実例に基づく帰納的学習による機械翻訳手法における遺伝的アルゴリズムの適用とその有効性

越前谷 博[†,☆] 荒木健治[†]
桃内佳雄[†] 栃内香次[††]

実用的な機械翻訳システムの実現に向け，これまで多くの研究がなされてきた．その主流となっているのが文法解析に基づく解析型の機械翻訳手法である．しかし，この手法は限定された文法規則で多様な言語現象に対処することの困難さから，良質な翻訳を十分に得られるものとはなっていない．これを解決する手法として，近年，実例や用例に基づく学習型の機械翻訳手法が研究されている．我々は，従来より実例から翻訳ルールを帰納的に学習し，翻訳を行う，帰納的学習による機械翻訳手法の提案とその評価実験を行ってきた．しかし，本手法は他の学習型の機械翻訳手法と同様に，翻訳ルールを得るために大量のデータを必要とするという問題点を有する．この問題点を解決するため，我々は以下の3点を目的に帰納的学習による機械翻訳手法への遺伝的アルゴリズムの適用を行った．①少量のデータから多くの翻訳ルールを得る，②最適な翻訳結果を高度に探索する，③正しいデータを与えることで良質な翻訳ルールを得る．この結果，システム全体が最適な翻訳を行うように進化し続けることが期待される．本論文では，遺伝的アルゴリズムの有効性を評価する．実験の結果，正翻訳率は52.8%から61.9%に増加し，新出語の抽出率は47.8%から75.2%に増加した．したがって，遺伝的アルゴリズムの適用が学習能力を向上させ，最適な翻訳結果の生成を可能にすることが確認された．

Application of Genetic Algorithms for Example-Based Machine Translation Method Using Inductive Learning and Its Effectiveness

HIROSHI ECHIZEN-YA,[†] KENJI ARAKI,[†] YOSHIO MOMOUCHI[†] and KOJI TOCHINAI[††]

There have been a lot of researches to realize practical machine translation systems. Among them, rule-based methods are considered to be representative ones. However, it is difficult to deal a wide variety of language phenomena with limited rules. To avoid this difficulty, the example-based approach has been proposed recently. We have proposed an example-based machine translation method, which acquires translation rules from examples using inductive learning, and have evaluated this method. However, this method needs a large amount of examples, like as the other example-based methods. To solve this problem, we applied genetic algorithms to the proposed method. The purposes of the paper are as follows: ① to get translation rules from data, ② to search for better quality translation results, ③ to get high quality translation rules by providing accurate data. In this paper, we evaluate the effectiveness of applying genetic algorithms through the experiments. As a result, the accuracy of the translation rate increased from 52.8% to 61.9%, and the extraction rate of new appearance words increased from 47.8% to 75.2%. It shows that the proposed method has made it possible to improve learning ability and to generate better quality translation results.

1. はじめに

社会の国際化の進展につれ，品質の高い翻訳を高速かつ経済的に行う実用的な機械翻訳システムの開発が強く望まれている．そうしたニーズに応えるべく，これまで多くの機械翻訳手法の研究が行われてきた．しかし，文法解析に基づく解析型の機械翻訳手法[1),2)]においては，多様な言語現象を有限個の文法規則で記述することの困難さや文法規則や辞書の巨大化にともな

† 北海学園大学工学部電子情報工学科
Department of Electronics and Information Engineering, Faculty of Engineering, Hokkai-Gakuen University
☆ 現在，北海道大学大学院工学研究科電子情報工学専攻
Presently with Graduate Course of Electronics and Information Engineering, Faculty of Engineering, Hokkaido University
†† 北海道大学工学部電子情報工学専攻
Division of Electronics and Information Engineering, Faculty of Engineering, Hokkaido University

う，それらの作成・改良の困難さが問題点としてあげられる．この問題点に対して，近年，実例や用例に基づく学習型の機械翻訳手法[3)~6)]がさかんに研究されている．しかし，この翻訳手法においては，翻訳ルールを獲得するために大量のデータが必要となり，良質な翻訳を行うためにはあらかじめ十分な量の実例を学習させなければならない．

我々は，人間の持つ言語獲得および知識獲得の能力を計算機上で実現することを目的とした研究を行っている[7),8)]．こうした立場から，我々は学習型の機械翻訳手法のひとつとして，システム自身が与えられたデータから自動的かつ帰納的に翻訳ルールを学習し，そこで獲得した翻訳ルールを用いて，翻訳を行う帰納的学習による機械翻訳手法の提案[9)]とその性能評価実験[10)]を行ってきた．その結果，それまでに得た翻訳ルールへ影響を及ぼすことなく，適切な翻訳例を追加していくことにより，より最適な翻訳が可能となった．しかし，この手法では最適な翻訳を行うために不可欠となる多様かつ良質な翻訳ルールを得るには，複数の類似した実例を必要とするため大量の学習データを与えなければならない．本論文では，従来の帰納的学習による機械翻訳手法の利点を損なうことなく，この問題点を解決する手法として，遺伝的アルゴリズムを適用した帰納的学習による機械翻訳手法を提案する．遺伝的アルゴリズムを適用することで，与えられた少量の実例を最大限に活用した翻訳ルール獲得能力の向上とより最適な翻訳結果の生成を図る．

遺伝的アルゴリズム[11)~13)]はバラエティに富んだ個体を作り出しながら環境に適応するための世代交代を繰り返すことで進化していくという，生物の進化の過程を模倣したものであり，最適化問題や探索問題に対して最適解を速やかに得るための手法である．本論文では，実例を含む翻訳ルールを個体として位置付け，交叉や突然変異を行うことで多様な翻訳ルールを生成する．そして，良質なデータを与えることで翻訳ルールに対する適応度の決定と淘汰を行う．このような処理の繰返しが，システム全体における世代交代となり，最適な翻訳を行うための進化を続けることになる．さらに，翻訳結果生成の際にも，抽出された翻訳ルールに対し遺伝的アルゴリズムの基本操作を適用することで，効率良く翻訳ルールを活用し最適な翻訳結果を導き出す[14),15)]．

このように帰納的学習による機械翻訳手法へ遺伝的アルゴリズムを適用することにより，システム自身が，与えられたデータを十分に活用した，より多様かつ良質な翻訳ルール獲得のための帰納的な学習を繰り返し，翻訳ルールを効率良く使用する最適な翻訳の実現に向けて成長し続ける．そして，より良い学習型の機械翻訳システムを構築していく．このような生成検査的な手法として，遺伝的アルゴリズムを用いた研究がいくつか行われている[16)~18)]．これらの研究は，実際の学習システムに十分応用されてはおらず，また，遺伝的アルゴリズムの適用においては，システムを制御するための最適なパラメータの検索やデータ構造としてどのような表現形式が最適であるかという点に観点がある[16),17)]．また，遺伝的オペレータを用いて新たな個体を生成する際に，遺伝すべき有用なものとは何であるかといった形質遺伝を重視したものもある[18)]．これに対して，本手法は，最適なパラメータやデータ構造，または，形質遺伝を重要視したものではなく，遺伝的アルゴリズムを実際の学習型の機械翻訳システムに対して適用する際の手法および問題点を明らかにしたという点で大きな意義を持つものである．本論文では，大量のデータを用いて行った評価実験とその考察結果に基づき，帰納的学習による機械翻訳手法における遺伝的アルゴリズムの有効性を確認する．

2. 処理過程

2.1 概　　要

本手法に基づくシステムの処理過程を図1に示す．図1における（　）内は，それぞれの処理部での遺伝的アルゴリズムの適用範囲を示している．本論文では，帰納的学習による機械翻訳手法における遺伝的アルゴリズムの有効性を明確にするために，具体的な学習型英日機械翻訳システムを構築して実験を行った．まず，原文として英文を入力する．その入力された英文に対し，翻訳部において，それまでに抽出された翻訳ルールへ遺伝的アルゴリズムの基本操作を適用することにより，最適な翻訳結果を生成する．翻訳部における遺伝的アルゴリズムの適用は，遺伝的アルゴリズムの代

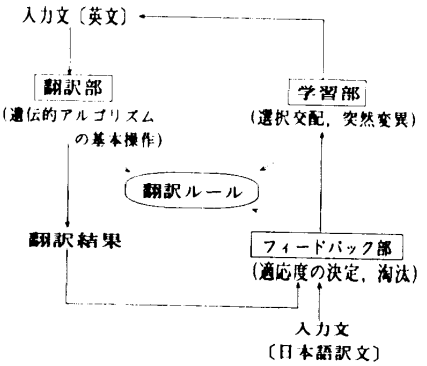

図1　処理過程
Fig. 1　Process.

