

## 日中機械翻訳における係り受け構造の可保留曖昧関係について

任 福 繼<sup>†</sup> 范 莉 馨<sup>††</sup>  
 宮 永 喜 一<sup>††</sup> 栄 内 香 次<sup>††</sup>

非同族言語間の機械翻訳では、原言語の係り受け構造の曖昧性をあらかじめ解消しなければならないとされている。日本語と中国語は非同族の言語であり、文全体の構造は異なるが、文のある部分ではその語順が同じである。この部分に対して、日本語の係り受け構造の曖昧性を残したまま翻訳しても中国語においても日本語の意味を復元されると考えられる。そして、あらかじめ与える訳文関数を用いてこの部分の翻訳を行うことができる。本論文では、上記のことについて日中機械翻訳における可保留曖昧関係 (RSA) を提案する。可保留曖昧関係は原言語の係り受け構造の曖昧性を解消しなくともその訳文を生成できる曖昧関係である。具体例として、並列助詞「と」と連体助詞「の」と名詞からなる名詞句および用言連体形からなる文の係り受け構造の可保留曖昧性について検討し、翻訳手法を提案する。技術論文 2919 文から上記可保留曖昧関係パターンをもつ 806 文を抽出し、翻訳実験を行った結果、正解率は 97.7% であった。これにより、本論文で提案した手法の有効性を確認することができた。

### On Reservable Structural Ambiguities in Japanese-Chinese Machine Translation

FUJI REN,<sup>†</sup> LIXIN FAN,<sup>††</sup> YOSHIKAZU MIYANAGA<sup>††</sup> and KOJI TOCHINAI<sup>††</sup>

In this paper a new concept, Reservable Structural Ambiguities (RSA), in Japanese-Chinese machine translation, is presented. The RSA are structural ambiguities of Japanese which can be translated into Chinese without being resolved. Usually, when translating from one language to another which doesn't belong to the same "language family", structural ambiguities must be resolved. However, Japanese and Chinese don't belong to the same family, their difference being mostly on sentence structure. However, some parts of their sentences are similar, making RSA possible, and this paper will focus on them. Some RSA patterns will be discussed and a method for generation of Chinese from Japanese sentences with RSA will be given. An experimental system based on this method was constructed and an experiment on 806 sentences which had RSA, taken from 2919 sentences from technical articles, was carried out. The result was a correct translation rate of about 97.7%, showing that the proposed method is quite effective.

#### 1. はじめに

一般に、非同族言語間の機械翻訳では、正しい訳文を得るために原言語の係り受け構造の曖昧性を解消しなければならない。その手法として、現在、実用化されている機械翻訳システムでは、構文解析と同時に動詞の格フレームと名詞の意味素性およびヒューリスティックなルールを用いた意味解析を行い、単語の意味や係り受けの曖昧性を解消している<sup>1)~3)</sup>。例えば、田村ら<sup>4)</sup>は意味解析を用いた並列名詞句の構造解析手

法を、稻垣ら<sup>12)</sup>は意味連結パターンを用いた係り受け解析手法を、また、向仲<sup>5)</sup>は技術文書の機械翻訳において常識と文脈情報をを利用して構文解析の曖昧性を解消する手法を、それぞれ提案している<sup>6)~8)</sup>。さらに、平井らは日本語係り受け構造の曖昧性検出において、検出の漏れを抑えながら、不要な曖昧性検出を少なくする方式を提案している<sup>5)</sup>。図 1 に、現在一般的な機械翻訳システムにおける係り受け構造の曖昧性処理の概念を示す。

20~40 語からなる普通の文では多品詞と係り受けの曖昧性を総合すると、統語解析の候補が億に上ることもありうることが天野により指摘されている<sup>10)</sup>。それゆえ、係り受け構造の曖昧性を解消するためには大量の処理時間を必要とする。さらに、「山田の家と寺が火事で焼けた」、「美しい谷間の百合」のような文に

† CSK 技術開発本部 MT 開発部

Machine Translation Development Dept., CSK

†† 北海道大学工学部電子工学科

Department of Electronics, Faculty of Engineering, Hokkaido University

おける係り受け構造の曖昧性を解消することは極めて困難である。特に、後の例文では、話者（書き手）自身、いかなる係り受け関係を表したいかを厳密には考えていない可能性がある。

本論文では、日中機械翻訳における可保留曖昧関係 (RSA) について考察する。可保留曖昧関係とは、係り受け構造の曖昧性を解消しなくても訳文を生成できる曖昧関係である。

