され，これをコード列という．

次に，得られたコード列を中国語のコード列に変換する．ここでは中国語の構文則にしたがって語順の変換が行われ，また，各コード元素は対応する中国語コード元素に変換される．

最後に，このコード列から中国語訳文を生成する．このとき，コードの性質にしたがって補助語の付加などの操作を行う．

本論文では，コードは正コードと副コードに分けられている．この二種類のコードには明らかな境界がないが，普通，時制・態・様相などを表すものを副コードとしている．付録1に本実験システムにおけるコードを示す．ここで，No.1～No. 43 は正コードであり，No. 44～No. 60 は副コードである．

日本語文のコード列を中国語文のコード列に変換するために，語順変換規則が用意されている．ここで，語順変換規則は，日本語文のコード列* によって中国語語順を決定するルールである．付録2に中国語基本文型のコード列と日本語の対照関係の一部を示す．本

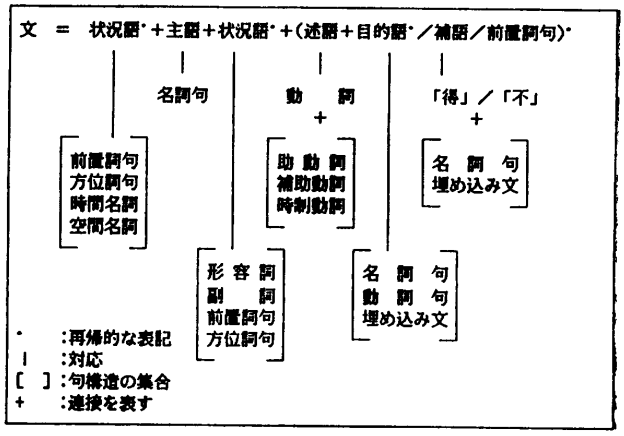


図3 中国語文の基本成分と構造
Fig. 3 Prototype structure and elements of Chinese sentence.

実験システムでは，語順変換規則を大規則と細規則に分けている．大規則は中国語文の主要素（例えば：主語，述語［謂語］，目的語［賓語］）の語順を決定する，すなわち基本文型の語順を決定する規則である．細規則は同一要素中の語順の確立，または FTM（付録1に示す）など任意成分の語順を決定する規則である．以下に，いくつかの大規則と細規則の例を述べる．

はじめに大規則の例を示す．

大規則1：[主語，目的語，述語]→（主語，述語，目的語）

大規則2：[主語，述語]→（主語，述語）

大規則3：[主語，目的語1，目的語2，述語]→（主語，述語，目的語1，目的語2）

ここで，[]はその内部の要素の順序が任意であることを意味し，()はその内部の要素の順序が固定されていることを意味する．例えば，大規則1によって，日本語文のコード列の中に，主語，目的語，述語要素があれば，主語→述語→目的語という中国語文の語順を決定することができる．

次に細規則の例を示す．

細規則1：[ΣSUB, ΣAGT]→(ΣSUB, ΣAGT)

細規則2：[ΣTIM, ΣSPA]→(ΣTIM, ΣSPA)

細規則3：[FSP, TSP]→(FSP, TSP)

ただし，ΣSUB＝SUB1, SUB2

ΣAGT＝AGT1, AGT2

ΣTIM＝TIM1, TIM2, FTM, TTM

ΣSPA＝SPA1, SPA2, FSP, TSP

上述のように，大規則により主語要素の語順が決められるが，もし主語要素を担当できるコードが複数個

* これは必ず付録1に示したコードの組合せでなければならない．この範疇以外はいまのところ処理できない．

あれば，細規則1によって，それらのコードの順序を決める．すなわち ΣSUB が必ず ΣAGT の前に位置しなければならない．“象は鼻が長い（象的鼻子長）”はこのような例である．

細規則2は時間に関する状況語は必ず空間に関する状況語の前に位置しなければならないことを示し，また，細規則3は状況語の空間起点は必ず状況語の空間終点の前に位置しなければならないことを示している．

