

## 遺伝的アルゴリズムにおける解析的知識を用いた 制約の有効性

工藤 晃一<sup>†</sup> 越前谷 博<sup>†</sup> 荒木 健治<sup>††</sup> 桃内 佳雄<sup>†</sup> 栢内 香次<sup>††</sup>

<sup>†</sup> 北海学園大学工学部, <sup>††</sup> 北海道大学大学院工学研究科

我々は、学習型機械翻訳手法への遺伝的アルゴリズムの適用の有効性を確認するために、遺伝的アルゴリズムを用いた実例からの帰納的学習による機械翻訳手法 (GA-ILMT) を提案した。しかし、遺伝的アルゴリズムの適用の問題点も同時に確認された。翻訳例の多様性の増加により、誤った翻訳例が多数生成され、正しい翻訳結果の生成が妨げられるのである。我々は、この問題を解決するために、翻訳例生成の多様性を制約し、翻訳例の生成を行う多段階交叉位置決定手法を提案した。本論文では、本手法を適用した GA-ILMT の評価実験より、学習型機械翻訳手法における遺伝的アルゴリズムへの解析的知識を用いた制約の有効性を示す。

## The Effectiveness of Constraints for Genetic Algorithms by Analytical Knowledge

Koichi Kudo<sup>†</sup>, Hiroshi Echizen-ya<sup>†</sup>, Kenji Araki<sup>††</sup>,  
Yoshio Momouchi<sup>†</sup> and Koji Tochinai<sup>††</sup>

<sup>†</sup> Hokkai Gakuen University, <sup>††</sup> Hokkaido University

We previously proposed a method of Machine Translation using Inductive Learning with Genetic Algorithms (GA-ILMT) to confirm the effectiveness of application of genetic algorithms for machine translation method with learning. Then we confirmed its effectiveness from experiments for GA-ILMT. However some problems for application of genetic algorithms are confirmed, too. To solve those problems, we proposed Multi-Stage Decision Method for Generation of New Translation Examples. This method constrains generations of new translation examples. In this paper, we show the effectiveness of constraints for genetic algorithms applied to machine translation with learning from experiments of GA-ILMT applying this method.

### 1 はじめに

現在、様々な機械翻訳システムが考案され、アプリケーションとして利用されている [1]。我々は、機械翻訳システムの更なる性能向上を目指し、学習により知識を獲得する能力を持った学習型機械翻訳システムの実現に向け、遺伝的アルゴリズムを用いた実例からの帰納的機械翻訳手法 (GA-ILMT) を提案した [2]。遺伝的アルゴリズムとは、多様な個体を生成しながら、周囲の環境に適応することで進化していく生物の進化過程のアルゴリズムである [3]。そして、様々な分野に適用され、その有効性は立証さ

れている [2, 3]。この手法では、与えられた翻訳例から自動的かつ帰納的に翻訳ルールを獲得し、そこで学習された翻訳ルールを用いて翻訳を行う。更に、遺伝的アルゴリズムの適用により、与えられた翻訳例から多様な翻訳例をシステムが自動的に生成し、大量の翻訳例を獲得することを可能にしている。そして、これまで行われた評価実験により、提案手法の適用の有効性 [2] は立証されている。

しかし、GA-ILMT に適用された遺伝的アルゴリズムには、様々な問題が残されている。本来、遺伝的アルゴリズムでは、多様な個体を生み出しながら

淘汰により適応度の低い個体は消滅し、適応度の高い個体のみが残る。GA-ILMTにおいても、個体である翻訳ルールを評価して適応度を決定し、淘汰処理を行うことは可能である。しかし、自然界では、与えられた環境の中に全ての個体が存在し、全ての個体が評価されるが、GA-ILMTでは、一つの翻訳結果の生成に対して適用された翻訳ルールのみを評価対象とするため、適応度を決定できない翻訳ルールが多数存在するのである。この問題を解決するためには、全ての個体を評価することが必要であるが、多くの翻訳結果の生成と時間を要する。そして、全ての個体の適応度が決定するまでの時間は膨大であるため、機械翻訳の実用面から考えるとこれは非常に困難である。したがって、適応度を決定することができない翻訳ルールが多数存在することになり、淘汰処理の機能を十分に発揮することができない。最も問題なのが、適応度を決定できない翻訳ルール中の誤った翻訳結果を誘発する無効な翻訳ルールの存在である。本来、このような翻訳ルールは、翻訳結果の生成により評価され、適応度が自動的に低下し淘汰されるべきである。しかし、GA-ILMTでは、このような翻訳ルールは、淘汰処理が不十分であるため、適応度が決定できなく淘汰できない。このように、すぐに淘汰されるべき翻訳ルールが、環境によって評価されずに適応度が決定できないため、淘汰が行われずに、次第に増加していくのである。そして、辞書が肥大化し、また、翻訳に対しても悪影響を与える。このような問題を解決するには、淘汰処理の改善が考えられるが、適応度の決定に問題があるため、改善は難しい。したがって、翻訳例の生成に対して制限を加え、淘汰処理に依存せずに、適応度が決定できない無効な翻訳例の生成を減少させることが有効であると考えられる。

