自然言語処理 72－7
（1989. 5. 19）

コード方式日中機械翻訳の実験システムJCMTの概要

任福継　宮永喜一　栃内香次

北海道大学工学部

筆者らはコード方式日中機械翻訳実験システムを開発している．コード方式は，従来の翻訳方式における種々の問題点の検討に基づいて提案した新しい機械翻訳方式である．コード方式では，原言語文の解析結果を，コードという，文の意味的な基本の単位となる要素と，その要素のもつ意味の組の集合で表示するものである．したがって，この方式は個別言語への依存性が少なく，多言語間の機械翻訳にも適していると考えられる．

本稿では，この方式による日中機械翻訳実験システムの概要を述べ，翻訳実験の結果を示す．

THE OUTLINE OF JAPANESE-CHINESE MACHINE TRANSLATION SYSTEM BASED ON CODE METHOD

Fuji Ren　Yoshikazu Miyanaga　Koji Tochinai

Faculty of Engineering, Hokkaido University

Kita 13, Nishi 8, Kita-ku, Sapporo 060, Japan

As a new system, a Japanese-Chinese machine translation system (JCMT) is constructed by use of a code method. In the method, intermediate results which are translated from original language become a set of codes. The code is a basic unit including the meaning of a sentence. The unit does not depend on a special language. Thus the method is convenient for multi-lingual machine translation.

This paper shows the outline of JCMT and some results of translation.

1．まえがき

機械翻訳の手法には，よく知られた二つのアプローチ，すなわち変換方式と中間言語方式（ピボット方式）がある．変換方式の利点は，言語情報の中でも比較的明確に記述しやすい構文情報を基に異なる言語を対応付け，翻訳できることである．すなわち異なる言語を対応付ける部分を少量の言語情報で形式的に行い，その他の部分はそれぞれの言語の特徴を活用する．変換方式の欠点として，変換規則がうまくコンパクトに記述できないとか，個別的になるという問題点がある．また，言語表現の一部のみを用いて対応付けするため，曖昧さが多く出てくるという問題点がある．一方，中間言語方式は，原言語文の解析結果を英語・日本語・中国語といった個別言語に依存しない中間言語で記述する機械翻訳方式である．その利点として，多言語間の翻訳を簡単に実現できることと，原言語文の解析結果が原言語とは無関係に表現されるので，原言語文の構造に依存しない，自然な翻訳文を生成できることなどがあげられる．この方式の欠点として，中間言語をどのように設計するかの指針が明確でない，すなわち，真に言語に依存しない中間言語を設定できないこと．また，原言語から中間言語への翻訳と，中間言語から目的言語への翻訳の二段の翻訳処理が必要になるという問題がある[1]．

本論文では，コード方式の機械翻訳を提案する．コード方式において，原言語を変換して得られる中間結果は文の意味的な基本の単位となるコードの集合である[2]．すなわち，まず，翻訳しようとする原言語文を，コード元素に分解してから文の意味的な基本の単位となるコードを生成する．この過程をエンコードという．そして，得られたコードの集合と，語順など目的言語の文法構造に従って，目的言語のコード列を作る．この過程をデコードという．最後に，このコード列から目的言語を生成する[3]．

本論文では，コード方式に基づく日中機械翻訳実験システムの概要を述べる．２章では中国語の特徴を述べ，３章ではコード方式の原理を説明し，４章では日中機械翻訳のアルゴリズムを述べ，５章では実験システムを紹介し，翻訳結果の例を示す．

2．中国語の特徴

中国語には，日本語のような「テ、ニ、オ、ハ」（格助詞）や，ヨーロッパの言語のような語尾変化がほとんどない[4,5,6]．以下，中国語の特徴の2，3について述べる．

2.1 語順

中国語には日本語の格助詞に相当するものがなく，語尾変化もないので，単語間の文法的関係は，主として，単語の並べかた（語順）によって表されている．したがって，中国語では，語順がきわめて重要である．例えば，’我教他’（私は彼を教える）の語順を逆にすると，’他教我’となって，意味がまったくちがってくる．