次に，例文により翻訳処理の流れを説明する．

例⑤　昨日，私は文房具店でテープを買った．

この文を解析して得られるコードを**表3**に示す．ここで，コード記号欄の () 内は各コード記号の持つ意味を表す．

ついで，表3のコード集合に対し，語順変換規則を用いて中国語文として正当な順序を求め，さらにコード元素の変換を行って，**表4**に示す中国語コード列を生成する．最後に，コードによって定められる補助語を付加して中国語文を生成する．その結果を以下に示す．

中国語訳文：我昨天 {在} 文具店買〈了〉磁帯．

訳文中で，{} の部分は格助詞“で”に対応する中国語補助語である．

3.2 次郎モジュールおよび常用文型の処理

第2章で述べたように，日本語の常用文型を中国語に翻訳する際，常用文型の訳文関数を用いて，中国語訳文を生成することができる．しかし，常用文型には多義性が存在するので，それを解消する必要がある．

次郎モジュールでは常用文型の関連部分の意味属性を利用して多義性を解消し，意味属性の細分類によって補助語の微細な差を区別でき，常用文型の意味を表す自然さと柔軟性のある中国語文を生成している[17),18)]．

なお，われわれは意味属性を通常の用法より広い意味で用いており，感情，時間，状態などの他，品詞の属性，形態なども含めている．意味属性を求めるために，品詞の意味分類情報の登録が必要となる．本実験システムでは，動詞の意味分類は IPAL[29)] にしたがい，体言の意味分類は文献27) などを参照して設定した．

多義性を持つ常用文型の訳文関数を式(4)に示す．

$$\text{常用文型L}\begin{cases}\text{関数}1 & \langle\text{条件}1\rangle \\ \vdots & \vdots \\ \text{関数}n & \langle\text{条件}n\rangle\end{cases} \qquad (4)$$

表3 例⑤の日本語コード
Table 3 Code of Japanese sentence.

コード記号	コード元素
TIM (時間)	昨日
SUB (動作の主体)	私
SPA (動作の場所)	文房具店
OBJ (動作の対象)	テープ
PRE (動作)	買う
0011 (動作の過去状態)	買う

表4 例⑤の中国語コード列
Table 4 Code string in Chinese.

コード記号	コード元素
SUB	我
TIM	昨天
SPA	文具店
PRE	買
0011	〈了〉
OBJ	磁帯

ここで，条件 i は訳文関数 i に対応する常用文型Lの関連部分の意味属性である．これを用い，常用文型の多義性を解消できる[17),18)]．

3.3 花子モジュールおよび補助語の推定

日本語の格助詞に対応して中国語文に現れる語を本論文では補助語という．一般に，日本語の格助詞に対応する中国語の補助語は一対多である．例えば，

例⑥　彼は　飛行機で　東京に　行く
　訳文：他　乗　飛機　去　東京

例⑦　彼は　食堂で　餃子を　食べる
　訳文：他　在　食堂　吃　餃子

の中で格助詞“で”に対応する補助語はそれぞれ“乗”と“在”である．前者は交通手段を表し，後者は動作場所を表す．さらに，

例⑧　彼は　自転車で　東京に　行く
　訳文：他　騎　自行車　去　東京

では，格助詞“で”に対応する補助語は“騎”である．例⑥と例⑧では，同様に交通手段を表すが，下位区分によって異なる補助語で表されている．これは中国語の特徴の1つである．それゆえ，このような複数の補助語候補から正しい補助語を推定することが問題になる．

上述の問題を解決する方法の1つは，太郎モジュールで異なる補助語に対応して異なるコードをつくることである．しかし，教科書，文献，文法書など約12,000文から格助詞を抽出し，対応する補助語につい

て考察を行った結果，格助詞“で”の補助語候補数は23個，格助詞“に”の補助語候補数は24個であった．したがって，これらのために多数のコードが必要となり，中国語のコード列の変換が複雑になって，翻訳の効率も悪くなる．

一方，例⑥～⑧からわかるように，格助詞“で”の補助語は異なるが，その語順は同じである．この特徴を検討して，本システムでは格助詞に対する補助語推定を，太郎モジュールからはずし，花子モジュールで行う．

一般に，

体言 i　格助詞 i …用言 j　（5）

（5）式のような日本語文では，{m(体言 i)，m（用言 j)} の組で補助語を推定できる．ここで，$m(x)$ は x の意味属性を表す．

花子モジュールには，格助詞補助語関連表（KAHOT）が用意されている[28]．補助語を推定する際，体言あるいは用言の意味属性などの本手法に必要な情報をKAHOT に登録する．例として格助詞“で”に関する KAHOT の構造を**表5**に示す．

表5 格助詞“で”に対する KAHOT の構造
Table 5 Organization of the KAHOT.

