

関連語を用いた文の分解に基づく中日機械翻訳システム

范 莉 馨[†] 任 福 繼^{††}
宮 永 喜 一[†] 栃 内 香 次[†]

中国語文の実際上の使用度から見れば、単文は極めて少なく、複文の方が多い。従って機械翻訳システムの実用という見地から言えば、複文の翻訳は極めて重要である。本論文では、中国語複文と日本語文との対応関係を検討し、関連語を用いて中国語複文をいくつかの単文に分解し、それぞれの単文の翻訳結果から最終の訳文を生成する手法を提案する。また、関連語の処理を効率的に行うために、多数の中国語教科書および科学技術文献から関連語に対応する訳文の文型を抽出してそれらを訳文関数と名付け、入力された中国語文中で関連語が識別されたなら、詳細な文法解析を行わず、訳文関数を用いて直接訳文を生成している。さらに、関連語の多義性を解消するため、関連語に関係する要素の意味属性を用いて多義性の解消を行っている。そして、提案手法に基づく中日機械翻訳システムを構築し、実験により本手法の有効性を確認した。

A Chinese-Japanese Machine Translation System Using Sentence Partitioning

FAN LIXIN,[†] REN FUJI,^{††} YOSHIKAZU MIYANAGA[†] and KOJI TOCHINAI[†]

For the machine translation from Chinese to another languages, it is important to analyze Chinese complex sentences because the major portion of Chinese sentences are complex. This paper proposes a method of Chinese-Japanese machine translation based on the relations between complex Chinese sentences and corresponding Japanese sentences. In this method, a Chinese complex sentence is decomposed to several simple sentences with "related words" that express semantic relations between simple sentences. Translations of each simple sentence are composed to a Japanese sentence by using of the characteristics of the related words. For efficient translation, a number of Chinese scientific documents and textbooks were investigated to extract relations between sentence patterns, related words and their Japanese translations, and the results were registered into a dictionary. By the use of these relations, complex sentences that have same patterns and related words in the dictionary are translated directly into Japanese without precise syntactic and semantic analysis. An experimental translation system was constructed and several experiments were carried out. The results show the effectiveness of the proposed method.

1. ま え が き

機械翻訳の研究は長い歴史をもち、いくつかの商用システムが提供されるようになってきた^{1)~6)}。中国では、機械翻訳に関する研究は1958年に始まり⁷⁾、1980年代に英中機械翻訳システムの本格的な開発が開始され、最近になって実用的な英中機械翻訳システムが実現している^{8),34)}。しかしながら、中日両言語間の機械翻訳に関しては本格的な研究が開始されたばかりであり、日英・英中言語間の機械翻訳と比較すると、未開拓の部分が極めて多い^{9)~17)}。

一般に、中国人が日本語文を翻訳する際、常用文型¹⁹⁾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れる。この方法に基づき、我々は先に日中常用文型機械翻訳システムを構築し、実験により、提案した処理手法の有効性が確かめられた¹⁹⁾。しかし、この手法をそのまま中日機械翻訳に適用することはできず、特に中国語複文の処理が困難である。

中国語文の実際上の使用度から見れば、単文は極めて少なく、複文の方が多いと考えられる^{21),22)}。従って機械翻訳の実用化という見地から言えば、複文の処理が極めて重要である。ここで、複文とは意味的關係のある二つまたは二つ以上の単文が集まったものである。複文を構成する単文の間はそれらの意味的關係を

[†] 北海道大学工学部
Faculty of Engineering, Hokkaido University

^{††} 広島市立大学情報科学部
Faculty of Information Sciences, Hiroshima City University

表す単語が繋がる。

複文は単文間の意味的關係によって付録1に示される基本種類に分けることができる。さらに、これらの基本種類から多重構造、派生により多数の複文が生成される。

中国語複文の一例を以下に示す。

例文①：因為我們是為人民服務的^A，所以我們如果有缺点和錯誤^B，就不怕別人批評指正^C。

例文①は二重複文で、この文には三つの単文がある。

単文A：“我們是為人民服務的”；

単文B：“我們有缺点和錯誤”；

単文C：“[我們] 不怕別人批評指正”。

その中で、単文Aと単文B、Cの間は原因・結果の關係であり、単文Bと単文Cの間は仮定の關係である。訳文①：私達は人民に奉仕するため^A、もし欠点と誤りがあったら^B、他人の批判や指摘を恐れない^C。

上記のような複文に対する機械翻訳手法はこれまでほとんど研究されていない。これは以下に示すような中国語の特徴により、中国語複文に対する構文解析、翻訳処理が非常に難しいためであると考えられる。

①中国語では、単語と単語の間には、単語を区別できるスペースがなく、日本語のように助詞、助動詞などの機能語を平仮名で表して構文要素を区切ることもない。

②中国語では、動詞、助動詞、形容詞などの語尾変化、すなわち活用現象もない。

③中国語では、兼用品詞（二つ以上の属性をもつ単語）が多い。また、兼用品詞は形態上の区別がない。

④英文法では、時制と動態は関連的に用いられている。また日本語では、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)を表す語・句(状況時)一で表される。従って、過去時制は過去の時間を表す語・句で表され、未来の時制は未来の時間を示す語・句で表される。

我々は大量の中国語複文の調査結果から、複文を構成する各単文の間には一般に特定の“連接要素^{*}”が存在することを見出した。例文①の中の下線(__ および ~) を付けた単語はこのような連接要素を示している。これは日中機械翻訳における“常用文型”¹⁹⁾に類似しているが、構造、意味ともに完全には一致しないので、本論文では、この連接要素を「関連語」と呼ぶ。

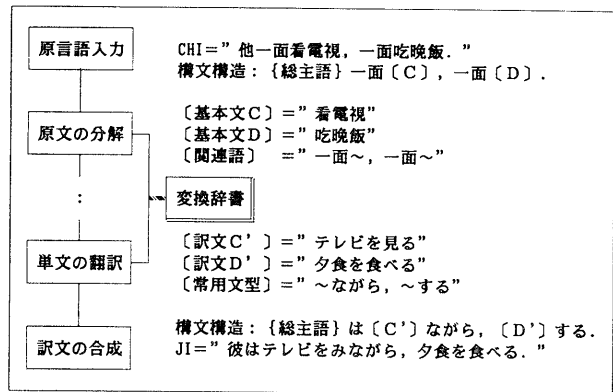


図1 中日機械翻訳システム概要図

Fig. 1 The concept of Chinese-Japanese machine translation system.