2.2 動詞

中国語の動詞の特徴としては，つぎの点があげられる．

（1）肯定・否定の形式で，疑問を表わすことができる．

例：山田看不看小説　　（山田さんは小説を読むか）

ここで，’不看’は’看’の否定形式である．

（2）動詞を重ねることによって，つぎのような意味を付加することができる．

・動作の時間の短いこと　［例a］

・動作が気の向くままに，何度かくり返されること　［例b］

・試みにやってみること　［例c］

重ねる方式はＡＡ式とＡＢＡＢ式がある．

看看	想想	試試	（ＡＡ式）
研究研究	休息休息	復習復習	（ＡＢＡＢ）

［例a］我想看看小説　　（私はちょっと小説を読みたい）

［例b］最近，復習復習生詞　　（近ごろ，新しい単語を復習している）

［例c］這本書不難，読読看　　（この本はあまり難しくないので，読んでみましょう）

（3）中国語の動詞は，英語・ドイツ語などとちがって，時制・人称・数によって，語尾が変化しない．その代り，時態助詞や補語を使って，動詞の変化・活用形をつくり，ある意味を付加することができる．

英語では，時制の変化は動詞の語形の変化や助動詞によって表わされる．また，日本語では，時制の変化は動詞の活用形で表わされる．しかし，中国語では，時制の変化は，通常，時間詞（時間を表示する語・句－－状況語）で表わされる．したがって，過去時制は過去の時間を表示する語・句で表わされ，未来時制は未来時制の時間を表示する語・句で表わされる．この場合に，動詞が変化しないのが中国語の特徴である．

また，動作の態を表わす動態には，将動態・開始態・進行態・持続態・完了態・経験態などがある．そして，動態は動詞の前（または後）に付く相応の単語で表示する．

例えば，我去年読過了這本書　（私は去年にこの本を読んでしまった）

（4）いくつかの動詞は連続に出現できる．また，日本語における動詞の可能形態、受動形態、使役形態などの文法現象は中国語では相応の動詞を用いて表現する．

例：食べる→吃　　食べさせる→譲吃

ここで，’させる’という形態は’譲’という動詞で表現する．

（5）動詞兼名詞

ある単語が動詞であると同時に名詞であるという文法現象は日本語、英語の中にも存在しているが，中国語においては，格助詞がなく，動詞の語尾変化もないので，この現象は，中国語文の機械解析、機械生成において大きな問題点となる．この問題の検討は別の機会に譲る．

2.3 複雑な補語

中国語文法においては，文の要素として，主語・述語（

謂語）・目的語（賓語）・限定語・状況語・補語がある．ここで，最も複雑な補語について説明する．

動詞・形容詞（日本語の形容詞・形容動詞に相当する）のあとについて，動詞・形容詞を修飾することばを，中国語文法では補語といっている．ふつう，補語には，動詞・形容詞・副詞・数量詞などが使われる．補語には，次のような種類がある．

・程度補語　・結果補語　・方向補語　・可能補語　・数量補語　・時間補語

3．コード方式機械翻訳の概要

定義1　コード元素　言語を構成する基本的な単位をコード元素という．

本論文で述べる日中機械翻訳実験システムでは，コード元素は文節である．

定義2　コード　コード元素またはコード元素の集合であり，それに文法あるいは意味的条件を表すコード記号が賦与されたものをコードという．

処理時間の減少と手順の簡単化のため，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はないが，ふつう，時制、動態、語感などを表すものを副コードとしている．