格助詞（で）	類別（A，B）	補助語候補数（15）		
順番	M（体言 i）	M（用言 j）	補助語	コード記号
1	場　所　0100	—	在	T001
2	交通手段（3類）0301	—	騎	T002
3	交通手段（2類）0302	—	乗	T003
4	交通手段（1類）0303	—	坐	T004
5	原　因　0700	—	因	T005
6	お　金　0801	—	花	T006
⋮				

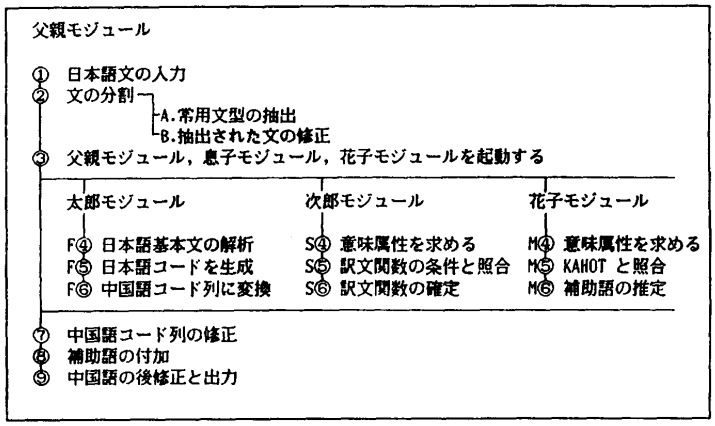


図4 日中機械翻訳システムの処理概要
Fig. 4 The process of Japanese to Chinese translation system.

3.4 日中機械翻訳システムの処理概要

現在のシステムでは，日本語文入力は文節分かち書きで行い，文節識別のための特別な処理は行っていない．**図4**にシステムの処理の流れを示す．

以下に例を用いてその処理概要を説明する．

例⑨　昨日，私は文房具店でテープしか買わなかった．

まず，文の分解を行う．

この例では常用文型は“しか…ない”である．また，“しか”をキーワードとして付属語テーブルに登録した．そこで，“しか…ない”という常用文型が確認された．それを抽出してから，前述のように抽出した後の文の修正を行う．本例文では二か所の修正が必要である．

①　キーワードのところに格助詞の付加

日本語において，“しか”があれば，本来その位置にあった格助詞は省略されることが多い．それで，“しか”を抽出したところで格助詞を付加しなければならない．後の動詞“買う”の他動詞の特性によって格助詞“を”を付加する．

②　“なかった”のところに用言の復元

ここで，2.2 節で述べたように，用言の復元を行うが，本例文では，“なかった”は“ない”の過去時態なので，この過去時態を用言に付加する必要がある．すなわち，復元の動詞は“買った”である．

以上により，基本文“昨日，私は文房具店でテープを買った”と常用文型“しか…ない”を得た．

ついで，太郎モジュールで基本文の解析，次郎モジュールで常用文型の変換，花子モジュールで補助語の推定を行う．

太郎モジュールと花子モジュールの入力は同じく，基本文“昨日，私は文房具店でテープを買った”であるが，次郎モジュールの入力は常用文型および常用文型で修飾された要素である．本例文では“しか…ない”および“テープ”である．これらの要素を利用して常用文型の多義性を解消できる[17]．

太郎モジュールで基本文を解析して，中国語のコード列を得る．以上の結果を表4に示す．

次郎モジュールでは，訳文関数の条件と照合して訳

文関数を推定する．常用文型“しか…ない”は多義性がないので，直接に登録された訳文関数をとる．これは“只”であり，修飾対象は“テープ”であるので，“只”の語順は動詞“買う”の直前である．

花子モジュールでは，格助詞に対応する中国語補助語を推定する．本例では，格助詞“は”と“を”の補助語が ϕ であり，“で”の補助語は，“文房具店”の意味属性によって“在”を推定した．ここで，“ϕ”は補助語なしを意味する．

さらに，父親モジュールでは，以上の情報を利用して，格助詞の補助語を付加し，常用文型の中国語表現を追加して，最後の中国語訳文を生成する．

訳文：我 昨天 在 文具店 只 買了 磁帯.