そこで、我々は、GA-ILMTの翻訳例の生成部に、様々な解析的知識を用いて翻訳例の生成に制約を与えることを目的とした多段階交叉位置決定手法 [4, 5] を適用することを提案した。翻訳例の生成段階で翻訳例の生成を制限することにより、淘汰処理に依存することなく将来的に適応度が低くなると考えられる翻訳例とそれらから抽出される翻訳ルールを減少させることが可能となる。したがって、翻訳ルールによる翻訳結果へのこのような影響を小さくすることが可能になる。しかし、多様な翻訳例を生成するという観点から考えると翻訳例の生成に対する知識による制約には問題がある。知識による制約は、将

来的に適応度が高くなると考えられる翻訳例の生成を制限することもあるためである。したがって、翻訳例の生成に対する知識の導入において、翻訳例の多様性の損失を最小限に抑え、かつ、淘汰されるべき翻訳例の生成を抑えることが必要である。この2つの点を同時に可能になったとき、遺伝的アルゴリズムへの解析的知識による制約は、有効なものになると考えられる。このように遺伝的アルゴリズムへの解析的知識の導入による制約の有効性は、自然言語処理分野におけるアプリケーションへの遺伝的アルゴリズムの利用という観点から評価すべきことである。

本論文では、GA-ILMTに本手法を組み込んだシステムを用いて行った評価実験とその結果より、本手法を適用していない従来のGA-ILMTと比較をし、学習型機械翻訳システムに適用された遺伝的アルゴリズムへの解析的知識の導入による制約の有効性についての考察を行う。

## 2 GA-ILMT

### 2.1 GA-ILMTの処理過程

従来のGA-ILMTは、帰納的学習による機械翻訳手法に対して遺伝的アルゴリズムを適用した学習型機械翻訳手法である。GA-ILMTにおける基本的な処理過程を図1に示す。GA-ILMTは、様々な言語に適用可能な学習型機械翻訳手法である。現在、我々の研究では、英日機械翻訳システムに適用している。翻訳例と遺伝的アルゴリズムにおける個体の染色体と遺伝子の対応を図2で示す。GA-ILMTでは、染色体として英文とその日本語訳文を組した翻訳例を対応させる。そして、染色体を構成している遺伝