上述の結果に基づき、本論文では、中国語複文と日本語文との対応關係を検討し、関連語を用いて中国語複文をいくつか独立な単文に分解し、さらに関連語の性質を用い、各単文の翻訳結果から最終の訳文を生成する手法を提案する。また、関連語の処理を効率的に行うために、多数の中国語教科書および科学技術文献から関連語に対応する訳文の文型を抽出してそれらを訳文関数と名付け、入力された中国語文中で関連語が識別されたなら、詳細な文法解析を行わず、訳文関数を用いて直接訳文を生成している。さらに、関連語の多義性を解消するため、関連語に關係する要素の意味属性を用いて多義性の解消を行っている。

以上の考察に基づく中日翻訳実験システムを構築した。その概要を図1に示す。さらに、実験により本手法の有効性を確認した。

以下、第2章では中日機械翻訳における関連語、第3章では文の分解に基づく中日機械翻訳手法、第4章では翻訳実験および検討、第5章では今後の課題について述べる。

2. 中日機械翻訳における関連語

本章では、本論文で提案した手法の中核である“関連語”について述べる。ここで、“関連語”は単文と単文を連接するための副詞、連接詞などの要素を一括

* 例外的には連接要素の省略された中国語複文も存在する。例えば中国語文“小張患了感冒, 没有去学校 (張さんは風を引いたので, 学校を休んだ)”。この文は次の三つの文と意味が同じである。

文①：“小張[因為]患了感冒, 没有去学校”

文②：“小張患感冒, [所以] 没有去学校”

文③：“小張[因為]患了感冒, [所以] 没有去学校”

本論文では、このような“連接要素”が省略された複文は処理対象外とし、検討は他の機会に譲る。

して指す。また、関連語に対応する日本語の訳語を常用文型²⁶⁾と言う。関連語は共起関係により一語性と多語性に分けられる。

2.1 一語性関連語

[定義1]: 中国語文においてただ一個所で出現し、一回しか出現しない関連語を一語性関連語と言う。

例文②: 那天天气不好, 因此我們没去長城。

例文③: 由於漢字結構複雜, 書写不方便。

上例中の「因此 (yinci)」, 「由於 (youyu)」は一語性関連語であり、その訳文関数²⁶⁾を〈1〉, 〈2〉式に示す。

A, 因此 B → A', そのために B'. ……〈1〉

由於 A', B → A' ので, B'. ……〈2〉

上式中の A', B' は中国語文 A, B の日本語表現を意味する。

上例の訳文は次のとおりである。

訳文②: その日は天气が悪く, そのために私達は万里の長城へ行かなかった。

訳文③: 漢字は構造が複雑なので, 書くのは便利ではない。

2.2 多語性関連語

[定義2]: 複文中の異なる個所に出現している複数の関連語を多語性関連語と言う。

例文④: 這只船不僅能在水上航行, 而且能在雪地航行。

例文⑤: 人数既多, 意見又不一致。

上例中の「不僅 (bujin) ~, 而且 (erqie) ~」と「既 (yi) ~, 又 (yu) ~」が多語性関連語であり、その訳文関数を〈3〉, 〈4〉式に示す。そのうち、A, B, C, D は中国語単文、特に C, D は主語のない単文を意味している。それぞれ対応する日本語訳文は A', B', C', D' で表す。

[主]不僅 C, 而且 D → [主] C' だけでなく, D'.

……〈3〉

既 A, 又 B. → A' し, B'. ……〈4〉

上例の訳文は次のとおりである。

訳文④: この船は水上を航行できるだけでなく, 雪原を航行できる。

訳文⑤: 人数が多いし, 意見が一致しない。

付録2に常用の一語性関連語と多語性関連語を示す。また、付録2の中の「多/一」欄の記号「可」は一語性関連語と多語性関連語との両方に分類され、ある場合は一語性、ある場合は多語性になるものを示す。例えば「因為 (yinwei)」, 「所以 (suoyi)」はこの種類

に属する。

2.3 関連語に対応する訳文の аспекト

次に関連語に対応する日本語訳文の аспекトの処理について述べる。

例文⑥: 既然他說忙, 我們就找別人。

例文⑥で「既然 (jiran) ~, 就 (jiu) ~」は多語性関連語である。例文⑥には時態をはっきり表す助動詞着 (zhao) / ている, 「了 (liao) / した, 「過 (guo) / したことがある」などの機能語が現れていないが、本システムでは関連語の性質により、訳文関数中に中国語文の аспекトなど意味を表す日本語文表現情報を入れ、深く意味解析を行わず訳文関数だけで対応する日本語訳文を生成することができる。すなわち例文⑥の訳文関数は次のようになる。

既然 A, 就 B. → A' 〈た〉からは, B'.

上式中、上記の аспекトの日本語文表現情報は「〈た〉」で、これは「動詞+た」、すなわち過去型を意味する。それに従って対応する日本語動詞の語尾変換を行うと、次に示す訳文が得られる。

訳文⑥: 彼が忙しいと言ったからには我々は他の人を頼むしかない。

3. 翻訳アルゴリズム

本章では関連語を用いた文の分解に基づく中日機械翻訳システムの流れを述べる。以下3.1では例文を用いて文全体の翻訳の流れ、3.2では関連語の抽出および中国文の分解、3.3では基本文の翻訳手順、3.4では関連語の変換および多義性解消の手法、3.5では訳文の合成について述べる。

3.1 翻訳の流れ

以下、例文を用いて中日翻訳システムの概要を説明する(図2を参照)。

例文⑦: 機械翻訳は自然言語処理的一部分, 所以自然語言處理的範圍更廣。

まず、形態素解析を行う。その結果を表1に示す。次に中間処理を行う。その結果を表2に示す。この処理過程では、単語「所以」の属性“c3”(cは関連語の属性記号、c1は一語性関連語、c2は多語性関連語、c3は両方ともに属する関連語の記号である)なので、スタックにこの「所以」を入れる。