3.5 家族モデルによる日中機械翻訳の特徴

家族モデルによる翻訳方式の主要な特徴が，①日本語文を基本文と常用文型に分けて，基本文の格関係を利用して解析を行い，日本語文のコード列を生成し，そして中国語文の構造に従って中国語コード列に変換すること，また，②常用文型についてあらかじめまとめた訳文関数で直接に中国語を生成し，多義性を持つ常用文型に対しては，その関連部分の意味属性を利用して多義性を解消することの2点であることは既に述べた．ここでは他の特徴を示す．

（1） 文の分解

家族モデルでは，常用文型の構造があらかじめシステムに組み込まれているので，文の分解が容易になる．また，文の分解によって，語と語との関係の間に不要な多義性の発生することが防止されている．特に中国語への変換ではこの点が非常に有効と考えられる．

（2） 解析と変換

解析主導型翻訳方式および変換主導型翻訳方式に関連して既に多数の提案がなされている[1),3),5),11),12),22),24)]．本論文の家族モデルは従来の手法を融合し日中両言語の特徴を利用した翻訳を行う手法である．すなわち，本方式では，基本文に対しては日本語文の解析処理と日中変換処理を融合し，また，常用文型に対しては訳文関数の条件を照合して直接に中国語文への変換を行っている．

それで，本方式は解析と変換の融合方式[22)]もしくは多段翻訳方式[5)]と類似した方式であると言うことができる．

（3） モジュールの相互独立性

家族モデルでは，4つのモジュールが相互独立に構成されており，基本文のコードおよび中国語コード列の変換規則，常用文型の訳文関数，補助語を推定するKAHOTなどは各自独立であるので，システムの改良，新たな規則の追加・変更が容易に実現する．

（4） 並列処理の可能性

家族モデルでは，太郎モジュール，次郎モジュール，花子モジュールは情報の変換がないので同時に並列処理できる．それで，並列処理の機械翻訳システムの構築が期待される．

4. 翻訳実験と考察

われわれは本論文で提案した家族モデルによる日中機械翻訳実験システムを構築した．実験システムでは基本文の解析用のコードは60種である．また常用文型は63種登録されており，そのうち多義性のある常用文型は29種である．常用文型を持つ約500文を用いて実験を行った．最後の結果は正翻訳率96%となった．なお，ここで正しい訳文とは，文法的にも意味的にも完全に正しいものと，厳密には語の用法などに誤りがあるが，人間が読んで理解する上では全く問題のないものとの両方を含み，前者は73.6%，後者は26.4%である．

格助詞の補助語の推定実験は，翻訳システム全体と分離して行い，評価データは補助語のみを指標として行った．格助詞を含む1200文の実験から，補助語を正しく推定できた正翻訳率が約95%であった．

実験によって本論文で提案した翻訳手法の正確性と有効性が確認された．多数の文を対象とし，また，広範な専門分野にわたる大規模な実験は行っていないが，教科書，文献などに記載されているような代表的な常用文型はほとんどすべて網羅しているので実用的には十分であると考える．

5. おわりに

中国語を母語とする人間が，日本語を勉強する際，常用文型が非常に重視されている．特に，日本語から中国語への翻訳を行うとき，常用文型をまとめて翻訳することにより容易に中国語訳文を生成できる．

本論文では，このような方法に基づいて文の分解による家族モデルを提案し，さらにこのモデルを用い日中機械翻訳実験システムを構築した．実験により本論文で提案した方式の有効性と正確性が確かめられた．