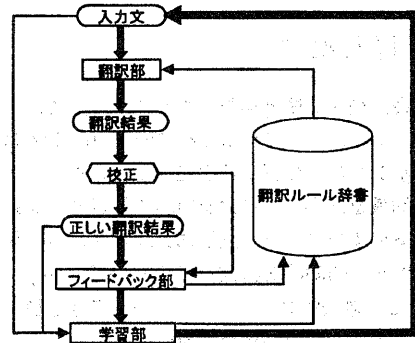


図1: GA-ILMTの処理過程

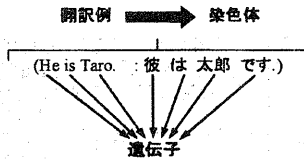


図 2: 染色体と遺伝子

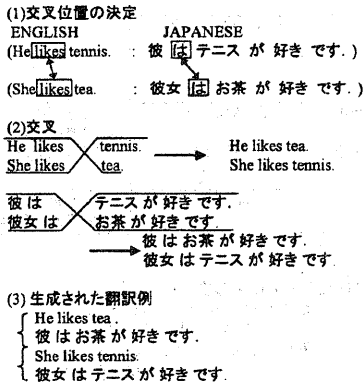


図 3: 従来の GA-ILMT における翻訳例の生成例

子には翻訳例を構成している単語を対応させる。原文として英文を入力すると、翻訳部において、学習部で既に獲得されている翻訳ルールに対する遺伝的アルゴリズムの基本操作を行い、翻訳結果を生成する。次に、フィードバック部において、ユーザーが与えた正しい日本語訳文を用い、翻訳部で用いら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、与えられた翻訳例（英文とその日本語訳文）に対して、交叉及び突然変異を行い多様な翻訳例を生成する。従来の GA-ILMT の翻訳例の生成では、それまでに入力された翻訳例から英文と日本語訳文のそれぞれにおいて、字面が完全に一致する文字列を共通部分を持つ 2 つの翻訳例を選択し、共通部分を交叉位置として一点交叉を行う。図 3 に示す例では、英文において「likes」、日本語訳文においては、「は」と「が大好きです」が共通部分として抽出される。そして、英文においては「likes」、日本語訳文においては「は」が交叉位置として抽出される。また、「が大好きです」を交叉位置とすると同じ日本語訳文が生成されるので翻訳ルールとしての登録は行わない。また、同時に翻訳ルールの抽出も行う。翻訳ルールの抽出は、2 つの翻訳例の共通部分と差異部分を多段階に抽出する。

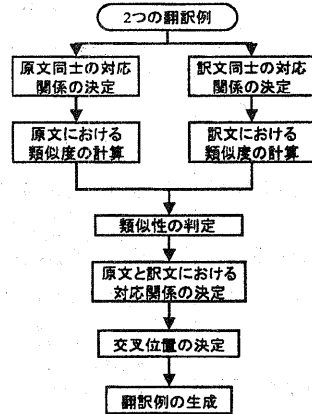


図 4: 多段階交叉位置決定手法の処理過程

## 2.2 多段階交叉位置決定手法

### 2.2.1 概要

GA-ILMT の学習部に適用される本手法は、翻訳例同士の類似性による交叉の制約と確実な対応関係を持つ交叉位置の単語を決定する手法である [4, 5]. 本手法の処理過程を図 4 に示す。本手法では、2 つの翻訳例における原文同士及び訳文同士の対応関係を多段階に確実性の高い順に決定し、次いで原文同士の類似度と訳文同士の類似度をそれぞれ計算する。類似度がしきい値未満の時は、その翻訳例を処理対象から除外し、交叉を行わない。しきい値以上の時は、処理を続行し、対応関係から交叉位置の候補を決定する。そして、その候補における原文と訳文の対応関係を決定し、この対応関係を持つ交叉位置の候補のみを交叉位置として交叉を行う。また、本手法では、英文と日本語訳文の両方で品詞情報を使用するため、入力文は、形態素解析を行っておく必要がある。原文は、Bril Tagger[6] を、訳文には、我々が開発した帰納的学習による形態素解析手法 [7] を用いる。また、対応関係の決定に用いる辞書とツールを表 1 に示す。

### 2.2.2 原文同士及び訳文同士の対応関係の決定

本手法では、交叉対象の候補となる 2 つの翻訳例における原文と訳文の構成単語に対して、次のような対応関係を決定する。

- (1) 出現位置が同じで字面が一致する単語間の対応関係。
- (2) 出現位置が異なり字面が一致する単語間の対応

表 1: 対応関係に用いた辞書とツール

種別	名称	語数	作成者	使用段階	使用目的
シソーラス	WordNet	126,516	Princeton University	3. 原形的一致 5. 上位概念の一致	不規則変化 英文の上位概念の検索
	分類語彙表	32,600	国立国語研究所	5. 上位概念の一致	日本語文の上位概念の検索
形態素辞書	Oxford Advanced Learner's Dictionary of Current English	70,646	University of London	3. 原形的一致	規則変化
英和辞書	英和・和英電索辞書	55,000	久保正治 (技術評論社)	4. 同一単語の訳語	訳語の取得
訳語リスト	ONE WORLD English Course 1 NEW EDITION の単語リスト	389	長谷川潔 他	4. 同一単語の訳語	訳語の取得
形態素解析ツール	Brill Tagger	1,429,865	Eric Brill	7. 品詞の一致	英文の品詞情報の取得
	帰納的学習による形態素解析手法	11,113	荒木健治 他	3. 表記と読み的一致	読みの取得
				7. 品詞の一致	日本語文の品詞情報の取得