表的なオペレータを抜粋して用いているのではなく，一連の処理手順を適用している．したがって，本システムは遺伝的アルゴリズムの枠組みをシステム全体と翻訳部の両方に取り入れたものとなっている．次いで，フィードバック部において，ユーザが与えた正しい日本語訳文を用い，翻訳部で使用された翻訳ルールに対する適応度の決定と淘汰を行う．そして，学習部において，与えられた英文とその日本語訳文からなる翻訳例に対し，選択交配と突然変異を行うことで多様な翻訳ルールを生成し，以後の翻訳に活用する．

このような処理を繰り返すことで，システム自身がより多様な翻訳ルールを生成するようになる．そして，良質な訳文を与えることで，環境に適応した良質な翻訳ルールを獲得し，より良い翻訳を行うようにシステムが進化する．

2.2 染色体と遺伝子

本節では遺伝的アルゴリズムを適用するにあたり，個体の染色体と遺伝子をどのように位置付けているかについて述べる．**図2**に示すように染色体には英文とその日本語訳文を組とした翻訳例を，そして，染色体を構成している遺伝子には翻訳例を構成している単語を位置付けている．したがって，個々の個体は可変長の染色体を持つことになるが，これは進化の過程で高等生物になるにつれ，より長く複雑な染色体を持つということを模倣している．また，翻訳例は翻訳ルールを抽出した後，その翻訳例自身も翻訳ルールとして登録される．これを原文の翻訳ルールと呼ぶ．

2.3 フィードバック部

フィードバック部では，まず，ユーザが与えた正しい日本語訳文に基づき，翻訳部で求められた翻訳結果に対する正誤の判定を行う．生成された翻訳結果が与えられた正しい日本語訳文と一致する場合，使用された翻訳ルールの正翻訳度数を1増加させる．一致しない場合には，誤翻訳度数を1増加させる．次いで，その値より，翻訳部で使用されたすべての翻訳ルールに対し適応度を決定する．適応度は以下の式(1)により得られる．

$$\text{適応度}\ (\%) = \frac{\text{正翻訳度数}}{\text{全翻訳度数}} \times 100 \qquad (1)$$

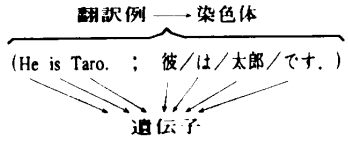

図2 染色体と遺伝子
Fig. 2 A chromosome and genes.

適応度は，各翻訳ルールにおける，それが翻訳に使用された際の正翻訳率を表している．そして，決定された適応度により淘汰を行う．淘汰を行う場合には，ある程度の翻訳ルールに対する試用期間を必要とすることが，予備実験より明らかとなった[19]．これは，学習が十分に行われていない段階では，日本語訳文において，「あなた」と「君」や「～です」と「～だ」等の表現の違いを認識できないため，使用された翻訳ルールに対し安易に誤りと判断してしまう場合が生じるためである．このような表現の違いから誤りと判断された翻訳ルールは，学習が十分に行われていない段階では低い適応度を持つことになる．しかし，学習の進行にともない，徐々に適応度が上昇し，淘汰の対象から外れることになる．そして，学習が十分に行われた段階で，複数の適応度の高い翻訳ルールを基に表現の違いを認識する[20]．したがって，このような試用期間を考慮し，淘汰の対象となる翻訳ルールは，全翻訳度数が5以上で，適応度が25%以下のものとした．

2.4 学 習 部

学習部では，翻訳例に対し選択交配と突然変異，そして，我々が先に提案した字面情報における共通部分と差異部分を多段階に抽出する手法[9]を用いることで多様な翻訳ルールの抽出を行う．本論文では，字面情報のみを用いることで，例外的な文章を処理することが困難であるという解析型の機械翻訳手法における問題点を克服できると考える．学習部における翻訳ルール抽出の処理過程を**図3**に示す．

選択交配は，それまでに入力された翻訳例から英文とその日本語訳文のそれぞれにおいて，共通部分を持つ2つの翻訳例を選択し，共通部分を交叉位置とする一点交叉を行う．共通部分は，英文とその日本語訳文ともに単語を最小単位として字面レベルで完全に一致する文字列である．**図4**にその具体例を示す．

まず，既存の翻訳例から共通部分を持つものを選択

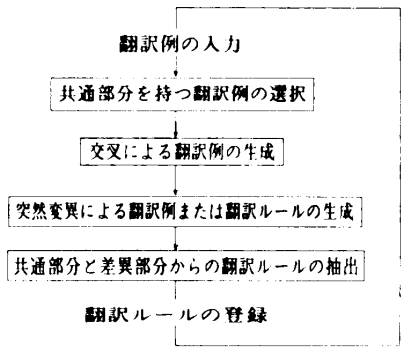

図3 翻訳ルール抽出の処理過程
Fig. 3 Extraction process of translation rules.

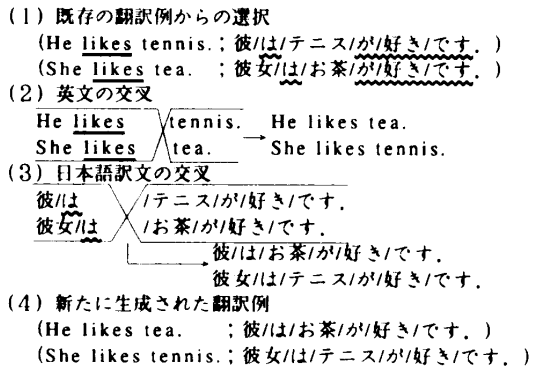

図4 交叉の例

Fig. 4 An example of crossover.

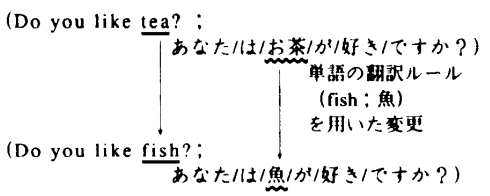

図5 突然変異の例

Fig. 5 An example of mutation.

する．図4に示す例では，英文においては「likes」，日本語訳文においては「は」と「が/好き/です」が共通部分として抽出される．したがって，英文においては「likes」，日本語訳文においては「は」を交叉位置とする一点交叉をそれぞれ行う．また，「が/好き/です」を交叉位置とする日本語訳文の交叉は，新たな日本語訳文を生成しないので翻訳ルールとしての登録は行わない．このように本手法では，交叉位置の決定を字面情報のみに基づき行っている．交叉位置を決定する際に，単語の意味的な対応関係をシステムに与えることは，結果的に解析型の機械翻訳手法の問題点を抱え込むことになり，本手法の利点を損なうものと考えられる．したがって，本手法では字面上での共通部分により交叉位置の決定を行う．

突然変異は，乱数を用い，突然変異率を2%として翻訳例の単語に対するランダムな変更を行う．遺伝子つまり翻訳例の単語をそれまでに抽出された単語の翻訳ルールや変数に置き換えることにより，新たな翻訳例や翻訳ルールを生成する．**図5**に翻訳例の単語が既存の単語の翻訳ルールに変更される突然変異の具体例を示す．また，突然変異は翻訳例に対するランダムな変更であるため，誤ったものを生成する場合がある．そのような翻訳例もしくは翻訳ルールはフィードバック部において誤りと判断され，適応度が低下していく．その結果，淘汰の対象となり消滅していく．

このように遺伝的アルゴリズムの基本操作を適用す

図6 共通部分と差異部分の抽出例

Fig. 6 An example of the extraction of common part and different parts.

図7 抽出された翻訳ルールの例

Fig. 7 Examples of the extracted translation rules.