本論文では処理する対象は単文である．また，常用文型中で呼応語が省略されたもの[21),26)]については処理できない．そして，日本語文を基本文と常用文型と

に分離した後に残る部分に関しても未検討である．また，現在のシステムでは，品詞自体の多義性を考慮していないので，システム全体としての翻訳能力はまだ十分ではない．これらの問題点の解決，および意味属性の範疇のより詳細な検討などが今後の課題である．

謝辞 日ごろ有益なご討論，ご助言をいただく研究室各位および北海道工業大学鈴木康広助教授，北海学園大学荒木健治助教授に感謝いたします．また，実験システムの構築を進めるにあたり種々ご助言をいただいた斎川勝男技官に感謝の意を表します．

参考文献

1) 野村浩郷，田中穂積：機械翻訳，bit 別冊，共立出版社 (1988).
2) 辻井潤一：機械翻訳システム，スペクトラム，Vol. 1, No. 11, pp. 48-57 (1988).
3) 古瀬 蔵，隅田英一郎，飯田 仁：変換主導型機械翻訳の実現手法，情報処理学会研究会報告，NL-80-8, No. 93 (1990).
4) 田中穂積：解析から合成までを融合した英日機械翻訳システム，日経エレクトロニクス，8月号，pp. 275-293 (1983).
5) 池原 悟，宮崎正弘，白井 諭，林 良彦：言語における話者の認識と多段翻訳方式，情報処理学会論文誌，Vol. 28, No. 12, pp. 1269-1279 (1987).
6) 高松 忍，西田富士夫：動詞パターンと格構造に基づく英日機械翻訳，信学論 (D)，Vol. J64-D, No. 9, pp. 815-822 (1981).
7) 石崎 俊，内田裕士：多言語間翻訳のための中間言語について，情報処理学会研究会報告，NL-70-3, No. 6 (1989).
8) 劉 涌泉：机器翻譯浅説，中国語文，12月号，pp. 575-577 (1958).
9) Liu, Z.: An Introduction to JFY-11 English-Chinese Machine Translation Algorithm, ZHONGGUO YUWEN (中国語文), May, pp. 216-220, July, pp. 279-285 (1981).
10) 天野真家，平川秀樹：英日機械翻訳用パーサについて，情報処理学会研究会報告，NL 32-1 (1982).
11) 田中穂積，辻井潤一，横山晶一，安川秀樹，鈴木克志，井佐原均，村田賢一，モニカ・ストラウス：より自然な翻訳へのアプローチ，自然言語処理技術シンポジウム，pp. 115-121 (1984. 11).
12) 佐藤理史，長尾 真：実例に基づいた翻訳，情報処理学会研究会報告，NL 70-9, No. 6 (1989).
13) 任 福継，宮永喜一，栃内香次：コード方式日中機械翻訳の実験システム JCMTS の概要，情報処理学会研究会報告，NL 72-7, No. 40 (1989).
14) 望月八十吉：中国語と日本語，光生館 (1974).
15) 揚 頤明，堂下修司：中国語の意味文法の構成とその処理系の作成，情報処理学会論文誌，Vol. 27, No. 2, pp. 155-164 (1986).
16) 任 福継，宮永喜一，栃内香次：日中機械翻訳システムにおける慣用表現のエンコードおよびデコード，信学技報，NLC89-30, A189-56, pp. 39-46 (1989).
17) 任 福継，宮永喜一，栃内香次：日中常用文型機械翻訳システム，信学論 (D-Ⅱ)，Vol. J 74-D-Ⅱ, No. 8, pp. 1060-1069 (1991).
18) 任 福継，范 莉馨，宮永喜一，栃内香次：JCMTS における慣用表現の解析手法，信学技報，NLC90-1, pp. 9-16 (1990).
19) 石 明徳：科技日語慣用句型，上海科学技術出版社 (1980).
20) 李 士俊，雷 跻九：日語常用詞彙及慣用型，未来出版社 (1985).
21) 首藤公昭，吉村賢治，武内美津乃，津田健蔵：日本語の慣用的表現について，情報処理学会研究会報告，NL-66-1, No. 38 (1988).
22) 田中穂積，井佐原均，安川秀樹：融合方式による機械翻訳システムの実験，情報処理学会研究会報告，NL 34-2, pp. 7-12 (1982).
23) 熊野 明，天野真家：英日機械翻訳システムの訳文生成について，情報処理学会研究会報告，NL 40-6, No. 40, pp. 1-6 (1983).
24) 奥 雅博：日本語慣用表現の分析と日英翻訳への適用，情報処理学会研究会報告，NL-62-2, No. 53 (1987).
25) 首藤公昭：文を構成する基本単位はなにか，数理科学，特集／計算言語，No. 309, pp. 29-34 (1989).
26) 周 敏西：省略及有関的几個問題，日語学習与研究，pp. 21-25 (1989. 4).
27) 荻野孝野：日本語の意味分類体系，計量国語学，Vol. 16, No. 3, pp. 95-112 (1987).
28) 任 福継，宮永喜一，栃内香次：格関係と意味属性による補助語の推定手法，北海道支部連大，No. 285, pp. 333-334 (1990. 10).
29) IPAL，計算機用日本語基本動詞辞書，情報処理振興事業協会技術センタ (1987).