次に、スタック中の要素を取り出して関連語テーブルの項目と照合する。また、複文分解が確認され、この関連語を抽出し、単文に分解する。その結果を以下に示す。

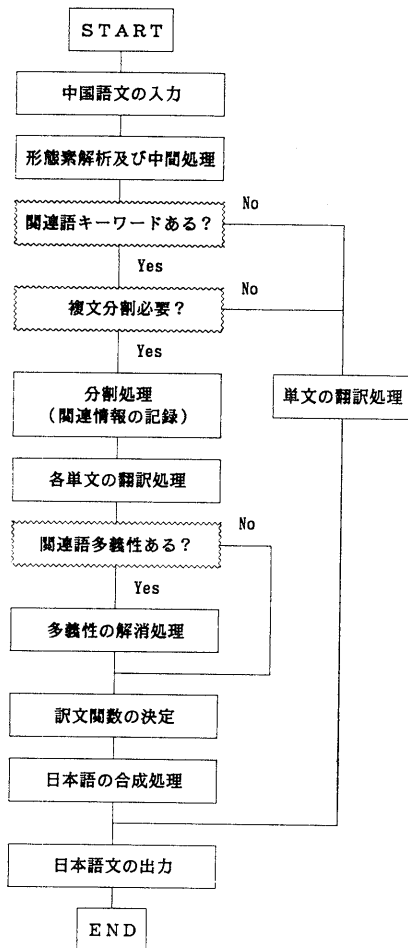


図 2 翻訳システムの流れ
Fig. 2 Flowchart of CJMT system.

単文 A：機械翻訳は自然言語処理的一部分。

単文 B：自然言語処理的範疇更廣。

関連語：所以

次は単文の翻訳処理を行う。その結果、下記の日本語訳文が得られた。

訳文 A'：機械翻訳は自然言語処理の一部分である。

訳文 B'：自然言語処理的範疇は更に広い。

引き続き関連語の変換処理を行う。関連語「所以」は多義性がないので、直接に訳文関数を取る。

A, 所以 B. → A', それで B'.

……〈7〉

表 1 例文②の形態素解析の結果

Table 1 The result of morphological analysis.

No	単語	属性記号	No	単語	属性記号
1	機器	n 4	10	所以	c 3
2	翻訳	v n	11	自然	a 2
3	是	v 7	12	語言	n 4
4	自然	a 2	13	處理	v n
5	語言	n 4	14	的	u 1
6	處理	v n	15	範疇	n 0
7	的	u 1	16	更	d 3
8	一	m 1	17	廣	a 1
9	部分	n 0			

表 2 例文②の中間処理の結果

Table 2 The result of compound words composition.

No	単語	属性記号	No	単語	属性記号
1	機器翻訳	n 9	7	自然語言處理	n 9
2	是	v 7	8	的	u 1
3	自然語言處理	n 9	9	範疇	n 0
4	的	u 1	10	更	d 3
5	一部分	n 9	11	廣	a 1
6	所以	c 3			

最後に、〈7〉式の訳文関数と単文の翻訳結果を用いて最終の日本語訳文を合成する。本例では形態的な変化がないので、訳文は次のとおりである。

訳文⑦：機械翻訳は自然言語処理の一部分である。それで自然言語処理的範疇は更に広い。

以下では各処理部の概要について述べる。

3.2 関連語の抽出および中国語文の分解

本節では、関連語の抽出および中国語文の分解アルゴリズムについて述べる。関連語を抽出するために、関連語テーブルを設けている。その構造を表 3 に示す。

表中、“キーワード 1”は文の分解および関連語の判断に使用される情報である。一語性関連語についてはそれ自身であり、多語性関連語については前半部分で

表 3 関連語テーブル情報 (一部)

Table 3 The connective words table.

関連語 キーワード 1 呼応ワード 2		複文分解の条件	日本語表現	訳文関数 類別	MC
不僅	而且	不僅 A, 而且 B。 S 不僅 C, 而且 D。	～だけでなく	③	—
一面	一面	S 一面 C, 一面 D。	～くしながら	④	—
才	—	—	—	—	20
……					

ある。“呼応ワード2”は多語性関連語の後半部分である。“MC”は一つの関連語が複数の日本語表現と対応する場合に使用される訳文条件テーブルへのポインタである。“訳文関数類別”には、対応する訳文関数の種類が記録されている²⁵⁾。また“複文分解の条件”には文の分解の条件を記録している。

以下、図1を用いてその処理の流れを説明する。

本例文には“一面”という単語は関連語を示す属性記号が辞書の属性欄に登録されているので、この“一面”を関連語テーブルのキーワードと照合する。本例では一番先の“一面”と関連語テーブルのキーワードとの照合が成功するので、引き続き呼応ワードとの照合を行う。表3から分かるように、キーワード“一面”の呼応ワードが“一面”である。これを検索対象として入力した文の中で探索する。勿論、本例では検索も成功する。次に、複文分解の条件を参照して、関連語“一面～一面”を抽出すると同時に、複文を二つの単文に分解する。ここに複文分解の条件の設定はシステム処理の効率化と実現の容易さのためである。

3.3 単文の翻訳手順

単文の翻訳手順は大別すると、①中国語解析、②中日変換、③日本語生成、の三つの部分から成り立っている。以下、各部分の要点を述べる。

3.3.1 中国語文解析

●形態素解析：

本システムでは最長距離マッチング法を用い形態素解析を行っている^{29),30)}。この手法の要点は、辞書中で字頭が同じ単語について、距離(語長)の長い順に入力文と照合を行うことである。それゆえ、入力文中に一致する単語があれば、必ず最長距離の語が見出されることになる。これにより、複合語の処理などが容易になる³¹⁾。