ることにより新たに生成された翻訳例と既存の翻訳例から，共通部分と差異部分を多段階に抽出する．**図6**にその具体例を示す．

共通部分と差異部分を多段階に抽出することで得られる翻訳ルールは，類似した翻訳例の組を必要とするため，その出現頻度の低さが従来より大きな問題点となっていた．しかし，遺伝的アルゴリズムの基本操作を適用することで，字面情報を最大限に活用した多様な翻訳例の生成が可能となり，この問題を解決する有効な手段になると考えられる．図4に示した2つの既存の翻訳例から抽出される最終的な翻訳ルールを**図7**に示す．

2.5 翻 訳 部

翻訳部では，それまでに抽出された翻訳ルールを用いて入力文に対する最適な翻訳結果を導き出す．そのために，翻訳部のみで1つの完結した遺伝的アルゴリズムのシステムを構築し，高度な探索を実現する．**図8**に翻訳部における処理過程を示す．

以下にそれぞれの処理の詳細を述べる．

(1) 初期集団の発生

それまでに抽出された翻訳ルールから，入力文の基礎的な構造を表現している文の翻訳ルール

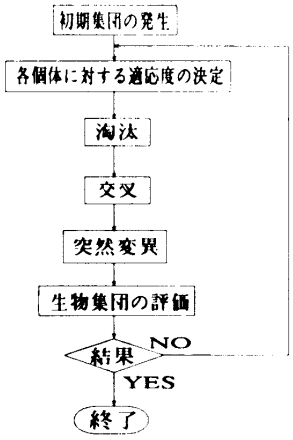

図8 翻訳部の処理過程
Fig. 8 Process of the part of translation.

(1) 原文
I am your teammate.
(2) 初期集団の発生
{ (I am @0.；私/は/@0/です.),
(@0 am your @1.；@0/は/あなたの/@1/です.),
(@0 am @1.；@0/は/@1/です.), ・・・}
(3) 交叉により生成された翻訳ルール
(I am your @1.；私/は/あなたの/@1/です.)
(4) 生物集団の評価
単語の翻訳ルール (teammate；チーム仲間)
代入
(I am your @1.；私/は/あなたの/@1/です.)
(I am your teammate.；私/は/あなたの/チーム仲間/です.)
(5) 翻訳結果
私はあなたのチーム仲間です.

図9 翻訳部における翻訳結果生成の例
Fig. 9 An example of the translation result processing.

を選択し，その集団を初期集団とする．入力文の基礎的な構造を表現している文の翻訳ルールとは，入力文に対し，文の翻訳ルールの変数以外のすべての単語が，入力文の単語の並びと同じ順序で含まれているものである．

(2) 各個体に対する適応度の決定
フィードバック部で使用した式 (1) を用いて各翻訳ルールに対する適応度を決定する．

(3) 淘汰
フィードバック部と同様に，全翻訳度数が5以上，適応度が25%以下の翻訳ルールに対して淘汰を行う．

(4) 交叉
学習部で用いた手法と同様に，共通部分を持つ2つの翻訳ルールを選択し，その共通部分を交叉位置とする一点交叉を行う．

(5) 突然変異
学習部と同様に，乱数を用い，突然変異率を2%として翻訳ルールに対するランダムな変更を行う．

(6) 生物集団の評価
文の翻訳ルールの変数部分に，既存の単語の翻訳ルールを代入し，入力文と一致するものが存在するかどうかを調べる．一致するものが存在した場合，入力文に対する翻訳結果が得られたとして翻訳部における処理を終了する．また，入力文と完全に一致しない，変数を含んだものが得られた場合，(2)～(6) を繰り返す．その結果，生成された英文に変化がなければ，その英文に対する日本語訳文を最適な翻訳結果として処理を終了する．

図9に翻訳部における処理過程の具体例を示す．

また，翻訳部では，1つの入力文に対し，複数の翻訳結果が生成される場合がある．その場合には，それらの翻訳結果に対し，以下に示す3つの方法を用いて優先順位を決定する．

①使用した翻訳ルールの適応度が最大のもの
②使用した翻訳ルールの抽象度が最小のもの
③使用した翻訳ルールの登録順位が最上位のもの

適応度はフィードバック部で使用した式 (1) より求めたものと同様である．また，抽象度は翻訳結果を生成する際に，基本となった翻訳ルールの単語数における変数の個数の占める割合である．抽象度は以下の式 (2) により得られる．

$$\text{抽象度}\,(\%) = \frac{\text{翻訳ルール中の変数の個数}}{\text{翻訳ルール中の単語数}} \times 100 \qquad (2)$$

3. 性能評価実験

3.1 評価方法

本手法において生成された翻訳結果に対する評価方法について述べる．有効な翻訳は以下の2つである．

①未登録語を含まない正翻訳
②未登録語を含む正翻訳

未登録語を含む正翻訳は，未登録語に名詞句や形容詞などの単語の翻訳ルールを与えることで容易に未登録語を含まない正翻訳を導き出せるものである．したがって，有効な翻訳率は以下の式 (3) により決定される．

$$\text{有効な翻訳率}\,(\%) = \frac{①+②}{\text{全翻訳数}} \times 100 \qquad (3)$$

無効な翻訳は以下の3つである．

③未登録語を含まない誤翻訳
④未登録語を含む誤翻訳

・有効な翻訳
①未登録語を含まない正翻訳
入力文 ：He is my teammate.
翻訳結果：彼は私のチーム仲間です.
彼は僕のチーム仲間だ.
②未登録語を含む正翻訳
入力文 ：You play baseball.
翻訳結果：あなたは@0をします.
@0は野球をします.
・無効な翻訳
③未登録語を含まない誤翻訳
入力文 ：That is my book.
翻訳結果：あれは私の本ではない.
④未登録語を含む誤翻訳
入力文 ：I don't like tea.
翻訳結果：私は@0が好きです.
⑤翻訳不能
入力文 ：Look at this picture.
翻訳結果：翻訳不能

図10 翻訳結果の評価方法の例
Fig. 10 An evaluation method of the translation results.

⑤入力文に対する基礎的な構造を表現している文の翻訳ルールが1つも存在せずにまったく翻訳が行えなかった翻訳不能

未登録語を含む誤翻訳は，名詞句や形容詞など以外の単語の翻訳ルールを必要とするもの，または，未登録語に名詞句や形容詞などの単語の翻訳ルールを与えても正翻訳を導き出せないものである．したがって，無効な翻訳率は，以下の式(4)により決定される．

$$\text{無効な翻訳率}\,(\%) = \frac{③+④+⑤}{\text{全翻訳数}} \times 100 \quad (4)$$

図10にそれぞれの具体例を示す．

また，生成された翻訳結果が複数存在する場合には，翻訳部で述べた方法により順位付けされた翻訳結果の上位1番から10番までを評価の対象とする．その結果，10個の翻訳結果中に有効な翻訳に該当するものが存在していた場合，その翻訳結果を有効な翻訳とする．

3.2 実験方法

まず，翻訳ルールを学習するために中学1年生用教科書ガイド・ワンワールド[21]に掲載されている英文とその訳文の1,010組（総文字数11,479文字）を学習データに用い，辞書を空の状態にして翻訳と学習を1文ずつ繰り返し行った．その結果，抽出された翻訳ルールを逐次，初期辞書に登録した．

次いで，翻訳評価データとして中学1年生用教科書ガイド・ニューホライズン[22]に掲載されている英文800文を用いて翻訳を行った．翻訳結果が生成された後，初期辞書作成時と同様に，1文ずつ正しい日本語訳文を与え学習を行った．

また，あらかじめ行った予備実験[14]より，入力されたすべての翻訳例に対する遺伝的アルゴリズムの適用は，翻訳ルールの抽出に膨大な処理時間を必要とすることが明らかとなった．これは実用的な機械翻訳システムを構築するうえで大きな弊害となる．そこで，我々は処理時間の短縮を目的としたヒューリスティックスの導入を行った．このヒューリスティックスは，人間が翻訳例を利用する場合，それまでに与えられたすべての翻訳例を同レベルで活用しているのではなく，最近学習したものを重点的に活用しているのではないかという直観に基づいている．本手法では，このヒューリスティックスを導入し，処理時間の短縮のため学習に使用する翻訳例を最近学習したものに制限した．

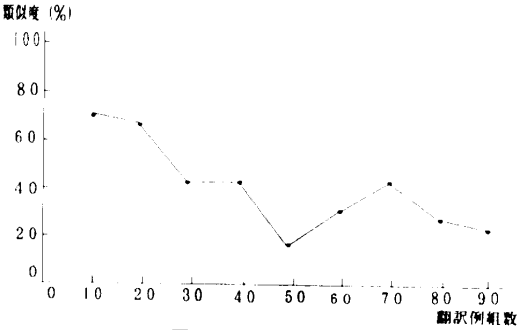

図11 類似度の推移
Fig. 11 Change of similarity degree.