付録 1 基本文解析用コード

No.	コード記号	助詞条件	付加条件	コード意味
1	SUB 1	は	略(下同)	主題 1
2	SUB 2	も		主題 2
3	AGT 1	は，が		動作者
4	AGT 2	は，が		動作物
5	OBJ 1	を		対象 1
6	OBJ 2	と，に		対象 2
7	ATT	の		限定
8	ADV	—		状況
9	TIM 1	に		時間 1
10	TIM 2	で		時間 2
11	CTN	と		内容規定

No.	コード記号	助詞条件	付加条件	コード意味
12	ASS	—		補助
13	FTM	から		時間始点
14	TTM	まで		時間終点
15	DUR	で		時間範囲
16	SPA 1	に		場所1
17	SPA 2	で		場所2
18	FSP	から		空間起点
19	TSP	まで		空間終点
20	SPT	を		経過場所
21	MAT 1	で		材料，部品1
22	MAT 2	から		材料，部品2
23	TOO	で		道具
24	MET	で		手段，方式
25	FRE	—		頻度
26	COM 1	より		比較1
27	COM 2	と		比較2
28	ACO 1	とともに		随伴1
29	ACO 2	に伴って		随伴2
30	CAU 1	ので		原因1
31	CAU 2	から		原因2
32	PUR	ため		目的
33	CON 1	ば		条件1
34	CON 2	で		条件2
35	DIR	へ		方向
36	PAR 1	と		受け手
37	PAR 2	に		与えて
38	RES 1	—		結果1
39	PRE	—		行為
40	QUA	—		数量
41	DEG	—		程度
42	INC	として		付帯関係
43	DES	のように		形容関係
44	RES	れる		尊敬
45	PLE	ください		丁寧
46	INV	しょう		勧誘
47	PAS	れる		受動
48	EMP	せる		使役
49	DEN	ない		否定
50	AGU	よう		推量
51	VOL	れる		自発
52	POS	れる		可能
53	EXI	てある		存在
54	HOP	たい		願望
55	ADV	ている		進行
56	QUE	か		疑問
57	ORD	せよ		命令
58	TRY	てみる		試行
59	HEA	そうだ		伝聞
60	HYP	もし…ば		仮定

注：助詞条件については代表的なものだけを示す．

付録 2　基本文のコード表現（一部）

文型名	例文（日本語の訳文）	コード表現		説明
		正コード	副コード	
自動詞文	天晴（空が晴れる）	SUB：天 PRE：晴	な し	日本語文のコード，語順と一致
他動詞文	薬水出現副作用（水薬が副作用を生ずる）	SUB：薬水 PRE：出現 OBJ：副作用	な し	語順は日本語文と異なる
兼語文（使役式文）	医生譲病人用中薬（医師が患者に漢方薬を飲ませる）	SUB：薬水 PAR：病人 PRE：用 OBJ：中薬	PC1：使役（譲） 類似語：使，叫，請	副コードPC1で使役を表示する
二客語文（双目的文）	張老師給小王鋼筆（張先生は王さんに万年筆を与える）	SUB：張老師 PAR：小王 PRE：給 OBJ：鋼筆	な し	兼語文のコードと一致するが，副コードPC1がない
受動式文	張老師被学生邀請（張先生が学生に招待される）	SUB：張老師 PAR：学生 PRE：邀請	PC2：受動（被） 類似語：受，挨，遭	副コードの有無で意味が完全に異なる