●中間処理

機械翻訳において文の解析を高精度で、かつ効率よく行うという観点から、できるだけ複合語単位で処理することが望ましい¹⁸⁾。例えば、

例文③：機器翻訳は自然言語処理的一部分。

3.3.4で述べるように、“機器翻訳”、“自然言語処理”、“一部分”を複合語とする処理が構文解析の際に極めて有効である。そのために、すべての複合語を固定化して辞書に登録することが考えられるが、これは不適当である。その理由は、

a. すべての複合語を登録することが実際に不可能であること；

b. 常に複合語として処理することができない場合がある；
という二つの理由からである。後者について次にその例を示す。

例文⑨：他用機器翻訳人民日報。

ここで、“機器翻訳”は例文⑧に示すように複合語である。一方、“翻訳”は動名詞(動詞, 名詞いずれにもなる多品詞語を動名詞という^{20),21)})である。もし“機器翻訳”という複合語を辞書に登録して常に用いることにすると、形態素解析の結果は[他, 用, 機器翻訳, 人民日報](彼, で, 機械翻訳, 人民日報)になり、これは誤りである。正しい結果は[他, 用, 機器, 翻訳, 人民日報](彼, で, 機械, 翻訳する, 人民日報)である。すなわち、この例文では“翻訳”は動詞で、文の述語になっている。それゆえこのような誤りを避けるためには、動名詞からなる複合語を固定的に辞書に登録せず、必要な場合にのみ複合語化することが望ましい。すなわち例文⑧の中の“機器”と“翻訳”は複合語“機器翻訳”を合成するが、例文⑨では合成せず、そのまま二つの単語として解析することが必要である。我々は大量の中国語テキストから抽出した単語間連関係に基づき、付録3のような複合語合成ルールを定めている²⁹⁾⁻³¹⁾。例えば、例文⑧における“機器”と“翻訳”については、複合語合成ルール No. 3 (付録3)により、“機器翻訳”を複合語として合成するが、例文⑨における“機器”と“翻訳”については、下記の例外規則により、合成をせず二つの単語として処理する。

規則：単文中に1個の動詞あるいは動名詞、形容動詞(形容詞, 動詞いずれにもなる多品詞語)しか存在せず、かつ文末の単語が形容詞でなければ、この動詞は文の述語であり、他の語と合成して複合語になることができない。

●構文解析

複合語合成ルールによる中間処理を行うので、複数の解釈が可能な単語、特に文の述語になりうる動詞の量が少なくなり、構文上の曖昧性を減少あるいは解消することができる。構文解析は主に文型照合法を用いて行われている^{23)-25),32),33)}。

ここで文型照合法とは、あらかじめ中国語文型のパターンを用意し、解析の際に、それと照合し、成功すれば直接文型を確定するという手法である。例えば、文型“～是～”のパターンとして、“名詞句/是/名詞句”と“名詞句/名詞句”がある。例文⑧は“～是～”

中国語構文：機器翻訳/是/(自然言語処理)的一部分。



日本語訳文：機械翻訳は自然言語処理の一部分である。



3.4 関連語の変換および多義性の解消

関連語の変換は訳文関数を用いて行っているが、関連語は多義性と多訳性が存在するので、それを解消する必要がある^{27),28)}。本手法では多義性と多訳性を厳密に分ける必要がないので、両方をともに多義性と呼ぶ。

例えば、中国語の関連語「才(cai)」は次に示す四つの日本語の意味をもつ。

- (1) やっと
- (2) <た>ばかり
- (3) わずかに
- (4)こそ…んだ。

例文⑮：試了幾回，才成功。

訳文⑮：何回か試みて，やっと成功した。

例文⑯：你才来就走，忙什么？

訳文⑯：あなたは来たばかりなのに，何を忙しくしているか？

例文⑰：這個小孩兒才六歲，懂得什么。

訳文⑰：この子はわずかに六才なんだから，何が分かっているか。

例文⑱：這樣作，才能得出正確的結果呢。

訳文⑱：このようにしてこそ正しい結果が得られるんだ。

これらの多義性は関連語の直前または直後にある単語（以下、それぞれ前、後共起語と言う）の意味属性を利用して解消することができる。

例文⑮中の「才」は一定の時間または時間間隔を経て、やっと得た結果を表し、その前共起語は時間または回数を表す単語である。本例文では「幾回(jihui)/何回か」である。従ってその訳語は“やっと”である。

例文⑯中の「才」はある動作が発生したばかりであることを強調するために用いられ、「才」の後には副詞「就」がある。その訳語は“<た>ばかり”である。

例文⑰中の「才」は数量の少ないことを表すもので、その後共起語は数量を表す単語である。本例では「六歲(liusui)/

六才」である。

例文⑱中の「才」は一定の条件がなければ、結果が出ないことを表す。その後共起語としては文末には語気助詞「呢」がつく。その訳語は“～してこそ…んだ”である。

本システムにおいて、関連語多義性の解消は文献26), 19)で提案した訳語条件テーブルを用いて行っている。すなわち、多義性をもつ関連語について訳語条件テーブルを用意し、前後共起語の意味属性と訳文関数との対をこのテーブルに記録している。関連語「才」の訳語条件テーブル構造を表4に示す。

3.5 訳文の合成

単文翻訳の結果と関連語の訳文関数を用い、日本語訳文の合成を行っている。この際考慮すべき主要な点は、次の2点である³⁵⁾。

(1) 総主語と小主語

[定義3]: 複文の主語を総主語と呼ぶ。本論文では括弧 { } で表記する。

[定義4]: 単文の主語を小主語と呼ぶ。本論文では括弧 [] で表記する。

例えば：下記二つの例文における関連語は「不僅～、而且～」である。しかし例文④における「這只船」は関連語「不僅」の前に位置するため、文の総主語になる。一方、例文⑨における「太郎」、「花子」はそれぞれ関連語の後ろに位置するため、単文の主語、すなわち小主語になる。上述のようにこの問題は関連語の位置情報を用いて処理している。

例文④：{這只船} 不僅能在水上航行，而且能在書地航行/{この船} は水上を航行できるだけでなく，雪原を航行できる。

例文⑨：不僅[太郎] 会網球，而且[花子] 也会網球。/テニスは[太郎] ができるだけでなく，[花子] もできる。

(2) 接続および形態変化

次に接続と形態変化処理について述べる。日本語は

表4 訳語条件テーブル（関連語「才」）
Table 4 The translation Words-Conditions table.

中国語表現	意味属性 (A)	意味属性 (B)	付加条件	日本語表現	ほか
A才B	出現回数	—	—	A'やっとB'	
	主 体	動作	—	A'<た>ばかりB'	
	—	数量	—	A'わずかB'	
	—	感嘆詞	—	A'こそB'んだ	

表 5 実験用文献および教科書一覧
Table 5 The documents of the experiment.