関係。

- (3) 表記の属性が一致する単語間の対応関係。
- (4) 同一の単語の訳語が存在する単語間の対応関係。
- (5) 上位概念が一致する単語間の対応関係。
- (6) 決定済みの対応関係に挟まれている単語間の対応関係。
- (7) 品詞の一致する単語間の対応関係。

以上の7段階で対応関係を決定する。但し、(3)の段階では、次のように原文と訳文で処理方法が異なる。

- 原文では、原形が一致する単語間の対応関係を決定。
- 訳文では、読みが一致する単語間の対応関係を決定。

上位の段階の対応関係ほど対応関係の確実性が高い。また、上位の段階で決定された対応関係は下位の段階では、処理対象としない。ここで使用する対応関係は、文同士の類似性を示すために重要と考えられるものを使用した。また、対応関係の順位は、確実性が高いと考えられる順に決定した。ここで決定された対応関係を持つ2つの単語は、対となって原文又は訳文の交叉位置の候補となる。

### 2.2.3 類似度の計算

次に、この処理部で用いられた翻訳例の対応関係から決定される文同士の類似度の計算について述べる。この類似度は、決定された各段階における対応

・原文同士の対応関係の決定と類似度

原文 1: Do you like rock music?

↓(3) ↓(6) ↓(1) ↓(6) ↓(1)

原文 2: Does he like pop music?

類似度 = 70.0

・訳文同士の対応関係の決定と類似度

訳文 1: あなたはロック音楽が好きですか?

\*(5) ↓(1) ↓(6) \*(1) ↓(1) ↓(1) ↓(1)

訳文 2: 彼はポピュラー音楽が好きですか?

類似度 = 82.5

図 5: 対応関係と類似度の例

関係の個数に基づいて式 (1) で計算される。

$$\text{類似度} = \frac{100.0}{S} \times (2.0 \times N_1 + 1.8 \times N_2 + 1.6 \times N_3 + 1.0 \times N_4 + 0.8 \times N_5 + 0.4 \times N_6 + 0.2 \times N_7 - 0.5 \times N_8) \quad (1)$$

S: 2つの翻訳例の原文又は訳文の単語の合計

$N_n$ : n 段階における対応関係の数 ( $n=1, \dots, 7$ )

$N_8$ : 対応関係がなかった単語数

この式において、全く同じ文同士の類似度は 100 となる。また、各項の係数は、上位の段階の対応関係の単語が多いと類似度が高くなり、下位の段階の対応関係や対応関係を持たない単語が多くなるほど類似度が低くなるように設定した。上位の段階の対応関係が多いほど、類似度が高くなり、2つの文の類似性が高いことになる。類似度は、原文同士及び訳文同士のそれぞれについて計算される。各項の係数は予備実験の結果より最適な値に設定されている。図 5 に、対応関係と類似度の計算の例を示す。それぞれの矢印は、対応関係が存在する単語を示す。()

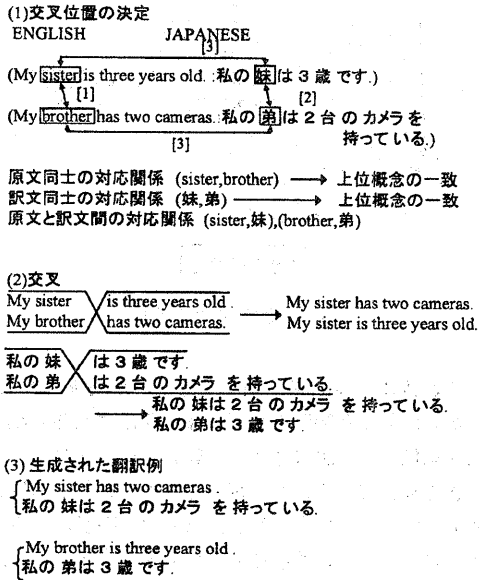


図 6: 翻訳例の生成例

内に示された数字は、対応関係の段階の数である。また、この各段階の対応関係の個数を(1)式に代入することにより、類似度が決定される。

#### 2.2.4 類似性の判定

類似性の判定について述べる。2つの翻訳例の原文と訳文の類似度の値により、この翻訳例を用いて交叉を行うかどうか判定する。原文と訳文における類似度が両方ともにしきい値以上である場合、この翻訳例を交叉対象とする。どちらか一方がしきい値未満である場合は、この翻訳例を交叉対象とはせずに、処理から除外する。このように類似性を判定して、しきい値以上の翻訳例同士を交叉させることにより、交叉を制約することが可能となる。