定義3　コード列　一組のコードで，コードの順序が確定されたものを，コード列という．

例：わたしたちはコード方式による日中機械翻訳の実験システムを開発している．

この文のコード元素を図1 に示す．

元素1 ： わたしたち
元素2 ： コード方式
元素3 ： 日中機械翻訳
元素4 ： 実験システム
元素5 ： 開発する

図1 コード元素

図2 にこれを解析して得られた正コードおよび副コードを示す．具体的な解析方法は次章で述べる．

多くの機械翻訳システムでは，入力文の分析結果から目的言語の表現を得るのに，入力文の分析結果の図式を目的言語での対応する図式に変換し（構造変換），そこから表層の構文を導き（構文合成）、形態素合成をして出力文を得ている．コード方式では目的言語での対応する図式を経ず，入力文の分析結果から，入力文の意味表現のコードだけを得る．この過程を原言語のエンコードという．

目的言語の側では，原言語文のコードに対し，目的言語の文として正当な順序を推定する．これは目的言語の構文規則および表現方式にしたがって，行なわれる．例えば，図2 の日本語文のコードに対して，目的言語が中国語であれば，図3 のコード列を得る．この過程をデコードという．

正コード	
コード記号	対応する単語
SUB	わたしたち
TOO	コード方式
OBJ	日中機械翻訳の実験システム
PRE	開発

副コード	
コード記号	関係する単語
0010	開発する

図2 正コードと副コード

コード列　（中国語）	
コード記号	対応する中国語
SUB	我們
TOO	編碼方式
PRE	開発
<0010>	<正在進行中>
OBJ	日中機器翻訳的実験系統

図3　中国語のコード列

さらに，得られたコード列に対して，目的言語の規則に従って訳文を生成する．このとき，コードの性質にしたがって単語の付加、訳文の修飾および時態、動態の付加などの操作を行う．図3 の中国語のコード列によって，図4 のような中国語文が得られる．

我們<正在>｛用｝編碼方式開発日中機器翻訳的実験系統

注：<>の部分は副ｺｰﾄﾞによって生成したものである
　｛｝の部分はｺｰﾄﾞTOO の条件から付加された単語である
　図4 生成した中国語訳文

なお，図2 の日本語文のコードに対して，目的言語が英語であれば，同じ手順で，図5 に示す英語コード列と図6 に示す英語訳文が得られる．

図7 に本機械翻訳システムの概念構造図を示す．

図7 において，A部とB部は原言語だけに関係し，C部とD部は目的言語だけと関係する

英語のコード	
コード記号	対応の英語単語
SUB	WE
PRE	DEVELOPE
<0010>	<BE-ING>
OBJ	EXPERIMENTAL SYSTEM OF JAPANESE -CHINESE MACHINE TRANSLATION
TOO	CODE METHOD

図5 英語のコード列

We <are> develop<ing> on experimental system of Japanese-Chinese machine translation {by the use of} code method