例えば、“山田の家と寺が火事で焼けた”という文では、次の二つの構文構造 A, B がいずれも成立する。

構造 A : (山田の家) と寺が火事で焼けた。

構造 B : 山田の (家と寺) が火事で焼けた。

しかしながら、中国語訳文は構造 A でも構造 B でも同一で、以下のようにになる。

中国語：由於火災山田的家和寺廟被燒了。

このような性質を持つ曖昧性を本論文では可保留曖昧関係という。すなわち、日本語と中国語は異なる言語族に属し、文全体の構文および表現方式が異なるが、部分的には構文および表現が類似している場合があり、このような部分に対しては、日本語の係り受けの曖昧性を残したまま中国語に翻訳することができる。しかし、当然であるが、すべての係り受け構造の曖昧性が可保留なのではなく、そのうち一部だけが可

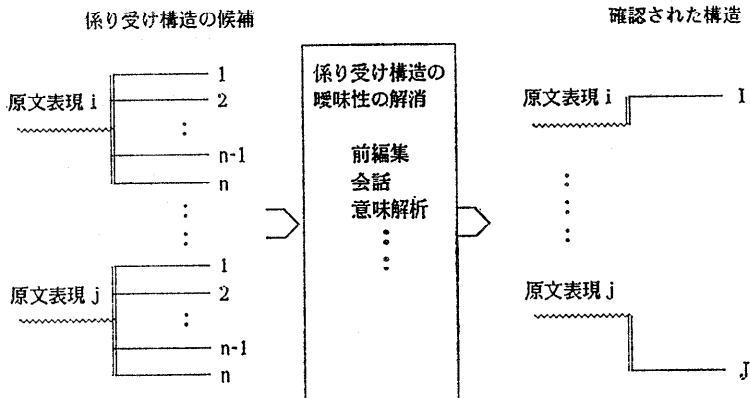
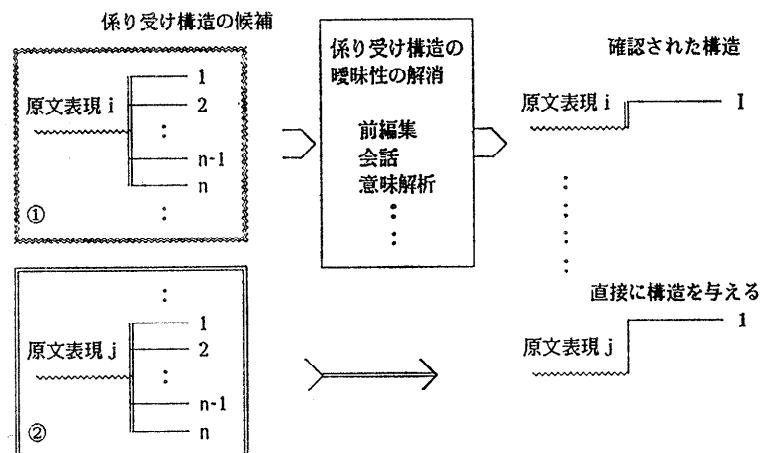


図 1 従来の係り受け構造の曖昧性の処理概念  
Fig. 1 The usual concept for processing structural ambiguities.



- ① 解消すべき係り受け構造の曖昧性
- ② 可保留曖昧関係

図 2 提案の係り受け構造の曖昧性の処理概念  
Fig. 2 The presented concept for processing structural ambiguities.

保留曖昧関係である。そこで、本論文では、日本語の係り受け構造の曖昧性を可保留曖昧関係と解消すべき曖昧関係に分け、可保留曖昧関係に対しては、その曖昧性を解消せずにあらかじめ与えた訳文関数（後に定義する）で直接に訳文を生成する手法を提案する。その処理概念を図2に示す。

## 2. 可保留曖昧関係

### 2.1 日本語の係り受け構造の曖昧性

一つの文節の係り先として二つ以上の文節が存在する場合、係り受け構造の曖昧性があるという。以下、日本語文における係り受け構造の曖昧性の例を示す。

(a) 並列助詞「と」、連体助詞「の」、名詞からなる名詞句の場合：

図3の例2-1では、文法的にも、意味的にも、二つの構文木が成立する。すなわち、「一太郎」の係り先として、「花子」と「子供」の二つの名詞が存在する。

(b) 形容詞からなる文の場合：

図3の例2-2では、形容詞「美しい」の係り先として、「楊貴妃」と「目」の二つの名詞が存在する。

(c) 動詞にかかる場合：図3の例2-3では、「ワープロで」の係り先として、「翻訳した」と「修正する」の二つの動詞が存在する。

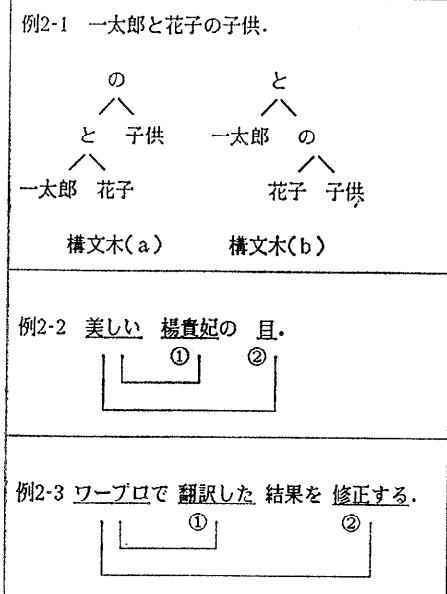


図3 係り受け構造の曖昧性の例  
Fig. 3 Examples of structural ambiguities.