本実験では，ヒューリスティックスを導入する際，学習に使用する翻訳例組数の決定を翻訳例の類似性に基づき行っている．また，本手法は，遺伝的アルゴリズムの適用を字面情報より行っているため，字面レベルでまったく類似性のない翻訳例の集団からは大きな効果は期待できない．そこで，本実験データにおいて，遺伝的アルゴリズムの適用が十分に活用される翻訳例の集団について調査を行った．**図11**は，中学1年生用教科書ガイド・ワンワールド[21]に掲載されている最初の10組の翻訳例を基準として，その後の10組ごとの翻訳例との類似度の推移を表したものである．類似度は以下の式(5)により決定される．

$$\text{類似度}\,(\%) = \frac{\text{共通単語数}}{\text{全出現単語数}} \times 100 \quad (5)$$

図11より，20組を境に類似度が大きく低下していることが確認できる．したがって，最適な翻訳例の集団を，基準として用いた10組を含む30組であると決定した．このヒューリスティックスの導入前と導入後のそれぞれに対し，中学1年生用教科書ガイド・ワンワールド[21]に掲載されている最初の100文を用いて予備実験を行った．その結果，ヒューリスティックスの導入による翻訳の精度と質への影響がないことを確認できた[15]．したがって，本実験では，学習に使用する翻訳例を最近学習した30組に制限した．このようなヒューリスティックスの導入が翻訳の精度および質

を低下させずに処理時間の短縮をもたらすものと考えられる．

3.3 実験結果

実験の結果，本手法の有効な翻訳率は61.9%となった．**表1**に遺伝的アルゴリズムを適用しなかった場合，**表2**に遺伝的アルゴリズムを適用した場合の実験結果を示す．その結果，遺伝的アルゴリズムを適用しなかった場合と比較し，有効な翻訳率は52.8%から61.9%に増加した．

また，本実験の結果における遺伝的アルゴリズムを適用しなかった場合と適用した場合の有効な翻訳率の推移を**図12**に示す．破線で表したものが遺伝的アルゴリズムを適用しなかった場合，実線で表したものが遺伝的アルゴリズムを適用した場合の有効な翻訳率の推移である．

表1 遺伝的アルゴリズムを適用しなかった場合の実験結果
Table 1 The result of the experiment without genetic algorithms.

		翻訳率	合計
有効な翻訳	正翻訳	40.3%	52.8%
	未登録	12.5%	
無効な翻訳	誤翻訳	8.9%	47.2%
	未登録	27.5%	
	翻訳不能	10.8%	

表2 遺伝的アルゴリズムを適用した場合の実験結果
Table 2 The result of the experiment using genetic algorithms.

		翻訳率	合計
有効な翻訳	正翻訳	47.0%	61.9%
	未登録	14.9%	
無効な翻訳	誤翻訳	14.7%	38.1%
	未登録	17.1%	
	翻訳不能	6.3%	

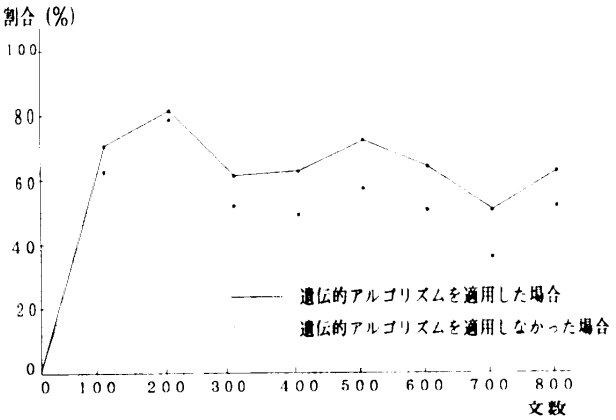

図12 有効な翻訳率の推移
Fig. 12 Change of the effective translation rates.

3.4 考察

3.4.1 本手法の有効性

本論文で提案する手法による実験の結果，遺伝的アルゴリズムの適用が，実例からの多様かつ良質な翻訳ルールの学習能力を向上させ，より最適な翻訳結果を導き出す学習型の機械翻訳システムの実現を可能にすることを確認できた．以下に本手法の有効性について述べる．

(1) 翻訳結果について

表1と表2に示すように，有効な翻訳率が52.8%から61.9%に増加した．これは，中学1年生用教科書ガイド・ニューホライズン[22]800文の翻訳において，無効な翻訳から有効な翻訳へ移行したものが95文，有効な翻訳から無効な翻訳へ移行したものが22文，結果として，有効な翻訳が73文増加したことに相当する．**表3**に無効な翻訳から有効な翻訳へ移行した95文の内訳とそれぞれの具体例を示す．**表4**には，有効な翻訳から無効な翻訳へ移行した22文の内訳とそれぞれの具体例を示す．

また，遺伝的アルゴリズムを適用することにより，校正の必要がない翻訳結果が増加し，翻訳の質においても向上していることが確認できた．具体的に，遺伝的アルゴリズムを適用しなかった場合と適用した場合のどちらにおいても有効な翻訳であった400文中で，翻訳の質が向上したものは63文に相当する15.8%であった．**表5**に，遺伝的アルゴリズムの適用により良質な翻訳となった63文の内訳とその具体例を示す．

(2) 翻訳ルールについて

こうしたより最適な翻訳結果の生成は，遺伝的アルゴリズムを適用することにより，多様かつ良質な翻訳ルールの抽出とそれらを効率良く使用することが可能になったためである．本論文で述べる良質な翻訳ルールとは，翻訳結果を生成する際に，他の翻訳ルールとの組合せを必要としない，より具象的な翻訳ルールである．本手法では，極論すれば，良質な翻訳を行うためには，あらかじめ人間が行った入力文とまったく同じ翻訳例を用意し，それらを翻訳ルールとして辞書に登録しておけばよいということになる．しかし，そのような翻訳例をすべて登録することは，現実的には不可能である．そこで，限られた翻訳例から良質な翻訳を行うために有効となる翻訳ルールを獲得することが要求される．こうした観点より，良質な翻訳ルールを他の翻訳ルールとの組合せを必要としない，より入力文と類似した翻訳ルールとして位置付けることで，翻訳結果の生成において誤った翻訳ルールとの組合せ処理が減少し，容易に良質な翻訳結果を導き出せるもの

表3 遺伝的アルゴリズムの適用の有効性
Table 3 The effectiveness of the application of genetic algorithms.

	文例	遺伝的アルゴリズムを適用しない場合の翻訳結果	遺伝的アルゴリズムを適用した場合の翻訳結果	文数(計) 95
誤翻訳（未登録無）→ 正翻訳（未登録無）	He is three years old.	彼は3枚の歳です.	彼は3歳です.	31
誤翻訳（未登録無）→ 正翻訳（未登録有）	He's helping with her homework.	彼は彼女の宿題をします.	彼は@0の宿題を手伝っています.	1
誤翻訳（未登録有）→ 正翻訳（未登録無）	Mike, is this your brother?	@0はこれはあなたの兄ですか?	マイク，こちらはきみの兄ですか?	20
誤翻訳（未登録有）→ 正翻訳（未登録有）	We are playing baseball.	@0は野球をしていますです.	@0は野球をしています.	33
翻訳不能 → 正翻訳（未登録無）	Hi, Ken.	翻訳不能	こんにちは，健.	3
翻訳不能 → 正翻訳（未登録有）	Yumi speaks English very well.	翻訳不能	@0はとても上手に英語を話す.	7

表4 遺伝的アルゴリズムの適用による誤り
Table 4 The mistakes by applying of genetic algorithms.