図6 生成した英語

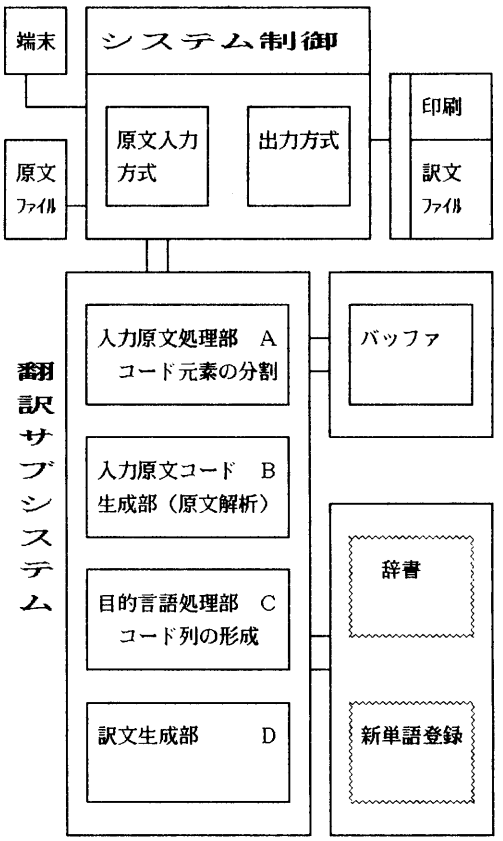


図7 翻訳システムの概念構造図

4．日中機械翻訳のアルゴリズム

本章ではコード方式日中機械翻訳システムの具体的な翻訳アルゴリズムを述べる．

4.1 日本語解析とコード生成

日本語のコード生成方法の要点は以下の二つである．

a．助詞を中心として，日本語文を解析する．

b．動詞の活用形から原形を推定し，活用形を区別する副コードを生成する．

本システムにおいてコード元素は文節に対応している．コード元素の分割は，特別な処理を行なわなず，日本語文の入力時，文節分かち書き入力を行ってスペースで区別する．

コード生成の際，まず，必要条件を判断する．必要条件を満足すれば付加条件を確認する．

表1にコードおよび対応する条件を示す．表1では，付加条件を省略したが，つぎに，OBJ というコードを例として付加条件を説明する．OBJ コードの意味は動作の対象と目的を表示するものである．この性質のコードは，日本語文の中に必必要の条件は格助詞'を'（'が'）がある，その上，ある付加条件を満足しなければならない．すなわち，格助詞'を'の場合，行為の動詞は他動詞でなければならない．以下の例で説明しよう．

表1　　コード表　（一部）

番号	記号	必要条件（助詞）	付加条件	コードの意味
1	SUB	が、は、も	略、下同	行為の主体
2	PRE	用言*,句尾（だ等）		行為（事）
3	OBJ	を（が）		行為の対象（目的）
4	ATT	の、動詞の原形		限定（修飾）
5	TIM	時間名詞＋に、で		時間
6	TFR	時間名詞＋から		時間の始点
7	TTO	時間名詞＋まで		時間の終点
8	DUR	時間、数詞＋で		限定の範囲
9	SPA	場所名詞＋に、で		場所
10	SFR	場所名詞＋から		場所の始点
11	STO	場所名詞＋まで		場所の終点
12	STH	を		動作進行経過の場所
13	MAT	材料名詞＋で、から		材料
14	TOO	道具名詞＋で		手段、道具
15	DIR	へ		目標

注：表中，必要条件の記述は主要な部分のみに限

*：用言→動詞、形容詞、形容動詞

例1:本を読む　　(讀書)

例2:勝利をかちとる　　(争取勝利)

例3:川が町の中を流れる　　(河流从街中間穿過)

例4:高い秋の空を鳥が飛んでいる　　(小鳥在秋天的高空中飛翔)

例1 、例2 の場合には，' 本' と' 勝利' はOBJ コードの必要条件と付加条件とも満足するので，' 本' と' 勝利' はOBJ コードである．しかし，例3 、例4 の中に，その行為動詞' 流れる' と' 飛ぶ' が自動詞であり，付加条件を満足しないので，' 町の中' と' 高い秋の空' はOBJ コードではなく，NO12のSTH コードである．

格助詞' が' の場合には，付加条件はやや複雑である．

つぎの条件中のいずれも満足すれば，OBJ コードになる．

#1.用言（動作）は' 希望' の意味を表示すること　　(例1)

#2.用言は' 好き' 、' 嫌い' の意味を表示すること　　(例2)

#3.用言は' 能力' の意味を表示すること　　(例3)

#4.用言は' 可能' 、' 不可能' の意味を表示すること　　(例4)

例1.わたしはあの本が読みたい　　(あの本)

例2.兄はスキーが嫌いだ　　(スキー)

例3.秋山は絵がうまい　　(絵)

例4.彼女は中国語が読める　　(中国語)

表２は用言に関係するコード説明表であり，表３は体言に関係するコード説明表である．次の章の説明の便利のため，表中に中国語の意味すなわち中国語生成時に付けなければならない単語を一緒に書いてある．