## 2.2 係り受け構造の曖昧性を持つ日本語文の中国語訳

前節の例文を用いて、日中両言語における係り受け構造の曖昧性の対応関係を説明する。

(a) 並列助詞「と」と連体助詞「の」の場合：

例2-1は係り受け構造の曖昧性の持つ日本語文であり、構文木(a)と構文木(b)に対応する中国語文の構文木をそれぞれ図4の構文木(a')と構文木(b')に示す。いずれも、同一の中国語文を得ることができる。

(b) 形容詞の場合：図3の例2-2の日本語文は図5のような中国語文に対応する。「美しい」の係り先が①、②のどちらであると考えても、同一の中国語文を生成することができる。

(c) 動詞にかかる場合：例2-3では、もし「ワープロで」の係り先を「翻訳する」とすれば、中国語訳文は図6の①であり、もし「ワープロで」の係り先を「修正する」とすれば、中国語訳文は図6の②である。したがってこの例では、両者は異なる文を生成する。

以上で述べたように、例文2-1、例文2-2では、い

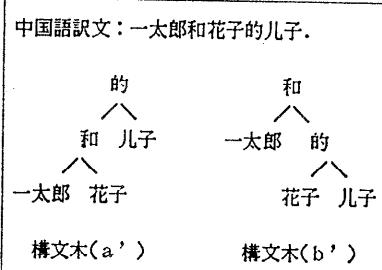


図4 中国語訳文の係り受け構造の曖昧性  
Fig. 4 An example of structural ambiguity in Chinese.

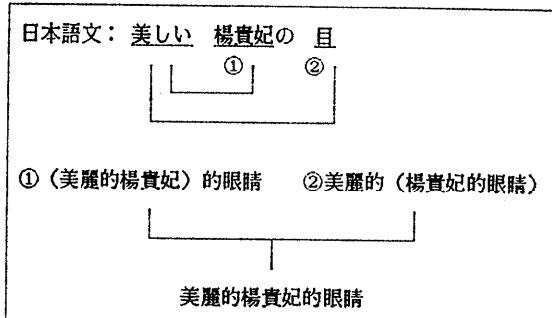


図5 例2-2の中国語訳文  
Fig. 5 The translated Chinese sentence of example 2-2.

それの係り受け構造でも、同一の中国語訳文を生成することができる。すなわち、このような係り受け構造の曖昧性は解消しなくてもかまわない。しかし、例文2-3では、係り先が異なると異なる中国語文になるので、この場合は曖昧性を解消しなければならない。

### 2.3 可保留曖昧関係およびその特徴

上に述べた例文2-1, 2-2のように、係り受け構造の曖昧性を解消しなくとも訳文を生成できる場合がある。前述のように、本論文ではこのような係り受け構造の曖昧性を可保留曖昧関係と呼ぶ。次に可保留曖昧関係の三つの主な特徴について述べる。

a. 可保留曖昧関係は必ず特定の二言語間に定義される。例えば、図7に示すように、例文2-4には係り受け構造の曖昧性があり、2種類の構文が存在する。この二つの構文はその中国語の訳文が同じであるので、日中機械翻訳においては可保留曖昧関係である。しかし、英語ではそれぞれ異なる文に対応するので、日英機械翻訳では可保留曖昧関係ではない。

b. 可保留曖昧関係は双方向性を持つ。図8により、この性質を説明する。図中の日本語文「前節で述べたシステムの機能」には、図中①, ②で示されるような係り受け構造の曖昧性がある。しかし、どちらの構文も同じ中国語訳文に対応するので、これは可保留曖昧関係である。また、この中国語文にも、図中(1), (2)で示されるような係り受け構造の曖昧性がある。

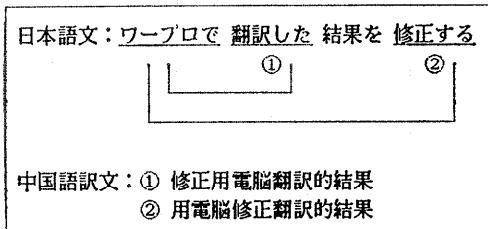


図6 例2-3の中国語訳文  
Fig. 6 The translated Chinese sentence of example 2-3.