	文例	遺伝的アルゴリズムを適用しない場合の翻訳結果	遺伝的アルゴリズムを適用した場合の翻訳結果	文数(計) 22
正翻訳（未登録有）→ 誤翻訳（未登録有）	Mike is an American boy.	@0はアメリカの少年です.	@0はアメリカの誰です.	6
正翻訳（未登録無）→ 誤翻訳（未登録無）	Jane's sister speaks English.	ジェーンの姉は日本語を話します.	ジェーンの姉ではありませんは日本語を話します.	10
正翻訳（未登録有）→ 誤翻訳（未登録無）	You have a nice computer.	あなたは@0コンピュータを持っています.	あなたはコンピュータを持っていません.	6

表5 遺伝的アルゴリズムの適用による良質な翻訳
Table 5 The high quality translations by applying genetic algorithms.

	文例	遺伝的アルゴリズムを適用しない場合の翻訳結果	遺伝的アルゴリズムを適用した場合の翻訳結果	文数(計) 63
正翻訳（未登録有）→ 正翻訳（未登録無）	Do you like sports?	あなたは@0が好きですか?	あなたはスポーツが好きですか?	11
正翻訳（未登録無）	I have two.	私は2持っています.	私は2つ持っています.	32
正翻訳（未登録有）	Yumi, this is Lucy.	由美，これは@0です.	由美，こちらは@0です.	20

と考えられる.

次に，具体的にどのような良質な翻訳ルールの抽出が可能となり，有効な翻訳を増加させたのかについて述べる．遺伝的アルゴリズムの適用により増加した良質な翻訳ルールは以下の3つに分類することができる.

①入力文と完全に一致する翻訳例，つまり原文の翻訳ルール

②従来では得られなかった良質な文の翻訳ルール

③従来では正確に切り出すことが困難であった単語の翻訳ルール

このような遺伝的アルゴリズムの適用により作り出された多くの良質な翻訳ルールを使用することで，より最適な翻訳結果の生成が可能となった．無効な翻訳から有効な翻訳となった95文以外の有効な翻訳においても，より良質な翻訳ルールを使用していた．表3における95文の翻訳結果が，これら3つの分類のいずれに該当しているのかについて，その内訳とそれぞれの具体例を**表6**に示す.

従来の手法では，学習部において与えられた翻訳例のみから字面情報を用いて共通部分と差異部分を抽出し，翻訳ルールの獲得を試みた場合，英文とその日本語訳文のそれぞれの差異部分が複数存在しているため明確に対応付けできず，良質な翻訳ルールの抽出が行えなかった．しかし，遺伝的アルゴリズムの基本操作を適用することで，与えられた翻訳例と類似した多くの翻訳例が自動的に作り出され，英文とその日本語訳文の差異部分が明確に対応付けられた翻訳例の出現頻度が向上した．そのため多くの良質な翻訳ルールの抽

表6　遺伝的アルゴリズムの適用による有効な翻訳
Table 6　The effective translations by applying of genetic algorithms.

原文の翻訳ルール：2文	遺伝的アルゴリズムを適用しなかった場合の翻訳過程	遺伝的アルゴリズムを適用した場合の翻訳過程
文例： He is three years old.	翻訳結果：彼は3枚の歳です. (@0 is @1 years old.； @0/は/@1/歳/です.) (He；彼) と (three；3/枚/の) を代入	翻訳結果：彼は3歳です. (He is three yaers old.； 彼/は/3/歳/です.)

原文の翻訳ルール (He is three years old.；彼/は/3/歳/です.) の抽出過程

① 既存の翻訳例
[1-1] (He is two years old.；彼/は/2/歳/です.)
[1-2] (My sister is three years old.；私の/妹/は/3/歳/です.)
② (is；は) を交叉位置とした交叉により生成された翻訳例
[1-3] (He is three years old.；彼/は/3/歳/です.)

文の翻訳ルール：78文	遺伝的アルゴリズムを適用しなかった場合の翻訳過程	遺伝的アルゴリズムを適用した場合の翻訳過程
文例： We are playing baseball.	翻訳結果：@0は野球をしていますです. (@0 are @1.；@0/は/@1/です.) (playing @0；@0/を/してい/ます) (baseball；野球)	翻訳結果：@0は野球をしています. (@0 are playing baseball.； @0/は/野球/を/してい/ます.)

文の翻訳ルール (@0 are playing baseball.；@0/は/野球/を/してい/ます.) の抽出過程

① 既存の翻訳例
[1-1] (Makoto and Akira are playing baseball.；真/と/明/は/野球/を/してい/ます.)
[1-2] (You are helping.；あなた/は/助けてい/るところだ.)
② (are；は) を交叉位置とする交叉により生成された翻訳例
[1-3] (You are playing baseball.；あなた/は/野球/を/してい/ます.)
③ 抽出された文の翻訳ルール
[1-1] と [1-3] より (@0 are playing baseball.；@0/は/野球/を/してい/ます.)

単語の翻訳ルール：15文	遺伝的アルゴリズムを適用しなかった場合の翻訳過程	遺伝的アルゴリズムを適用した場合の翻訳過程
文例： Whose cat is this?	翻訳結果：これは誰の猫を飼っているですか? (Whose @0 is this?； これ/は/誰の/@0/ですか?) (cat；猫/を/飼ってい/る) を代入	翻訳結果：これは誰の猫ですか? (Whose @0 is this?； これ/は/誰の/@0/ですか? (cat；猫) を代入

単語の翻訳ルール (cat；猫) の抽出過程

① 既存の翻訳例
[1-1] (My brother is two years old.；私の/弟/は/2/歳/です.)
[1-2] (My cat is three years old.；私の/猫/は/3/歳/です.)
② (is；は) を交叉位置とする交叉により生成された翻訳例
[1-3] (My cat is two years old.；私の/猫/は/2/歳/です.)
③ [1-1] と [1-3] より抽出された単語の翻訳ルール (cat；猫)

出が可能となった.

次に，遺伝的アルゴリズムを適用することにより，翻訳ルールがどのように変化したかについて述べる. **表7**に，遺伝的アルゴリズムの適用により，無効な翻訳から有効な翻訳へと移行した翻訳結果に対する初期集団で発生した翻訳ルール数を変数の数別に表したものを示す. **図13**には，その具体例を示す. 図13に示す※印のついた翻訳ルールは，遺伝的アルゴリズムを適用することにより得られた，入力文の構造をより正確に表現している翻訳ルールである. 表7より，遺伝的アルゴリズムを適用しなかった場合に比べ，翻訳ルール数が約2倍に増加している. この中で，誤った翻訳ルールは，遺伝的アルゴリズムを適用しなかった場合では約35%，適用した場合では約42%であったことから，誤った翻訳ルールの数は増加したことになる. したがって，有効な翻訳率が増加したのは，図13の具体例が示すように，初期集団において発生した適用可能な翻訳ルールの数が増加したためというより，入力文の構造をより正確に表現している翻訳ルールが出現するようになったためと考えられる.

さらに，**図14**には，遺伝的アルゴリズムを適用した場合における学習量に対する翻訳ルール数の変化を示す. 図14より，1,000文を境にして傾きが緩やかになっていることが確認できる. これは，最初の1,010

表7 遺伝的アルゴリズムの適用により有効な翻訳となった翻訳結果に対する初期集団の翻訳ルール数の内訳
Table 7 Details of the number of the translation rules in the initial groups for the translation results which genetic algorithms changed into the effective translation results.