No	書名	著者	出版社	出版年月
1	实用現代漢語語法	劉月華, 他	中国外語教学与研究出版社	1986年10月
2	現代漢語語法教程	陳国梁	中国西安市西安交大出版社	1986年05月
3	JAPANESE IN THIRTY HOURS BY EIICHI KIYOOKA 中訳本	清岡映一著, 周炎輝訳	中国湖南科学技術出版社	1981年05月
4	中日交流標準日本語 2 中級	中国人民教育出版社と日本光村 図書株式会社共同編著	中国人民教育出版社	1990年01月
5	中国語文 3 月号	中国語文雑誌社編集部	中国国際図書貿易総公司	1983年03月
6	中国語文 4 月号	中国語文雑誌社編集部	中国国際図書貿易総公司	1983年04月

表 6 翻訳の例
Table 6 The illustrations of translation.

No	中国語文	日本語訳文
01	這也是新書, 那也是新書。	これも新書だし, あれも新書だ。
02	人數既多, 意見又不一致。	人数が多いし, 意見が一致しない。
03	他們一邊跳舞, 一邊唱歌。	彼らは踊りを踊りながら, 歌を唱う。
04	他不是看書, 就是睡覺。	彼は本を読んでなければ, 寝る。
05	他不是日本人, 是中国人。	彼は日本人ではなく, 中国人だ。
06	我一說那件事, 他就明白了。	私があの事を話すと, 彼はすぐ分かった。
07	他不但是文学家, 而且還是政治家。	彼は文学者だけでなく, 政治家だ。
08	因為天氣不好, 所以我們沒去杭州。	天氣が良くないので, 杭州へ行かなかった。
09	天氣很惡, 因此我們沒去長城。	天氣が悪く, そのために万里長城へ行かなかった。
10	既然有事, 也可以回去。	用事があるんだから, 帰ってもいい。
11	如果下雨, 我們就不去參觀。	雨が降れば, 私達は見学に行かない。
12	假如我有計算機的话, 就不会錯了。	もしも私は電子計算機があったら, 間違いはずがない。
13	他雖然後悔, 但是没有方法挽救。	彼は後悔するが, しかし救済する方法がない。
14	就是下大雨, 我也必須去。	たとえ大雨が降っても, 私は行かなければならない。
15	只有掌握日語, 才能很好地研究日本文学。	日本語をマスターしてこそ, 日本文学をよく研究できる。
16	我只要有時間, 就和妳一起去。	私は時間がありさえすれば, あたなと一緒に行く。
17	無論買什麼, 都要講價錢。	何をかうと問わず, いつも値ぎろうとする。
18	不管有什麼困難, 我們都能克服。	どのような困難があろうとも, 私達は克服できる。
19	寧可失去自由, 不願放弃真理。	自由を失うとも, 真理を捨てることはしない。
20	与其隨便下結論, 不如認真討論。	やたらに結論を出すより, 真剣に討論する方がいい。

中国語と異なり、動詞、助動詞、形容詞の接続関係により形態的な変化をすることがある。例えば、図1に示した例文「他一面看電視，一面吃晚飯。」では関連語「一面～，一面～」の訳文関数は〈5〉式で示される。

[主語]一面C，一面D。→C'ながら，D'。

……〈5〉

ここで：C=看电视；C'=テレビを見る；

D=吃晚飯；D'=夕食を食べる；

もし〈5〉式をそのまま使うと、「彼はテレビを見るながら，夕食を食べる。」という誤った訳文になる。

すなわち「ながら」の接続関係によって、動詞「見る」の形態を変える必要がある。本システムでは変化すべき情報を直接訳文関数文に登録し、形態変化処理を行っている。上の例では、〈5〉式の中でC'における文末の動詞を連用形に変換する記号を用意し、それにより処理している。

[主語]一面C，一面D。→C'くししながら，D'。

……〈6〉

訳文：彼はテレビを見ながら，夕食を食べる。

4. 翻訳実験と検討

前章で述べた手法に基づく中日機械翻訳実験システムを構築し、実験により本論文で提案した手法の有効性を確認した。

実験用の辞書には、実験に用いた例文に出現する単語を中心に3,270語を登録した。なお、この辞書には品詞自身の多義性は考慮していない。各単語が1対1に登録している。

表5に示す資料から関連語150個を抽出し、手作業で関連語テーブルおよび訳語条件テーブルを作成した。そのうち、一語性関連語が26語、多語性関連語が99語で、その他が25語であった(付録2参照)。

システムを上述のように初期設定し、関連語を含む複文200文に対して翻訳実験を行った。その結果、全

文の正翻訳率は74%、関連語だけの正翻訳率は91%であった。後者は関連語に対応する日本語の表現が正しいことを意味し、前者は関連語だけでなく、文全体の日本語表現が正しいことを意味する。

前者の正翻訳率が低い原因は、①品詞自身の多義性と、②実験システムでは長い状況語および補語に対する変換処理能力が弱いことなどであるが、①は誤訳文の約30%、②は誤訳文の約65%であった。また、関連語変換の誤りはすべて多義性をもつ関連語で生じている。その原因は前共起語と後共起語の意味属性の誤判定のためであり、①訳語条件テーブルに登録された意味属性の項目の精度不足、②意味属性の推定機能の不足によるものと考えられる。しかし、関連語について中国語教科書などに記載されている代表的なものを網羅しているため、実用的には十分であると考えられる。

表6にいくつかの翻訳例を示す。実験結果と以上の分析から、関連語の翻訳は極めて良好であると判断され、本論文で提案した関連語を用いた文の分解に基づく中国語複文翻訳手法の有効性が確認された。

5. む す び

本論文では中国語文の構造上重要な接続役を果たす関連語の特性について検討し、この関連語を用いた文の分解に基づく中日機械翻訳手法を提案した。さらに、この手法に基づく実験システムを構築し、実験により本手法の有効性を確認することができた。