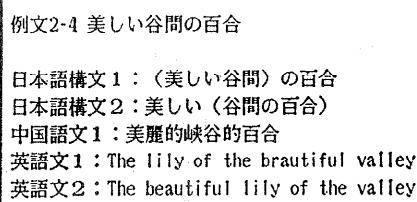


図7 可保留曖昧関係の特徴1  
Fig. 7 A feature of RSA.

しかし、(1), (2)のどちらの構文も同じ日本語訳文に対応するので、これも可保留曖昧関係である。

c. 可保留曖昧関係に対して、あらかじめ与える構文構造によって訳文を生成することができる。この点について、4章で述べる。

### 3. 日中機械翻訳における可保留曖昧関係のパターン

前節で述べたように、日中機械翻訳において、係り受け構造の曖昧性のすべてが可保留ではない。それゆえ、曖昧性を持つ係り受け構造の中から可保留曖昧関係を抽出することが必要である。

本章では、可保留曖昧関係パターンを①並列助詞「と」、連体助詞「の」、名詞からなる句；②形容詞連体形、名詞、並列助詞「と」、連体助詞「の」、名詞からなる句；③動詞連体形、名詞、並列助詞「と」、連体助詞「の」、名詞からなる句の3種類に大別し、その各々について検討する。

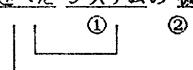
#### 3.1 「NとNのN」および「NのNとN」パターン

並列助詞「と」、連体助詞「の」、名詞からなる句の係り受け構造の曖昧性についてはすでに多数の研究があり、そのパターンの分類と出現頻度も報告されている<sup>4, 12)</sup>。これを表1に示す。我々の検討によれば、表1に示される17種のパターンはすべて可保留曖昧関係である。この17種のパターンはさらに以下の3種に分類される。

パターン1：NとNのN [のN]\*

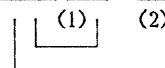
パターン2：[Nの]\* NのNとN

日本語：前節で述べた システムの 機能



中国語訳文①②：在前節叙述的系統的功能  
(a) 日本語→中国語

中国語：在前節 叙述的 系統的 功能



日本語訳文(1)(2)：前節で述べたシステムの機能  
(b) 中国語→日本語

図8 可保留曖昧関係の特徴2  
Fig. 8 A feature of RSA.

パターン3：[Nの]\* NのNとNのN [のN]\*

ここで（以下同様）Nは名詞、「と」は並列助詞、「の」は連体助詞であり、また、

[x]\* は  $\underbrace{x \ x \dots x}_{m}$

を意味する。ここで、mは0または任意の正整数である。

### 3.2 「AN の N」および「AN と N」パターン

「美しい谷間の百合」という文はしばしば係り受け構造の曖昧性の代表的な例としてあげられる<sup>11)</sup>。この曖昧性は日中機械翻訳において可保留曖昧関係である。このような形容詞（形容動詞を含む）に関する可保留曖昧関係パターンには以下に示す2種類がある。

パターン4：AN と N

パターン5：AN の N [のN]\*

ここで、Aは形容詞の連体形である。なお、ここで連体形には形容詞の過去形の連体形も含める。

### 3.3 「VN の N」および「VN と N」パターン

動詞に関する可保留曖昧関係パターンには以下の2種類がある。

パターン6：VN と N

パターン7：VN の N [のN]\*

ここで、Vは動詞の連体形である。

以上、本章で述べた可保留曖昧関係パターンを表2にまとめて示す。また、それぞれの例文を表3に示す。

表1 “の” “と” パターンの分類および出現頻度  
Table 1 The classification and the frequency  
of “no” and “to” patterns.

No.	パ タ ー ン	頻度
1	AとBのC	69
2	AとBのCのD	17
3	AとBのCのDのE	3
4	AとBのCのDのEのFのG	1
5	AのBとC	29
6	AのBとCのD	24
7	AのBとCのDのE	7
8	AのBとCのDのEのF	2
9	AのBとCのDのEのFのG	1
10	AのBのCとD	3
11	AのBのCとDのE	3
12	AのBのCとDのEのF	2
13	AのBのCとDのEのFのG	2
14	AのBのCのDとEのF	1
15	AのBのCのDとEのFのG	1
16	AのBのCのDのEとFのG	1
17	AのBのCのDのEとFのGのHのI	1