	遺伝的アルゴリズムを適用しなかった場合	遺伝的アルゴリズムを適用した場合
変数なし	1	8
変数1つ	353	588
変数2つ	461	1,144
変数3つ	66	28
合計	881	1,768

例1：入力文
(My name is Kazuo.；僕の/名前/は/和夫/です．)
遺伝的アルゴリズムを適用しなかった場合
(My @0 is @1.；私の/@0/は/@1．)
(@0 is @1.；@0/は/@1/です．)
(@0 is @1.；@0/は/@1．)
遺伝的アルゴリズムを適用した場合
※ (My name is @0.；僕の/名前/は/@0/です．)
(My name is @0.；私の/名前/は/@0/です．)
(@0 name is @1.；@0/名前/は/@1/といいます．)
(@0 name is @1.；@0/名前/は/@1/です．)
(My @0 is @1.；私の/@0/は/@1/です．)
(My @0 is @1.；@0/は/@1/です．)

例2：入力文
(We are playing baseball.
；私達/は/野球/を/して/います．)
遺伝的アルゴリズムを適用しなかった場合
(@0 are @1.；@0/は/@1/です．)
(@0 are @1.；@0/は/@1．)
(@0 are @1.；@0/は/@1/ます．)
遺伝的アルゴリズムを適用した場合
※ (@0 are playing baseball.
；@0/は/野球/を/して/います．)
(@0 are playing @1.；@0/は/@1/を/して/います．)
(@0 are @1.；@0/は/@1/です．)
(@0 @1 baseball.；@0/は/野球/を/@1．)
(@0 playing baseball.；@0/は/野球/を/して/います．)
(@0 playing @1.；@0/は/@1/を/して/います．)

※：遺伝的アルゴリズムの適用により新たに得られた翻訳ルール

図13 初期集団の例
Fig. 13 Examples for the initial groups.

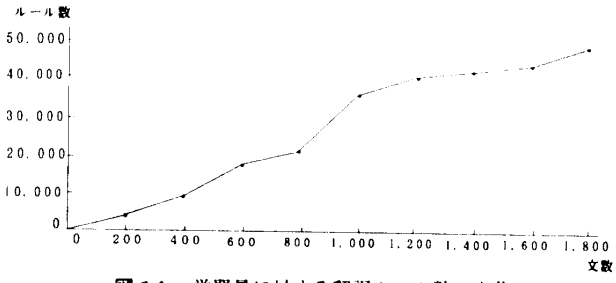

図14 学習量に対する翻訳ルール数の変化
Fig. 14 Change of the number of the translation rules for the learning quantity.

文が学習データとして使用した中学1年生用教科書ガイド・ワンワールド[21]に掲載されている文章であり，その後の800文が翻訳評価データとして用いた中学1年生用教科書ガイド・ニューホライズン[22]に掲載されている文章であったことから，最初の1,000文に新出語が集中し，翻訳ルールが大幅に増加したと考えられる．また，表8には，学習した翻訳ルールを変数の数別に表したものを示す．表8に示すように，遺伝的アルゴリズムを適用することにより，翻訳ルール数の倍率は7.7倍となった．これは，遺伝的アルゴリズムの適用により，多くの翻訳例が生成されたためである．誤った翻訳ルールは，遺伝的アルゴリズムを適用しなかった場合では約42%，適用した場合では約46%であった．この結果より，遺伝的アルゴリズムを適用することで，正しい翻訳ルール数が増加したことを確認できた．

表8 翻訳ルールの内訳
Table 8 Details of the translation rules.

	遺伝的アルゴリズムを適用しなかった場合	遺伝的アルゴリズムを適用した場合	倍率
変数なし	4,915	33,806	6.9倍
変数1つ	1,191	12,091	10.2倍
変数2つ	245	2,902	11.8倍
変数3つ	19	341	17.9倍
変数4つ	1	12	12.0倍
合計	6,371	49,152	7.7倍

また，本手法は学習機能に基づいた機械翻訳手法であるため，ユーザや対象分野に適応することが可能である．したがって，本手法によるシステムがユーザや対象分野に動的に適応していくことで，類似した翻訳例の学習が行われ，ルール数の爆発やsparsenessの問題を解決できると考えられる．複文の翻訳においては，理論的には，単文の翻訳ルールを組み合わせることにより，その翻訳は可能になるものと考えられる．しかし，複文に対する実験は今回行っていない．したがって，これらの点については今後研究を進め，別の機会に報告する予定である．

(3) 学習について

図12より，遺伝的アルゴリズムを適用しなかった場合に比べ，全般的に有効な翻訳率が増加していることを確認できた．これは遺伝的アルゴリズムの適用がある特定の文法を持った文章に有効なのではなく，多様な文章に対して有効であることを示している．また，有効な翻訳率の上下の動きは，名詞句や形容詞などの未登録語ではなく，与えられても容易には正翻訳を導き出すことができない初めて与えられた動詞の単語数の増加にほぼ対応している．したがって，学習量を反映しているといえる．図12の300文の位置においては，それまで与えられていない一般動詞の出現の増加が，700文の位置においては，原型としてはすでに与えられているが，三人称単数により動詞が変化した形で与えられた単語の出現の増加が原因となり翻訳率が低下した．しかし，このような原因はいずれも学習不

足によるものであるため，学習を十分に行うことで翻訳率を上昇させることが可能となる．

次に，学習量に対する翻訳率の伸びについて述べる．図12の有効な翻訳率の推移は，名詞句や形容詞以外の未登録語の出現の影響を大きく受けており，学習が十分行われているとはいえず，学習量に対する翻訳率の伸びを十分に示したものとはなっていない．そこで，十分な学習が行われ，未登録語が非常に少ないと考えられるbe動詞の文章のみを対象に翻訳率を調査した．**図15**にその調査結果を示す．図15より，be動詞においては少量の学習量でかなり高い翻訳率が得られることを確認できた．また，100文以降において翻訳率がさほど急激に向上していないのは，100文以降ではbe動詞および否定語の短縮形などの単語を含む文章が，無効な翻訳の約90%を占めていたために，それらの単語が未登録語となり，翻訳率の向上の妨げになったということである．

さらに，本手法における学習の容易性について述べる．本実験で使用した全翻訳例において，英文とその日本語訳文中の単語が1対1に対応するものは314組であった．そのうち正しく辞書に登録されていたものは，遺伝的アルゴリズムを適用しなかった場合では150組，遺伝的アルゴリズムを適用した場合においては236組であった．したがって，新出語の抽出率は，47.8%から75.2%に増加した．これは，本手法が比較的高い率で新出語を学習できていることを示している．また，抽出された単語の組が何度目の出現時に得られたのかを**表9**に示す．表9より，1度目から3度目の間で抽出された単語の組は，遺伝的アルゴリズムを適用しなかった場合で約73%，遺伝的アルゴリズムを適用した場合では，約87%であった．したがって，遺伝的アルゴリズムの適用により，少ない出現回数で新出語を獲得できていることが確認できた．また，抽出できなかった単語の組においても，それらが何度出現していたのかを**表10**に示す．表10より，抽出できなかった単語の組において，1度もしくは2度しか出現していないものは，遺伝的アルゴリズムを適用しなかった場合で約57%，遺伝的アルゴリズムを適用した場合で約77%であった．この結果より，抽出できなかった単語の組においては，その原因が学習不足によるものであると考えられる．また，それらの単語に対しては少量の学習量で抽出することが可能になると考えられる．

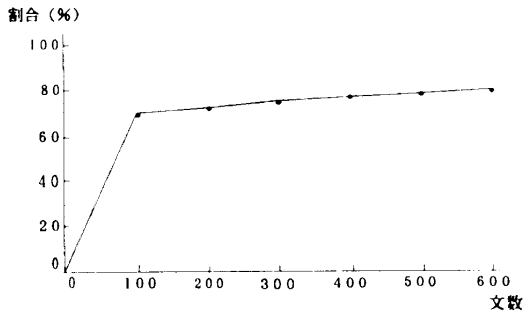

図15 be動詞の翻訳率の伸び
Fig. 15 Change of the translation rate for 'be'.

(4) 翻訳部における遺伝的アルゴリズムの有効性

翻訳部に対して遺伝的アルゴリズムの基本操作を適用することで効率良く翻訳結果を生成することが可能となった．**図16**にその具体例を示す．従来では，個々

表9 抽出された単語の組の出現回数の内訳
Table 9 Details of the appearance frequency for the set of the extracted words.

	遺伝的アルゴリズムを適用しなかった場合	遺伝的アルゴリズムを適用した場合
1度目	63	134
2度目	27	40
3度目	20	32
4度目	15	19
5度目	6	6
6度目以降	19	5
合計	150	236

表10 抽出できなかった単語の組の出現回数の内訳
Table 10 Details of the appearance frequency for the set of the words which could not be extracted.