つぎに，活用形を区別する副コードを生成するため，動詞の活用形から原形を推定する方法を述べる．

実験システム中に，副コードの生成と原形の推定を行なうため，９つのサブルーチンを用意してある．ここでは，その概要を述べる．

☆　還元の詞類：動詞、助動詞、形容詞、形容動詞

☆　還元の方法：

1.変化した単語の尾部に助動詞を見つける

2.本助動詞の接続方法を調べる

3.上の接続方法によって原形を還元する

以下の点を注意しなければならない．

1.される及びさせる

' される' と' させる' の直前が２個の漢字単語なら，' する' に還元する．そうでなければ，サ変と五段活用動詞一形（未然形）との二種類変化ルール（図8 に示す）で還元し，辞書を引いて推定する．

例えば，'教育される'に対し，'教育する'に還元する．そして，'促される'に対し，'促す'に還元する．

2.音便変化

日本語において，助動詞の接続中での始めの仮名が'た'、'て'である時，'す'と'する'語尾の五段活用動詞以外に，五段活用動詞に接続すると，音便変化がある．そして，'た'、'て'の濁音化もある．表４は音便の還元ルールである．

表２　　用言関係のコード　（一部）

記号	必要条件	付加条件	意味*
1001	ようになっている	略	成了
1002	ようにしている		成為
1003	から、ので		因為
1004	ている		ー
1005	ために		為了
1006	ため		為
1007	ことができる		能（）
1008	ことにより		根據
1009	ことになる		成
1010	べきである		応該
1011	という		称着…的
1012	ければならない		必須
1013	につれて		随着
1014	ように		那様
1015	ような		那様的

＊：コードの意味の説明は簡略化してある，示されている単語は中国語文を生成する時に本コードに対応して付けなければならない単語である．

表３　　体言関係のコード　（一部）

記号	必要条件	付加条件	意味*
0001	において	略	在…中
0002	について		関於
0003	についても		関於…也
0004	については		関於
0005	によって		根據
0006	により		由於
0007	として		作為
0008	に対して		対於
0009	にわたり		渉及
0010	としている		作為…的
0011	のようになる		象…那様
0012	とともに		和…同時
0013	のために		為了
0014	である		…
0015	だけ		只

＊：表２と同じ

表4　音便変化の還元ルール

い→	く	っ→	う	ん→	ぬ
	っ		つ		ぶ
	ぐ(だ,で)		る		む

例えば，行なった

↓

行なっ ……た→過去助動詞

↓

行な［う／つ／る］（表4）

↓ ……辞書を引く、複数単語がある処

行なう ……最終の推定

動詞の活用形から原形を推定すると同時に，副コードを生成する．上例の場合，過去の意味を表示する副コードが生成されている．

形容詞、形容動詞の還元も同様に処理する．図8に用言の還元法をまとめて示す．

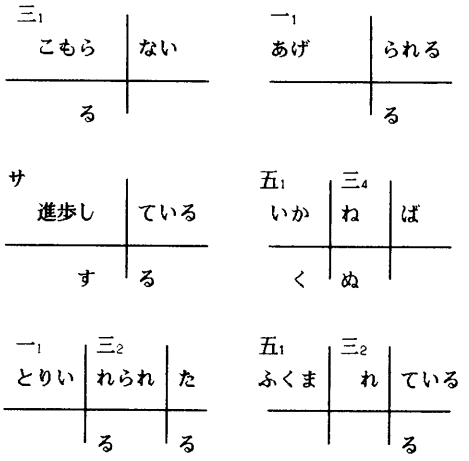


図8 用言の還元

図8 において，’五’：五段活用動詞　’一’：上，下一段活用動詞　’サ’：サ変動詞　’三’：上の三つの種類（どちらも可）　’1,2,3,4,5’：動詞の五つの基本活用形

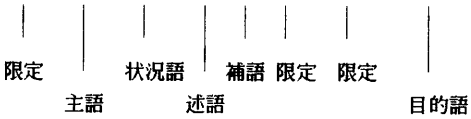
4.2 中国語文の生成

4.2.1 文の要素の構造（語順）

第2章で述べたように，中国語においては，文の要素間の直接組合せ関係は語順によって表示される．中国語の動詞述語文を例として，文の要素の基本語順を図9 に示す．図10は中国語文の一例である．