### 4. 可保留曖昧関係の処理方式

本章では、我々が開発中の日中機械翻訳システムにおける可保留曖昧関係の処理概要を述べる。

前述のように、機械翻訳では一般的に、係り受け構造の曖昧性があれば、それを解消しなければならない。しかし、もし係り受け構造の曖昧性が可保留曖昧関係であれば、その曖昧性の解消をせずに、訳文関数を用いて訳文を直接生成することができる。ここで訳文関数とは可保留曖昧関係である日本語文の各々のパターンに対応する中国語文のパターンのことである。例えば：

例文 4-1：ラジオと車のバッテリー

この例文のパターンは

「N<sub>1</sub> と N<sub>2</sub> の N<sub>3</sub>」……………<1>

であり、「と」は並列助詞、「の」は連体助詞である。それで、このパターンは可保留曖昧関係である。

例文 4-1 の中国語訳文は「收音機和汽車的蓄電池」であるが、この訳文を形式化すれば、

「M<sub>1</sub> 和 M<sub>2</sub> 的 M<sub>3</sub>」……………(1)

になる。この(1)が日本語パターン <1> に対応する訳文関数である。ここで M<sub>i</sub> は日本語 N<sub>i</sub> に対応する中国語名詞である (i=1, 2, 3)。

可保留曖昧関係の各パターンに対する訳文関数を表2に示す。ここで、用言を含む句では、日本語文の語順と中国語文の語順が異なることを考慮しなければならない。例えば：「ご飯を食べる」という日本語句に対応する中国語句は「吃(食べる) 飯(ご飯)」である。

例文 4-2：骨を食べる犬と猫。

この例文のパターンは、目的語（骨を）を除くと「VN と N」……………<2>

である。ここで、「V」は動詞「食べる」を示している。この動詞「V」に対応する中国語動詞「D」は「吃」である。もし、目的語の位置を変えずに

「D 的 M 和 M」……………(2)

をそのまま適用すると、中国語訳文は

骨頭吃的狗和猫

となる。これは誤った訳文である。

そこで、動詞、形容詞の場合、その語のみならず、その語が関連する文節を対象として処理を行う。

例えば、例 4-2 の場合、「D」の位置は目的語の前に移動し、「吃骨頭(骨を食べる)」となる。表2ではこのような場合には「'」をつけ、「D'」と表記している。

目的語がさらに複数の文節からなる、例えば「(魚の骨を)食べる犬と猫」のような場合は、目的語の解析を行ったうえで、「D'」として「吃魚的骨頭」を生成する必要がある。これら用語の語順変換処理の詳細は別の機会に述べる。

次に、下記の例文 4-3 について、図 9 にその処理の

流れを示す。

例文 4-3：反政府の仏教系と政府支持のカトリック系労組の分裂メーデー。

この例文では、「と」は並列助詞、三つの「の」は連体助詞であり、表 2 のパターン 2 に対応する。そこで、訳文関数 2 を用いて中国語文を生成する。

表 2 可保留曖昧関係の訳文関数と出現頻度  
Table 2 The translation functions and frequencies of RSA.

No.	パターン	訳文関数	出現頻度
1	N と N の N [の N]*	M 和 M 的 M [的 M]*	98
2	[N の]* N の N と N	[M 的]* M 的 M 和 M	65
3	[N の]* N の N と N の N [の N]*	[M 的]* M 的 M 和 M 的 M [的 M]*	45
4	AN と N	X' 的 M 和 M	8
5	AN の N [の N]*	X' 的 M 的 M [的 M]*	128
6	VN と N	D' 的 M 和 M	25
7	VN の N [の N]*	D' 的 M 的 M [的 M]*	443

表 3 可保留曖昧関係パターンの例文  
Table 3 Examples of RSA patterns.

No.	パターン	例文
1	N と N の N [の N]*	記憶部と演算部の科学演算レジスター ラジオと車のバッテリー
2	[N の]* N の N と N	世界の未来の科学と技術 体言の構成と機能情報
3	[N の]* N の N と N の N [の N]*	企業の中の技術革新の進展と職種の変化 反政府の仏教系と政府支持のカトリック系労組の分裂メーデー
4	AN と N	便利な武器と道具 綺麗な桜花と桃花
5	AN の N [の N]*	美しい谷間の百合 不思議な光線の強度
6	VN と N	目をもつ狼と鯨 記憶する金属と宇宙工学
7	VN の N [の N]*	前節で述べたシステムの機能 語の接続関係を利用した日本語文章の誤り検出

最後の訳文      控制器將信号送到存儲器和計算器的科学計算寄存器

日本語：	反政府の仏教系と政府支持のカトリック系労組の分裂メーデー
パターン：	N1 の N2 と N3 の N4 の N5
訳文関数：	M1 的 M2 和 M3 的 M4 和 M5
中国語訳文：	反政府的佛教派和支持政府的迦里庫工会的分裂集会

図 9 可保留曖昧関係の訳文生成  
Fig. 9 An example of the translation process for RSA.