	遺伝的アルゴリズムを適用しなかった場合	遺伝的アルゴリズムを適用した場合
1度目	59	38
2度目	34	22
3度目	19	7
4度目	13	3
5度目	11	3
6度目以降	28	5
合計	164	78

(1) 原文
This is an album.
(2) 初期集団の発生
{ (This is @0.；これ/は/@0/です．)，
(@0 is an @1.；@0/は/@1/です．) }
(3) 交叉により生成された翻訳ルール
(This is an @1.；これ/は/@1/です．)
(4) 生物集団の評価
単語の翻訳ルール (album；アルバム) を代入
(This is an @1.；これ/は/@1/です．)
(5) 翻訳結果
これはアルバムです．

図16 翻訳部における遺伝的アルゴリズムの有効性
Fig. 16 The effectiveness of genetic algorithms in the part of translation.

表 11 淘汰された翻訳ルールの精度
Table 11 Precision of the translation rules selected.

	占有率（誤翻訳ルール数/淘汰数）
変数なし	97.4% (2,361/2,425)
変数 1 つ	97.4% (1,444/1,483)
変数 2 つ	96.8% (638/659)
変数 3 つ	71.4% (20/28)
合計	97.1% (4,463/4,595)

の文の翻訳ルールの変数部分へ単語の翻訳ルールを単に代入していたため，既存の単語の翻訳ルールとして（album; アルバム）のみが存在していた場合，図 16 に示す初期集団で選択された 2 つの文の翻訳ルールからは，翻訳結果として「@0 はアルバムです.」が得られるだけであった．しかし，遺伝的アルゴリズムの適用により，他の翻訳ルールを効率良く活用し，最適な翻訳結果「これはアルバムです.」の生成が可能となった．

(5) 淘汰処理について

本手法の実験におけるフィードバック部において淘汰された翻訳ルールの精度についての調査を行った．**表 11** に調査結果を示す．その結果，全体で誤った翻訳ルールが約 97%を占めていたことから，淘汰された翻訳ルールに対しては，高い精度で評価されていることが確認できた．

(6) ヒューリスティックスの導入

ヒューリスティックスの導入により，学習部における翻訳ルールの抽出処理が，ごく最近学習した翻訳例のみに制限されたことで，処理時間が大幅に減少した．

3.4.2 他手法との比較

学習型の機械翻訳手法には，実例からシステム自身が規則を抽出し，その規則を用いて翻訳行う手法[3),5)]と，大量の事例を格納して入力文と事例データベース中の文との距離計算を行い，類似した事例を模倣することにより翻訳を行う，アナロジーによる手法[4),6)]がある．これらの手法は，翻訳率および翻訳の質の向上のために，膨大な量の学習データを必要とするという問題点を抱えている．本手法では，この問題点を解決するために遺伝的アルゴリズムを適用した．具体的には，翻訳例に対し，交叉や突然変異を行うことにより新たな翻訳例を生成し，それにともない多くの翻訳ルールを作り出す．また，獲得した翻訳ルールに対し，淘汰を行うことで，良質な翻訳ルールの集団を維持する．さらに，翻訳結果を生成する際にも，遺伝的アルゴリズムの一連の処理手順を適用し，より良質な翻訳結果の生成を行う．今回の実験は，遺伝的アルゴリズムを適用した学習型の機械翻訳手法の第一段階として，その適用の有効性を実証するために行ったものである．

また，実験結果について従来の手法と比較した場合，古瀬らの手法[6)]では，「国際会議に関する問い合わせ」という特定の分野における対話文 1,056 文の翻訳正解率は 60～80%，総計で 70.5%となる結果が示されている．しかし，この手法が日英の翻訳に適用したものであるのに対し，本手法は英日の翻訳を対象としているため明確な比較はできない．しかし，本手法は翻訳例の字面情報から翻訳ルールを帰納的に学習し，それらを用いて翻訳を行うため，理論的には，すべての言語間での翻訳が可能である．したがって，日英の翻訳においても，英日の翻訳における実験結果とほぼ同じ性能を示すと考えられる．さらに，本手法が適応能力を備えたものであるため，対象分野に適応することが可能であり，古瀬らの手法[6)]における実験結果と同等な結果が得られるものと考えられる．また，その他の手法[3)~5)]は，その基本的な枠組みを与えているにとどまっており，翻訳に関する実験結果は示されていない．これに対し，本手法の実験は中学 1 年生レベルの簡単な英文の翻訳であるが，特定の分野を対象としたものではなく，また，具体的な評価実験を行い，その結果，正翻訳率として 61.9%を得ている．そして，本手法は，従来の手法に比べ，高い学習能力を持ちながら，ユーザあるいは対象分野に速やかに適応することが可能であるという点で非常に有望な手法であるといえる．

3.4.3 問 題 点

遺伝的アルゴリズムを適用することにより，有効な翻訳から無効な翻訳となったものは，表 4 で示したように 22 文であった．その原因は，遺伝的アルゴリズムの適用により誤った翻訳ルールの抽出が行われたためである．この遺伝的アルゴリズムの適用による誤った翻訳ルールの抽出の原因は以下の 2 つに分類することができる．

①交叉において英文とその日本語訳文の持つ性質が誤って継承された翻訳例が生成されるため
②日本語訳文において表現の違いを認識できずに，交叉の際に誤った翻訳例が生成されるため

表 4 における 22 文の翻訳結果が，これらの 2 つの分類のいずれに該当しているのかについて，その内訳とそれぞれの具体例を**表 12** に示す．表 12 に示す文の性質の誤った継承による翻訳例の生成は，英文の交叉において継承される文の性質と日本語訳文の交叉において継承される文の性質が対応しなかったものである．英文の交叉では否定文と肯定文の性質が交叉前の文の性質をそのまま継承しているが，日本語訳文では

表 12 遺伝的アルゴリズムの適用による誤った翻訳
Table 12 The mistakes translations by applying of genetic algorithms.

文の性質の誤った継承：5 文	遺伝的アルゴリズムを適用しなかった場合の翻訳過程	遺伝的アルゴリズムを適用した場合の翻訳過程
文例： Jane's sister speaks Japanese.	翻訳結果：ジェーンの姉は日本語を話します. (@0 speaks Japanese.; @0/は/日本語/を/話し/ます.) (Jane's @0；ジェーンの/@0) (sister；姉)	翻訳結果：ジェーンの姉ではありませんは日本語を話します. (@0 speaks Japanese.; @0/は/日本語/を/話し/ます.) (Jane's @0；ジェーンの/@0) (sister；姉/ではありません)

誤った翻訳ルール（sister；姉/ではありません）の抽出過程

①既存の翻訳例
[1-1]（She is not my sister.；彼女/は/私の/姉/ではありません.）
[1-2]（You are my classmate.；あなた/は/私の/級友/です.）
②（my；は/私の）を交叉位置とした交叉により生成された翻訳例
[1-3]（She is not my classmate.；彼女/は/私の/級友/です.）
[1-4]（You are my sister.；あなた/は/私の/姉/ではありません.）
③抽出された誤った単語の翻訳ルール
[1-1] と [1-3] より（sister；姉/ではありません）

表現の違いによる誤り：17 文	遺伝的アルゴリズムを適用しなかった場合の翻訳過程	遺伝的アルゴリズムを適用した場合の翻訳過程
文例： Mike is an American boy.	翻訳結果：@0 はアメリカの少年です. (@0 is an @1.；@0/は/@1/です.) (American @0；アメリカの/@0) (boy；少年)	翻訳結果：@0 はアメリカの誰です. (@0 is an @1.；@0/は/@1/です.) (American @0；アメリカの/@0) (boy；誰)

誤った翻訳ルール（boy；誰）の抽出過程

① 既存の翻訳例
[1-1]（Who is this boy?；この/少年/は/誰/ですか？）
[1-2]（Is this Judy's brother?；こちら/は/ジュディの/弟/ですか？）
②（this；は）を交叉位置とした交叉により生成された翻訳例
[1-3]（Is this boy?；こちら/は/誰/ですか？）
[1-4]（Who is this Judy's brother?；この/少年/は/ジュディの/弟/ですか？）
③ 抽出された誤った翻訳ルール
[1-2] と [1-3] より（boy；誰）