本手法の要点としては、1)長い複文を短い単文に分解すること、2)関連語を用いて複文の分解を行うこと、3)関連語についてその訳文条件を調べ、訳文関数を求めること、4)入力文中で関連語が識別されたなら、詳細な文法解析を行わず訳文関数を用いて直接訳文を生成すること、5)関連語に関係する要素の意味属性によって関連語の多義性を解消すること、が挙げられる。本手法の主な貢献としてはいままで解決できなかった中国語複文の解析および日本語訳に対し有効な方法を提案したものと考えられる。

複文中の関連語が省略されたものについては本手法ではまだ処理できない。これを解決するためには、省略された関連語の復元が必要であるが、これについては引き続き実データから情報を抽出し、推定アルゴリズムを開発する予定であり、今後適当な機会に改めて報告したい。また、現在の実験システムでは、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。さらに、長い状況語、補語の解析方法および訳文関数の種類、訳文条件の精密化および意味属性の体系化、意味属性範疇のより詳細な検討などが今後の課題である。

参 考 文 献

- 1) 長尾 真: 機械翻訳サミット, オーム社(1989).
- 2) Hutchins, W. J.: Machine Translation and Machine-aided Translation, *J. Doc.*, Vol. 34, No. 2, pp. 119-159 (1979).
- 3) Snell, B. M. (ed.): *Translating and the Computer*, North-Holland (1979).
- 4) 野村浩郷, 田中穂積: 機械翻訳, bit 別冊, 共立出版社 (1988. 9).
- 5) 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).
- 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) 劉 涌泉: 機器翻譯淺說, 中国語文 1958, 12月号, pp. 575-577 (1958).
- 8) Liu Zhuo: An Introduction to JFY-II English-Chinese Machine Translation Algorithm, *ZHONG-GUOYUWEN* (中国語文), pp. 216-220, May, 1981; pp. 279-285, July, 1981 (1981).
- 9) 石田 俊, 内田裕士: 多言語間機械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 (Jan. 1982).
- 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) 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, Vol. NL 72-7, No. 40 (1989. 5).
- 12) 李 公緯: 日本語表現の特徴について(その一), 日語学習与研究, pp. 7-15 (1989. 1).
- 13) 李 統漢: 日譯中的中文表達, 日語学習与研究, pp. 31-35 (1989. 4).
- 14) 寺田栄男, 孫 東恢, 田町常夫: 簡易型中日機械翻訳実験システムについて, 情報処理学会研究会報告, NN-75-8 (1990. 1).
- 15) 望月八十吉: 中国語と日本語, 光生館 (1974).
- 16) 今富正巳: 中国語 日本語翻訳の要領, 光生館 (1976).
- 17) 大原信一: 中国語と英語, 光生館 (1973).
- 18) 辻井潤一: 辞書の構成と機械翻訳, 情報処理学会自然言語処理研究会報告, Vol. 126, No. 10 (1985. 10).
- 19) 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. J74-D-II, No. 8, pp. 1060-1069 (1991).
- 20) 陳 国梁: 現代漢語語法教程, 西安交通大学出

- 版社(1986).
- 21) 劉月華ほか: 実用現代漢語語法, 中国外語教学与研究出版社(1986.10).
- 22) 三野昭一: 中国語文法の基礎, 三修社(1987).
- 23) 范莉馨, 任福継, 宮永喜一, 栃内香次: 構文構造の特徴に着目した中日機械翻訳システム, 平成3年電気関係学会北海道支部連合大会, 363, pp. 417-418 (1991.4).
- 24) 范莉馨, 任福継, 宮永喜一, 栃内香次: 中日機械翻訳における構文構造の特徴に着目した訳文の生成, 電子情報通信学会研究会 NLC 91-30, PRU 91-65, pp. 25-32 (1991.10).
- 25) 范莉馨, 任福継, 宮永喜一, 栃内香次: 構文構造に基づく中日機械翻訳システム, 第44回情報処理学会全国大会論文集(3), pp. 3-137-138 (1992.3).
- 26) 任福継, 范莉馨, 宮永喜一, 栃内香次: 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).
- 27) 任福継, 宮永喜一, 栃内香次: 慣用表現の訳文生成について, 信学全大, D 92 (1990.3).
- 28) 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, NLC 90-1, pp. 9-16 (1990.5).
- 29) 范莉馨, 任福継, 宮永喜一, 栃内香次: 中国語文中の複合語の生成について, 電子情報通信学会言語理解とコミュニケーション研究会, NLC 90-1, AI 90-29 (1990.5).
- 30) 范莉馨: 中国語文解析システムにおける複合語の自動抽出に関する研究, 北海道大学修士論文 (1991.3).
- 31) 范莉馨, 任福継, 宮永喜一, 栃内香次: 中日機械翻訳における中国語複合語の自動合成について, 情報処理学会論文誌, Vol. 33, No. 10, pp. 1103-1113 (1992).
- 32) 香坂順一: 文法からの中国語入門, 光生館 (1991.9).
- 33) 邵桂鳳, 鎌田清一郎, 河口英二, 安在弘幸: 中国語文解析における文型パターンの作成とその利用について, 情報処理学会自然言語処理研究会, 89-4, pp. 25-32 (1992.5).
- 34) 陳肇雄, 高慶獅: 智能化英漢機譯系統 IMT/EC, 中国科学-A, 1989, No. 2, pp. 186-194 (1989).
- 35) 范莉馨, 任福継, 宮永喜一, 栃内香次: 文の分解に基づく中日機械翻訳システム, 信学技報, NLC 93-35, pp. 33-40 (1993.7).