実際には多くの場合、上述のパターンは文の一部（例えば、ある文要素）であり、システムではこの部分の訳文（または訳語）を記憶し、一つの要素として文全体の翻訳を行う。

以下、例文 4-4 によりその概要を示す。

例文 4-4：制御部が信号を記憶部と演算部の科学演算レジスタへ送出する。

まず、入力の日本語文を形態素解析してコード要素を生成する<sup>14,15)</sup>。そして、RSA 検出器を用いて可保留曖昧関係が成立する部分を検出する。本例では、“記憶部と演算部の科学演算レジスタ”が検出される。さらに、検出する可保留曖昧関係部分について、RSA 翻訳器を用いて翻訳を行う。それと同時に、システムはこの結果を一つの要素として、目的言語のコード列に変換し、訳文を生成する。その詳細は文献 14), 15), 16) に譲る。図 10 に以上の処理手順を示す。図 10 では“RSA 处理”は RSA 検出器と RSA 翻訳器を用いる可保留曖昧関係の処理を意味し、“通常の翻訳”は文献 15) で提案した家族モデルを用いた翻訳処理を意味する。

次に、RSA 検出器と RSA 翻訳器について簡単に述べる。

RSA 検出器とは可保留曖昧関係パターンをコード要素列と照合し、一致する部分を検出するものである。例として、例文 4-4 に対する日本語文のコード要素列を表 4 に示す。この例では、表 4 の No. 3～No. 5

と表 2 のパターン No. 1 との照合が成功し、これが可保留曖昧関係として検出される。ついで、この部分を一つの要素（表 5 に e(t1) で示す）として、日本語文の要素列を短縮する。その結果を表 5 に示す。

RSA 翻訳器とは検出された可保留曖昧関係部分について、訳文関数を用いて訳文を生成するものである。例文 4-4 では、検出された部分は“記憶部と演算部の科学演算レジスタ”で、パターンおよび訳文関数は表 2 の No. 1 である。この訳文関数を用い、中国語訳文“存儲器和計算器的科学計算寄存器”が生成され

表 4 例文 4-4 のコード要素  
Table 4 Code elements of Japanese sentence.

No.	コード要素	助詞
1	制御部	が
2	信号	を
3	記憶部	と
4	演算部	の
5	科学演算レジスタ	へ
6	送出する	

表 5 短縮したコード要素  
Table 5 Shortened code elements.

No.	コード要素	助詞
1	制御部	が
2	信号	を
3	e(t1)	へ
4	送出する	

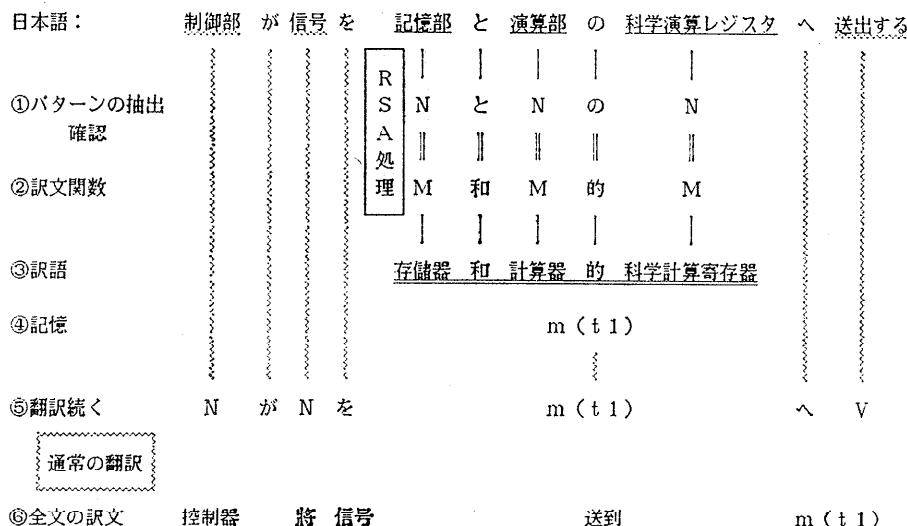


図 10 翻訳の手順  
Fig. 10 The process of translation.