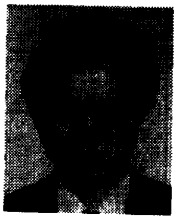
付録 3　実験に用いた常用文型の一覧表

No.	常用文型	多／一	多義性
1	あえて…ない	多	○
2	あながち…ない	多	○
3	あまり…ない	多	○
4	あやうく…ところだった	多	×
5	いがいには…ない	多	○
6	いかなる…ても	多	×
7	いかにも…らしい	多	○
8	いくら…ても	多	○
9	いささか…ない	多	×
10	いっこう…ない	多	×
11	かつて…ない	多	×
12	かならずしも…ない	多	○
13	けっして…ない	多	×
14	さえ…ば	多	×
15	さほど…ない	多	×
16	さらに…ない	多	×
17	しか…ない	多	×
18	すこしも…ない	多	×
19	すんでのことに…ところだった	多	○
20	ぜんぜん…ない	多	×
21	それほど…ない	多	×
22	たいして…ない	多	×
23	たえて…ない	多	○
24	たとえ…ても	多	×

No.	常用文型	多/一	多義性
25	ちっとも…ない	多	○
26	ちょうど…のようだ	多	○
27	ということは…ということである	多	×
28	どうか…てください	多	×
29	どうしても…ない	多	×
30	とうてい…ない	多	○
31	とても…ない	多	○
32	どんなに…ても	多	×
33	ないかぎり…ない	多	×
34	なかなか…ない	多	○
35	ひとり…だけでなく…である	多	×
36	まるで…ようだ	多	○
37	いうまでもない	一	○
38	うちに	一	○
39	かもしれない	一	×
40	からといって	一	○
41	からには	一	×
42	かわりに	一	○
43	…たことがある	一	×
44	ことができる	一	×
45	しかたがない	一	○
46	ずにはいられない	一	○
47	だけでなく	一	×
48	つもりだ	一	○
49	てはいけない	一	○
50	というよりむしろ	一	×
51	というわけではない	一	○
52	ないとはかぎらない	一	○
53	なければならない	一	×
54	におうじて	一	○
55	にかんして	一	×
56	にさいして	一	×
57	にしたがって	一	○
58	にすぎない	一	×
59	にちがいない	一	×
60	にはあたらない	一	×
61	にもかかわらず	一	○
62	にもとづいて	一	○
63	ばかりでなく	一	×

注：「一」は一語性常用文型を表す．
「多」は多語性常用文型を表す．
「○」は多義性のあることを意味する．
「×」は多義性のないことを意味する．

(平成3年1月11日受付)
(平成3年6月13日採録)



任　福継（正会員）

1982年中国北京郵電学院電信工程部計算機と通信専攻卒業．1985年同大学院計算機応用専攻修士課程修了．1986年中国科学院博士課程入学，1987年中退来日．1991年北海道大学工学研究科電子工学専攻博士課程修了．工学博士．計算機科学，自然言語処理，特に機械翻訳の研究に従事．

范　莉馨（正会員）

1984年中国北京郵電学院無線通信専攻卒業．同年，中国郵電部郵電科学研究院に勤務．助理工程師．1991年北海道大学大学院電子工学専攻修士課程修了．現在，同大学院博士後期課程在学中．自然言語処理，機械翻訳に関する研究に従事．電子情報通信学会会員．

宮永　喜一（正会員）

1956年生．1981年北海道大学工学部電子工学専攻修士修了．工学博士．現在，北海道大学工学部電子助教授．並列計算機システム，ディジタル信号処理等の研究に従事．電子情報通信学会，日本音響学会，IEEE 各会員．

栃内　香次（正会員）

昭和14年生．昭和37年北海道大学工学部電気工学科卒業．昭和39年同大学院工学研究科修士課程修了．現在同工学部電子工学科教授．工学博士．自然言語処理，音声情報処理および信号処理プロセッサなどの研究に従事．電子情報通信学会，日本音響学会各会員．