付録 1 中国語複文の一部

No	複文類型	例文	意味
01	並列複文	人数既多, 意見又不一致。	人数が多いし, 意見が一致しない。
02	对比複文	他每天不是看书, 就是睡觉。	彼は毎日日本を読んでいなければ, 寝ている。
03	承接複文	我一收到汇款, 就给休回信。	私は為替で送金を受け取ると, すぐ君に手紙で返事をする。
04	累加複文	我们不但取宝, 而且作科学考察。	私達は貴重な経験を吸収するだけでなく, 科学的に視察する。
05	選択複文	或者打個全勝, 或者打個全敗。	完全に勝つか, 或は完全に負けるかである。
06	因果複文	因为天气不好, 所以我们没去杭州。	天気がよくなかったので, それで私達は杭州へ行かなかった。
07	转折複文	他虽然后悔, 但是没有方法挽救。	彼は後悔しているが, しかし救済する方法がない。
08	条件複文	只有掌握了日语, 才能很好地研究日本文学。	日本語をマスターしておいてこそ日本文学をよく研究することができる。
09	無条件複文	民族性無論好坏, 改变都是不容易的。	民族性はよい悪いを問わず, 改めるのが容易ではない。
10	假定複文	假果我手里有一部电子计算机, 就不会算错了。	もしも私達の手元に電子計算機が一台あったら, 計算ちがいはするはずがない。
11	讓步複文	即使意見通不过, 也不该放棄它。	たとえ意見が通らなくても, それを放棄すべきではない。
12	取舍複文	与其匆匆忙忙交卷, 不如多花点时间修改一下。	そそくさと急いで答案を出すより, 少し余計に時間をかけて手直ししてみ方がよい。
13	目的複文	为了不影響工作, 他近来减少了社会活动。	仕事に影響をしないために, 彼は最近社会活動を減少した。
14	時間複文	自從回国以后, 身体就好起来了。	帰国してから身体が良くなりはじめました。
15	連鎖複文	構成複文の単文越多, 複文の構造越複雜。	複文を構成する単文が多ければ多いほど複文の構造はますます複雑になる。

付録 2 常用の一語性と多語性関連語

No	関連語	意味	多/一	多義	No	関連語	意味	多/一	多義
1	又...，又...	...し，	多	×	55	既...，那...	...であるのだから，...	多	×
2	也...，也...	...し，	可2	×	56	既是...，那...	...であるのだから，...	多	×
3	既...，又...	...し，	多	×	57	既然...，那...	...であるのだから，...	多	×
4	既...，也...	...である上に，...である	多	×	58	如果...，就...	もしも...ならば，...である	可1	×
5	一边...，一边...	...しながら，...する	多	×	59	要是...，就...	もしも...ならば，...である	可1	×
6	边...，边...	...しながら，...する	多	×	60	假如...，就...	もしも...ならば，...である	可1	×
7	一面...，一面...	...しながら，...する	多	×	61	倘若...，就...	もしも...ならば，...である	可1	×
8	一方面...，一方面...	...一方では...一方では	多	×	62	假使...，就...	もしも...ならば，...である	可1	×
9	有时...，有时...	時には...時には	多	×	63	倘若...，就...	もしも...ならば，...である	多	×
10	是...，不是...	であって，...でない	多	×	64	尚使...，就...	もしも...ならば，...である	多	×
11	不是...，是...	でなくて，...である	多	×	65	若是...，就...	もしも...ならば，...である	多	×
12	不是...，而是...	でなくて，...である	多	×	66	倘若...，一定...	もしも...ったら，...しない。	多	×
13	不是...，就是...	でなければ...である	多	×	67	尚使...，一定...	もしも...ったら，...しない。	多	×
14	不...，就...	...でないで，...である	多	×	68	尚或...，一定...	もしも...ったら，...しない。	多	×
15	不...，就是...	...でないで，...である	多	×	69	惘然...，可是...	...であるけれども，しかし...	可2	×
16	...，就...		一	○	70	惘然...，但是...	...であるけれども，しかし...	可2	×
17	越...，越...	...ければ...ほど	多	×	71	惘然...，不遇...	...であるけれども，しかし...	多	×
18	S...，S就...	...と，...する。	一	×	72	惘然...，也...	...であるけれども，しかし...	多	×
19	...，才...		一	○	73	惘然...，却...	...であるけれども，しかし...	可1	×
20	一...，就...	...すると，すぐ	可1	×	74	惘然...，還...	...であるけれども，しかし...	可1	×
21	不但...，而且...	...だけでなく，...	可1	×	75	固然...，可是...	確かに...であるが，しかし...	多	×
22	不单...，而且...	ただ...だけでなく，...	多	×	76	固然...，但是...	確かに...であるが，しかし...	多	×
23	不僅...，而且...	...だけでなく，...	多	×	77	固然...，也...	確かに...であるが，しかし...	多	×
24	不光...，而且...	ただ...だけでなく，...	多	×	78	尽管...，可是...	...であるけれども，しかし...	多	×
25	不只...，而且...	ただ...だけでなく，...	多	×	79	尽管...，但是...	...であるけれども，しかし...	多	×
26	不但...，並且...	...だけでなく，...	可1	×	80	尽管...，還是...	...であるけれども，しかし...	多	×
27	不单...，並且...	ただ...だけでなく，...	多	×	81	尽管...，却...	...であるけれども，しかし...	多	×
28	不僅...，並且...	...だけでなく，...	多	×	82	尽管...，還...	...であるけれども，しかし...	多	×
29	不光...，並且...	ただ...だけでなく，...	多	×	83	尽管...，也...	たとえ...でも	多	×
30	不但...，還...	...だけでなく，...	可1	×	84	縱然...，也...	たとえ...でも	多	×
31	不僅...，還...	ただ...だけでなく，...	多	×	85	任凭...，也...	たとえ...でも	多	×
32	不光...，還...	ただ...だけでなく，...	多	×	86	那怕...，也...	たとえ...でも	多	×
33	不但...，也...	ただ...だけでなく，...	多	×	87	就是...，也...	たとえ...であっても，...	可	×
34	不只...，也...	ただ...だけでなく，...	多	×	88	既使...，也...	たとえ...であっても，...	多	×
35	不僅...，更...	ただ...だけでなく，...	多	×	89	即使...，也...	たとえ...であっても，...	多	×
36	不光...，更...	ただ...だけでなく，...	多	×	90	即或...，也...	たとえ...であっても，...	多	×
37	不但...，更...	ただ...だけでなく，...	多	×	91	...，也...	たとえ...であっても，...	多	×
38	不僅...，也...	ただ...だけでなく，...	多	×	92	就算...，也...	たとえ...であっても，...	多	×
39	不光...，也...	ただ...だけでなく，...	多	×	93	只要...，就...	...しさえすれば	多	×
40	...，蓋...	...だけでなく，...	一	×	94	只要...，才...	...しさえすれば	多	×
41	...，連...也...	...だけでなく，...	多	×	95	只消...，就...	...しさえすれば	多	×
42	因為...，所以...	...だから...	可	○	96	只須...	...でさえあれば	一	×
43	因為...，以致...	...だから...	可1	○	97	只有...，才...	...してこそ...する。	多	×
44	由于...，所以...	...のために，...	可1	×	98	只有...，才能...	...してこそ，...ことができる。	多	×
45	由于...，因此...	...のために，...	可1	×	99	除非...，才...	...してこそ...する。	多	×
46	由于...，以致...	...のために，...	可1	×	100	除非...，不然...	...なら別だが，そうでなければ...	多	×
47	既...，就...	...であるのだから，...	多	×	101	除非...，否則...	...なら別だが，そうでなければ...	多	×
48	既是...，就...	...であるのだから，...	多	×	102	除非...，要不...	...なら別だが，そうでなければ...	多	×
49	既然...，就...	...であるのだから，...	多	×	103	若要...，除非...	...なら...するより他はない	多	×
50	既...，也...	...であるのだから，...	多	×	104	不論...，都...	...を問わず，	多	×
51	既是...，也...	...であるのだから，...	多	×	105	無論...，都...	...を問わず，	多	×
52	既然...，也...	...であるのだから，...	多	×	106	無論...，也...	...を問わず，	多	×
53	任凭...，也...	...に関わりなく，必ず...	多	×	107	寧可...，也不...	...するよりは...する方がよい	多	×
54	任凭...，都...	...に関わりなく，必ず...	多	×	108	寧可...，決不...	...しても，...しない。	多	×