る。図10中のm(t1)はこの訳を示すものである。

## 5. 実験および考察

4章で述べた可保留曖昧関係の処理アルゴリズムを、我々が開発中の日中機械翻訳実験システム<sup>9),15)</sup>のサブシステムとして組み込み、本方式の有効性と正確性を確認する実験を行った。

### 5.1 実験

表6に示す情報処理に関する論文、研究報告および科学技術に関する論説、人物伝記および教科書から可保留曖昧関係の文を抽出して実験を行った。

#### (1) データの情報

- 総文数=2,919文
- 総文字数=197,400文字
- 平均文長=197,400文字/2,919文=68文字/文
- 抽出された可保留曖昧関係文=806文

- 可保留曖昧関係文の出現率=812文/2,919文  
 $\times 100\% = 27.8\%$

表2中の出現頻度はこのデータにおけるパターン別の出現頻度である。

#### (2) 抽出の方法

表6に示す資料の中の一部は計算機で自動的に抽出したが、ほかの資料はすべて手作業で抽出した。なお、抽出したのは可保留曖昧関係に関連する部分である。

#### (3) 実験結果

本実験は翻訳システム全体と分離して行った。評価指標はパターン関連部分の訳語の正確度である。実験結果は2919文中の67文が誤りであった。したがって正訳率は97.7% ((2919-67)/2919) であった。表7に翻訳例の一部を示した。その中にNo.5とNo.6は語順の交換が必要な例である。

表6 抽出実験用資料  
Table 6 Collected data for the experiment.

No.	著者名	出處	文数	No.	著者名	出處	文数
1	島津ほか	情処論 Vol. 27, No. 2	242	10	鈴木ほか	情処論 Vol. 29, No. 4	177
2	平井ほか	情処論 Vol. 31, No. 10	269	11	荒木ほか	情処論 Vol. 28, No. 4	128
3	林ほか	情処論 Vol. 31, No. 10	113	12	任ほか	情処研報 NL 72-7	143
4	向仲ほか	情処論 Vol. 31, No. 8	100	13	任ほか	信学技報 VLD 88-63	137
5	野中ほか	情処論 Vol. 31, No. 8	192	14	任ほか	信学技報 NLC 89-30	139
6	柄内ほか	信学技報 EC 82-20	124	15	任ほか	信学技報 NLC 90-2	122
7	柄内ほか	情処論 Vol. 24, No. 2	124	16	張ほか	“科学与人類” 機械工業出版社	578
8	柄内ほか	情処論 Vol. 26, No. 4	133	17	岡田	“季節風と日本人” 東大外国語学校教材	42
9	柄内ほか	情処論 Vol. 27, No. 3	156				

注: 情処論=情報処理学会論文誌; 情処研報=情報処理学会研究報告; 信学技報=電子情報通信学会技術研究報告

表7 翻訳の例文  
Table 7 Examples of translation.

パターン	日本語文	中国語訳文
1	記憶部と演算部の科学演算レジスタ ラジオと車のバッテリー	記憶部和演算部的科学計算寄存器 收音机和汽车的蓄電池
2	世界の未来の科学と技術 体言の構成と機能情報	世界的未来的科学和技术 体言的構成和機能信息
3	企業中の技術革新の進展と職種の変化 反政府の仏教系と政府支持のカトリック系労組の分裂メーテー	企業中的技术革新的进展和職种的变化 反政府的佛教派和支持政府的加里庫工会的分裂集会
4	便利な武器と道具 綺麗な桜花と桃花	方便的武器和道具 漂亮的樱花和桃花
5	美しい谷間の百合 不思議な光線の強度	美麗的山間的百合 不思議的光線的强度
6	目をもつ猿と鯨 記憶する金属と宇宙工学	具有眼睛的猴子和鲸鱼 記憶的金属和宇宙工学
7	前節で述べたシステムの機能 語の接続関係を利用した日本語文章の誤り検出	在前节叙述了的系統的功能 利用了語的接続関係的日本語文章的錯誤検出

表 8 誤った翻訳の例  
Table 8 Examples of mistaken translation.

No.	日本語文	誤った中国語訳文
1	入力と同一の三組	輸入和同一的三組
2	論文一篇と同程度の長さ	論文一篇和同程度的長度
3	W1 と同音の語 W2	W1 和同音の語 W2
4	変換辞書の初期収録語と同一分野の文献	変換辞書の初期収録語和同一領域的文献

## 5.2 考 察

表 2 に示した可保留曖昧関係パターンについて、5.1 節で述べたように、抽出された可保留曖昧関係文は 812 であり、可保留曖昧関係文の出現率は 27.8% (可保留曖昧関数文/総文数 = 806/2919) である。ここで、一つの文の中に複数の可保留曖昧関係パターンがある場合は、複数の可保留曖昧関係文であるとした。

論文、教科書など 2919 文 (約 197400 文字) から表 2 に示した可保留曖昧関係パターンを抽出し、実験を行った結果、正訳率は 97.7% であった。

本手法では可保留曖昧関係に対して従来のような曖昧性の解消、係り受けの同定をせず、あらかじめ与えられる訳文関数を用いて訳文を生成している。それで、従来の手法と比べて、表層の処理で十分な翻訳精度が得られるので高速性が期待できるなどの利点が考えられる。また、本手法のその他の特徴および性能の定量的検討は今後の課題である。