図9 中国語文の基本語順

（荒木的）朋友＜昨天＞買到了（一本）（新出版的）雑誌



昨日荒木の友人は一冊の新出版の雑誌を買ってしまった．

図10　中国語文の構造の一例

コード列を生成するために，中国語の基本文型のコード順序（語順を代表するもの）を登録した表を用意してある．

4.2.2 中国語常用文型のコード表現

表5に中国語で常用する文型のコード表現及び日本語との対応関係を示す．

5・実験システム

本日中機械翻訳実験システムＪＣＭＴはPC9801上に開発しているが，北海道大学大型計算機センターのHITAC682Hへも移植されている．

5.1 辞書の作成

現在，我々の手許には，中国語漢字用のシステムがないので，日本語の漢字を用いて，中国語の単語をつくっている．もし，ある中国語単語の漢字がなければ中国語の同音字で代用し，前後に括弧を付けて区別する．現在，辞書の容量は2300語で，そのうち，23個の漢字だけが同音字で代用されている．

表6，7に日中辞書の一部を示す．なお，表7は名詞，動詞，形容詞等の細かい意味属性を示す表で，表6のM部とリンクしている．

5.2 プログラム言語

実験システムでは，pro FORTRAN-77で作成されている．

5.3 システムの構造

図11にシステムの構造図を示す．

★ 原言語部

原言語は日本語である．この部分では，入力の日本語文を解析して，コードを形成する．活用形処理部には次の6個の処理サブルーチンがある．

☆テンス処理
☆受動態処理
☆使役態処理
☆否定態処理
☆希望態処理
☆可能態処理

★ 辞書管理部

この部分で，変化した活用形から原形を探し，原形の訳

語を見つける．そして，単語品詞の性質による再確認という過程を含む．新出単語をNEWWODというファイルに登録する．

★ 入出力管理部

入出力方式は以下の二種類がある．

(1) キーボードで入力，画面とプリンターで出力

(2) ファイルで入出力

★ 目的言語部

まず，コード列を形成する．そして，副コードによって訳文を生成する．

5.4 翻訳実験

以下に示す文献を用いて翻訳実験を行った．

a.Japanese in Thirty Hours, Eiichi Kiyooka, 中訳本，湖南科学技術出版社，1980.

b.c.d.情報処理学会論文誌，第24巻第2号,第26巻第4号,第27巻第3号，栃内，他．

e.f.g.h.i.本論文末尾の参考文献1,2,3.5,7.

表5 中国語常用文型のコード （一部）

文型名	例の文 （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	無し	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	無し	語順は日本語文と異なる
兼語文 （使役式文）	医生讓病人用中薬 （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 （讓） 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	無し	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

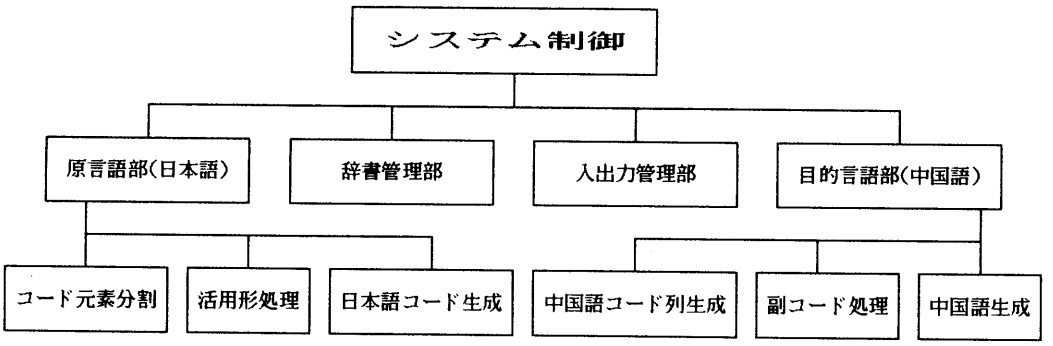


図11 日中機械翻訳システム構造図

表6　日中辞書

JWORD	JTY	M	CWORD	CTY	未用
サイクル	n	6	循環	n	
割く	v5	2	割開	v	
さようなら	tt		再見	tt	
直に	ad		馬上	ad	