付録 2 の続き

No.	関連語	意味	多/一	多義	No.	関連語	意味	多/一	多義
109	不管..., 也...	...であろうとも, ...	多	×	130	寧可..., 也要...	たとえ...でも, ...	多	×
110	不管..., 都...	...であろうとも, ...	多	×	131	寧可..., 也得...	たとえ...でも, ...	多	×
111	是..., 還是...	..., 或は...	多	×	132	与其..., 不如...	...するくらいなら, むしろ...する	多	×
112	或者..., 或者...	..., 或は...	可1	×	133	与其..., 寧可...	...するくらいなら, むしろ...する	多	×
113	或是..., 或是...	..., 或は...	可1	×	134	与其..., 寧肯...	...するくらいなら, むしろ...する	多	×
114	要..., 要...	...か, それとも...か	可1	×	135	連..., 都...	ただ...だけでなく...	多	×
115	寧願..., 不...	...しようとも...しない。	多	×	136	連..., 也...	...さえも	多	×
116	寧願..., 不願...	...しようとも...しない。	多	×	137	尚且..., 何況...	...である上に...である。	多	×
117	寧肯..., 不...	...しようとも...しない。	多	×	138	...至少	少なくとも	一	×
118	寧肯..., 不願...	...しようとも...しない。	多	×	139	...況且		一	×
119	寧可..., 不願...	...しようとも...しない。	多	×	140	先..., 然後...	...しするのち	可1	×
120	...以便	...できるように	一	×	141	..., 否則...	そうでないと	一	×
121	...用以	...するように用いる	一	×	142	...的時候	...する時	一	×
122	在...的過程	...うちに, ...	一	×	143	剛...	...したばかり	一	×
123	在...之内	...うちに, ...	一	×	144	自从..., 就...	...から...はじめて	可2	×
124	趁...的時候	...うちに, ...	一	×	145	每逢...	になるたびごとに	一	×
125	...好	...しやすいように	一	○	146	為了...	のために	一	×
126	...省得	...しないように	一	×	147	為...	のために	一	×
127	...免得	...しないように	一	×	148	為要...	のために	一	×
128	...以免	...しないように	一	×	149	為是...	のために	一	×
129	...以妨	...するように	一	×	150	為的是...	のために	一	×

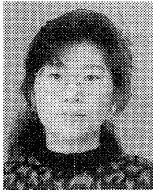
注: 「一」は一語性関連語を表す。26語;
 「多」は多語性関連語を表す。99語;
 「S」は同一主語であることを意味する;
 「可1」は前項を省略される多語性関連語を意味する。19語;
 「可2」は後項を省略される多語性関連語を意味する。4語;
 「可」は両方とも省略される多語性関連語を意味する。2語;
 「○」は多義性のあることを意味する;
 「×」は多義性のないことを意味する。

付録 3 複合語合成ルール

品詞名	No	接続表記	複合語属性	優先順位	合成の例
名詞 n	1	n*/n*/n*	n 9	B	科学文化水平
	2	n*/n*	n 9	B	春夜、青年節
	3	n*/vn	n 9	D	機器翻訳、
動詞 v	4	vn/n*	n 9	D	翻訳文献、
	5	vn/vn	n 9	D	生活経歴、
形容詞 a	6	a1/n*	n 9	B	大雪、軽工業
	7	a2/n*	n 9	B	自然経済、
	8	an/n*	n 9	B	重要科学、
	9	a*/vn	n 9	D	重大発明、
	10	av/n*	n 9	D	対関係、
数詞 m	11	m*/n*	n 9	B	一九四九年
	12	m*/m*	m 9	B	三百六十五、
	13	m*/q*	m q	C	一回、172cm
代名詞 r	14	r*/r*	r 9	B	我們大家
	15	r*/q*	r 9	B	那幅、這首
	16	r*/h2	r 9	B	那一带

(平成 5 年 10 月 12 日受付)

(平成 6 年 9 月 6 日採録)



范 莉馨 (正会員)

1963年生。1984年北京郵電大学無線通信専攻卒業。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工学研究科電子工学専攻修士課程修了。

1994年同大学院博士課程修了。工学博士。CSK 研究員、大連理工大学計算機技術研究所助教授兼務。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、ACSEJ 各会員。



任 福継 (正会員)

1959年生。1982年北京郵電大学電信工部卒業。1985年同大学院計算機応用専攻修士課程修了。同年同大計算機工部講師。1986年中国科学院博士課程。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。CSK 研究員を経て、現在広島市立大学情報科学部助教授。1992年より大連理工大学計算機学部教授兼務。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、言語処理学会、AAMT、ACSEJ 各会員、ACSEJ 会長。



宮永 喜一 (正会員)

1956年生。1981年北海道大学大学院工学研究科電子工学専攻修士課程修了。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。



栢内 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。