もちろん、本手法により検出される可保留曖昧関係が完全に翻訳できるわけではない。表 8 にいくつかの誤った訳文を示した。その誤りの原因はパターンの判断、とくに並列助詞 “と” の確認を誤ったためであると考えられる。表 8 に示した例文の “と” は通常の並列助詞 “と” の機能と異なる。このような問題の解決策としては並列助詞の確認ルールの付加あるいは名詞の意味属性の利用などが考えられる。また、現在のシステムでは、多品詞および品詞自体の多義性を考慮していないので、長い並列句および多品詞問題などを含む多数の文を対象とした場合、本手法の適用性をさらに検討する必要があると考える。

## 6. む す び

日中両言語は同族の言語とは考えられていない<sup>13)</sup>。しかし、文のある部分、例えば、並列助詞 “と” と連体助詞 “の” からなる文では両言語の構造は同じである。本論文では、機械翻訳の観点から上述の特徴に着目し、日中機械翻訳における可保留曖昧関係およびそ

の翻訳方式を提案した。さらに、実験により本論文で提案した方式の有効性と翻訳精度を確認することができた。

今後の課題として、可保留曖昧関係パターンの系統的な抽出方式の検討、並列助詞 ‘と’ の有効な判断手法および翻訳システム全体への組み込みなどを考えていく。

**謝辞** 日ごろ有益なご討論、ご助言を頂く研究室各位に感謝いたします。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頂いた斎川勝男技官に感謝の意を表します。

## 参 考 文 献

- 1) 牧野武則：機械翻訳、オーム社 (1989).
- 2) COLING-90, Computational Linguistics (1990).
- 3) 向仲 順：技術文書の機械翻訳における常識と文脈情報の利用、情報処理学会論文誌、Vol. 31, No. 8, pp. 1168-1173 (1990).
- 4) 田村直良、田中穂積：意味解析に基づく並列名詞句の構造解析、情報処理学会研究会報告、NL 59-2 (1987).
- 5) 平井章博、梶 博行、芦沢 実：機械翻訳向け前編集のための日本語係り受け構造の曖昧性検出方式、情報処理学会論文誌、Vol. 31, No. 10, pp. 1425-1437 (1990).
- 6) 有田英一、福島正俊、進藤靜一：日英機械翻訳システムにおけるプリエディットについて、情報処理学会研究会報告、NL 48-7 (1985).
- 7) 長尾 真、田中伸佳：制限文法に基づく文書作成支援システム、情報処理学会研究会報告、NL 44-5 (1984).
- 8) Tomita, M.: Sentence Disambiguation by Asking, *Computers and Translation*, Vol. 1, No. 1, pp. 39-51 (1986).
- 9) 任 福継、宮永喜一、柄内香次：日中常用文型機械翻訳システム、信学論 (D-II), Vol. JT 4-D-II, No. 8, pp. 1060-1069 (1991).
- 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) 東条 敏：自然言語処理入門、近代科学社 (1988).
- 12) 稲垣博人、壁谷喜義、小橋史彦：意味連結パターンを用いた係り受け解析、情報処理学会研究会報告、NL 67-5 (1988).
- 13) 望月十八吉：中国語と日本語、光生館 (1974).
- 14) 任 福継、宮永喜一、柄内香次：コード方式日中機械翻訳の実験システム JCMT の概要、情報処理学会研究会報告、NL 72-7 (1989.5).
- 15) 任 福継、茫 莉馨、宮永喜一、柄内香次：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、情報処理学会論文誌、Vol. 32, No. 10,

- pp. 1249-1258 (1991).
- 16) 任 福継, 宮永喜一, 栄内香次: 意味属性による中国語補助語の推定アルゴリズム, 情報処理学会論文誌, Vol. 32, No. 11, pp. 1374-1382 (1991).  
 (平成3年4月1日受付)  
 (平成5年5月12日採録)



任 福継 (正会員)

1959年生. 1982年中国北京郵電学院電信工程部卒業. 1985年同大学院計算機応用専攻修士課程修了. 1987年中国科学院博士課程中退來日留学. 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. 現在CSK技術開発本部研究員. 1992年より中国大連理工大学教授兼務. 計算機科学, 人工知能, 自然言語処理等の研究に従事. 人工知能学会会員. ACSJ (The Association for Chinese Scientists in Japan) 会長.



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. 助理工程师. 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工学専攻修士修了. 工学博士. 現在, 北海道大学工学部電子工学科助教授. 並列計算機システム, デ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.



栄内 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. 現在同大学院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.