否定文と肯定文の性質が交叉前の文の性質とは異なる性質を継承している．その結果，英文の性質とその日本語訳文の性質が対応していない誤った翻訳例が生成されることになる．このような誤った翻訳例から抽出される翻訳ルールは，誤った性質を継承することになる．誤った翻訳ルールは，基本的にはフィードバック部において淘汰の対象となり消滅していくが，試用期間中に使用された場合，誤翻訳を生成する原因となる．今後は，無効な翻訳結果生成の原因となるこのような誤った翻訳ルールの抽出を防ぐための交叉手法を取り入れる必要がある．また，日本語訳文の表現の違いにより生成された翻訳例から抽出される誤った翻訳ルールは，学習を十分に行うことにより解消されていくが，まだ，誤った翻訳ルールとしての評価を受けていない段階で使用することにより誤翻訳の原因となった．

4. おわりに

本論文では，帰納的学習による機械翻訳手法へ遺伝的アルゴリズムを適用することの有効性を大量のデータを用いて行った性能評価実験の結果より述べた．その結果，遺伝的アルゴリズムの適用が，遺伝的アルゴリズムを適用しなかった場合と比べ，少量のデータからの多様かつ良質な翻訳ルールの学習能力の向上と効率の良い最適な翻訳結果の生成を可能にするために有効であることが確認された．有効な翻訳率は 52.8%から 61.9%に増加した．また，無効な翻訳においては，その約 47%が名詞句や形容詞などの未登録語ではなく，与えても容易に正翻訳が得られない動詞の単語の初めての出現によるものであったことから，学習を十分に行うことで，翻訳率を向上させることが可能とな

る．そして，本手法における新出語に対する抽出率が47.8%から75.2%に増加したことから，新出語に対する学習能力も向上していることが確認できた．

また，本手法に基づくシステムは既存の機械翻訳システムに比べるとかなり翻訳率が劣っている．しかし，本手法は，次に述べる3つの点より，従来の手法を超えるものであると考えられる．第1に，本手法は，例外的な文章を処理することが困難であるという解析型の機械翻訳手法の問題点を解決する学習型の機械翻訳システムを実現している．第2に，遺伝的アルゴリズムの適用により，学習型の機械翻訳手法の膨大な量の学習データを必要とするという問題点を解決し，正翻訳率および翻訳の質を向上させることができる．第3に，本手法は，あらかじめ規則を与える必要がなく，また，遺伝的アルゴリズムを適用することにより，入力された翻訳例からシステム自身が知識を学習し，それらを利用する能力が向上する．その結果，従来の学習型の機械翻訳手法に比べ学習能力が高く，ユーザあるいは対象分野に柔軟に適応することが可能となる．このようなアプローチは従来の手法とは異なるものである．したがって，本手法は精度および質の高い実用的な機械翻訳システムの実現に向け有効な手法になると考えられる．

今後は，誤翻訳を防ぐために淘汰の精度向上や交叉手法の検討などの遺伝的アルゴリズムの適用レベルでの改良を行う．そして，より多くの実用レベルの文章の翻訳を可能にする，実用的な学習型の機械翻訳システムの実現に向けて研究を進める予定である．

参考文献

1) 長尾　真：機械翻訳サミット，オーム社 (1989).
2) 野村浩郷（編）：言語処理と機械翻訳，講談社 (1991).
3) 赤間　清：帰納的学習システム LS/1 による翻訳の学習，人工知能学会誌，Vol.2, No.3, pp.341–349 (1987).
4) 佐藤理史：MBT2：実例に基づく翻訳における複数翻訳例の組合せ利用，人工知能学会誌，Vol.6, No.6, pp.861–871 (1991).
5) 野美山浩：事例の一般化による機械翻訳，情報処理学会論文誌，Vol.34, No.5, pp.905–912 (1993).
6) 古瀬　蔵，隅田英一郎，飯田　仁：経験的知識を活用する変換主導型機械翻訳，情報処理学会論文誌，Vol.35, No.3, pp.414–425 (1994).
7) 荒木健治，栃内香次：帰納的学習による語の獲得および確実性を用いた語の認識，電子情報通信学会論文誌，Vol.J75-D-ll, No.7, pp.1213–1221 (1992).
8) 荒木健治，高橋祐治，桃内佳雄，栃内香次：帰納的学習によるべた書き文のかな漢字変換手法の適応能力の評価，電子情報通信学会信学技報，NLC 94-3, pp.17–24 (1994).
9) 荒木健治，栃内香次：多段階共通パターン抽出法を用いた翻訳例からの帰納的学習による翻訳，情報処理北海道シンポジウム'91, pp.47–49 (1991).
10) 内山智正，荒木健治，宮永喜一，栃内香次：帰納的学習による機械翻訳手法の評価実験，情報処理学会研究報告，NL 93-4, pp.23–30 (1993).
11) Goldberg, D.E.: *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley (1989).
12) 北野宏明：遺伝的アルゴリズム，産業図書 (1993).
13) 安居院猛，長尾智晴：ジェネティックアルゴリズム，昭晃堂 (1993).
14) 越前谷博，荒木健治，桃内佳雄：帰納的学習による機械翻訳手法への遺伝的アルゴリズムの適用，北海学園大学工学部研究報告，22, pp.275–283 (1995).
15) 越前谷博，荒木健治，桃内佳雄，栃内香次：実例からの帰納的学習による機械翻訳手法における遺伝的アルゴリズムの有効性について，情報処理学会研究報告，NL 107–10, pp.75–82 (1995).
16) De Jong, K.：Learning with Genetic Algorithms: An Overview, *Machine Learning*, Vol.3, No.3, pp.121–138 (1988).
17) De Jong, K. and Spears, W.:Learning Concept Classification Rules Using Genetic Algorithms, *IJCAI'91*, pp.651–656 (1991).
18) 山村雅幸，小野貴久，小林重信：形質の遺伝を重視した遺伝的アルゴリズムに基づく巡回セールスマン問題の解法，人工知能学会誌，Vol.7, No.6, pp.1049–1059 (1992).
19) 越前谷博，荒木健治，桃内佳雄：遺伝的アルゴリズムを用いた帰納的学習による機械翻訳手法，平成6年度電気関係学会北海道支部連合大会講演論文集，No.162 (1994).
20) 内山智正，荒木健治，宮永喜一，栃内香次：帰納的学習による機械翻訳手法の改良，平成5年度電気関係学会北海道支部連合大会講演論文集，No.339 (1993).
21) 教科書ガイド教育出版版ワンワールド1，日本教材，東京 (1991).
22) 教科書ガイド東京書籍版ニューホライズン1，あすとろ出版，東京 (1991).

(平成7年9月21日受付)
(平成8年5月10日採録)

越前谷 博（学生会員）

1967年生．1991年北海学園大学工学部電子情報工学科卒業．1996年同大学院工学研究科電子情報工学専攻修士課程修了．現在，北海道大学大学院工学研究科電子情報工学専攻博士後期課程在学中．自然言語処理，機械翻訳の研究に興味を持つ．1994年度情報処理学会北海道支部奨励賞受賞．電子情報通信学会会員．

荒木 健治（正会員）

1982年北海道大学工学部電子工学科卒業．1988年同大学院博士課程修了．工学博士．同年，北海学園大学工学部電子情報工学科助手．平成元年同講師．1991年同助教授．機械学習を用いた自然言語処理の研究に従事．電子情報通信学会，言語処理学会，ACL等各会員．

桃内 佳雄（正会員）

1942年生．1972年北海道大学大学院工学研究科精密工学専攻博士課程単位取得退学．同大学工学部勤務を経て，現在，北海学園大学工学部電子情報工学科教授．工学博士．自然言語の理解と生成に関する研究に従事．電子情報通信学会，言語処理学会，計量国語学会，人工知能学会，日本認知科学会，ACL等各会員．

栃内 香次（正会員）

1939年生．1962年北海道大学工学部電気工学科卒業．1964年同大学院工学研究科電気工学専攻修士課程修了．現在同工学部電子情報工学専攻教授．工学博士．主として自然言語処理，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．