注：JWORD:日本語単語
JTY:　日本語単語の品詞の種類
M:　単語の属性（表7）
CWORD:中国語単語
CTY:　中国語単語の品詞の種類

表7　Mの説明

JTY	Mの記号	意味
n（名詞）	1	時間性質を表わす名詞
	2	場所性質を表わす名詞
	3	材料性質を表わす名詞
	4	道具性質を表わす名詞
	5	生物性質を表わす名詞
	6	ー
	7	身体部位の名詞
	8	動物の名詞
	9	名詞またサ変動詞
	10	物理特性の名詞
	11	人名
	：	：
v（動詞）	1	自動詞
	2	他動詞
	3	自、他動詞
a（形容詞）	1	形容詞
	2	形容動詞
：		

5.5 実験例・結果

以下に，(a) 入力の日本語文と(b) 出力の中国文の例を示す．

(1) (a) 荒木さんは　優秀な　成績を　収めるので，先生に　褒められる．

(b) 荒木因為取得優秀的成績被老師表揚

注：本例は受動文型の翻訳例である．

(2) (a) わたしは　妹に　電灯を　つけさせる．

(b) 我讓妹妹開電灯

注：本例は使役文型の翻訳例である．

(3) (a) 鈴木さんは　中国の　歴史を　研究したい．

(b) 鈴木想研究中国的歴史

注：本例は希望を表現する翻訳例である．

(4) (a) 電子機器講座の　先生と　学生は　十二月十六日に　忘年会を　行なったか．

(b) 電子機器教研室的老師和学生在十二月十六日進行了忘年会否

注：本例は疑問を表現する翻訳例であるが，中国語標準語の疑問詞は本計算機で不存在なので，'否'で代わる．'否'は中国語のある方言の疑問詞である．

(5) (a) 専門分野で　よく　使われる　漢字語は　限られるという　我々の　予想を　確認するために，現実の文献について　調査を　行なった．

(b) 為了確認被専業領域経常的使用的漢字語被限定這様的我們的予想関於現実的文献進行了調査

注：本例は過去を表現する翻訳例である．

本システムの翻訳性能を定量的に評価するために，文献a)から640 文を用いて翻訳実験を行なった．その結果は

意味に正しい訳文数:506

うち，完全に正しい訳文数:367

誤訳文数:134

誤訳文の中，半分は単語が辞書に存在しないために発生したものである．

他の文献を用いる実験は進行中である．

6．おわりに

以上，コード方式の機械翻訳手法を提案し，この方式を用いて日中機械翻訳実験システムを構築した．さらに，実験結果を報告した．今後，コード種類の拡張，複文のコード生成，対話学習機能の検討及び性能評価などを予定している．

参考文献

［1］野村浩郷，田中穂積，機械翻訳，bit 別冊，1988，共立出版．

［2］任福継，宮永喜一，栃内香次，コード方式日中機械翻訳システムの構想，昭和63年電気関係学会北海道支部連合大会論文集，268,pp.322-323．

［3］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について，1989年電子情報通信学会春季全国大会論文集，D310．

［4］陳国梁，現代漢語語法教程，西安交通大学出版社，1986．

［5］三野昭一，中国語文法の基礎，三修社，1987．

［6］劉月華，他，実用現代漢語語法，外語教育与研究出版社，1986．

［7］長尾真，他，科学技術庁機械翻訳プロジェクトの概要，情報処理，OCT.1985. Vol.26. no.10.