#### 2.2.5 原文と訳文の対応関係の決定

原文の単語と訳文の単語における対応関係の決定には、英和辞書[8]を使用する。辞書から訳文における交叉位置の候補となる単語を検索し、対応する原文における交叉位置の候補となる単語を決定する。

#### 2.2.6 交叉位置の決定と翻訳例の生成

交叉位置の決定には、原文同士及び訳文同士の対応関係と原文と訳文の対応関係の情報を用いる。次の条件を満たす交叉位置の候補の単語を交叉位置として決定する。原文及び訳文の交叉位置の候補で

ある単語が、それぞれの翻訳例の中で原文と訳文の対応関係が存在するとき、交叉位置として決定される。これにより、原文と訳文における交叉位置は、結合される。そして、この後、決定された交叉位置で一点交叉を実行し、新たな翻訳例を生成する。図6は、類似度がしきい値以上である場合の交叉位置の決定と翻訳例の生成の例である。図で示されている[1],[2],[3]は、それぞれ原文同士、訳文同士、原文同士と訳文同士の対応関係を示している。原文同士では、上位概念の一致により、"sister"と"brother"の対応関係が決定される。また、訳文同士においても上位概念の一致により、"妹"と"弟"の対応関係が決定される。これらが交叉位置の候補となり、原文と訳文の対応関係の決定により"sister"と"妹"、"brother"と"弟"の対応関係が決定される。これらの交叉候補の組合せは、交叉位置となる条件を満たすため、この位置で一点交叉が実行される。交叉位置より右側の部分が、原文と訳文において入れ替えられる。そして、図で示されるような新翻訳例が生成される。

### 3 評価実験

#### 3.1 評価方法

本手法に基づいたシステムの翻訳結果に対する評価方法について述べる。翻訳結果は以下の2つの評価方法で分類する。

- (1) 正翻訳
- (2) 誤翻訳

正翻訳は、翻訳結果が原文の正しい訳文であるときの翻訳である。このとき未登録語を含んでいる翻訳結果において、未登録語に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。したがって、未登録語を含んでいる翻訳結果も正翻訳に含めることにする。誤翻訳は、翻訳結果が原文の正しい訳文ではないときの翻訳である。また、適用可能な翻訳ルールが全くないときに翻訳不能となる場合も誤翻訳に含む。正翻訳率は以下の式(2)で決定される。

$$\text{正翻訳率} [\%] = \frac{\text{正翻訳数}}{\text{全翻訳数}} \times 100 \quad (2)$$

また、生成された翻訳結果が複数存在する場合には、順位付けされて出力される翻訳結果の最上位から10位までを評価対象とする。10個の翻訳結果中に正

表 2: 実験結果

システム	正翻訳数	誤翻訳数	正翻訳率 [%]
(A)	317	480	39.8
(B)	370	427	46.4

(A):従来の GA-ILMT  
(B):本手法を適用した GA-ILMT

表 3: 生成された翻訳例数

システム	正翻訳例数	誤翻訳例数	合計	精度 [%]
(A)	1,404	4,430	5,834	24.1
(B)	516	396	912	56.6

(A):従来の GA-ILMT  
(B):本手法を適用した GA-ILMT

翻訳に該当するものが存在していた場合、その翻訳結果を正翻訳とする。

### 3.2 実験方法

本手法を適用した GA-ILMT と比較のために従来の GA-ILMT について実験を行う。実験には、中学校 1 年生用教科書ガイド・ワンワールド [9] に掲載されている英文と日本語訳文の 794 組を使用した。実験では 1 文ごとに翻訳を行い、翻訳終了時に正翻訳文を与え、学習も同時に行った。また、多段階交叉位置決定手法におけるしきい値は、予備実験 [10] より決定された最適値である 55 点とした。

### 3.3 実験結果

実験結果を表 2 に示す。従来の GA-ILMT と本手法を適用した GA-ILMT における翻訳率は、それぞれ 39.8% と 46.4% となった。表 3 に実験で生成された翻訳例の総計を示す。生成された翻訳例は、正しい翻訳例と誤った翻訳例と 2 種類に分類した。従来の GA-ILMT と本手法を適用した GA-ILMT で生成された翻訳例の精度は、それぞれ、24.1%, 56.6% となった。この精度は、式 (3) で計算される。

$$\text{精度 [\%]} = \frac{\text{正翻訳例数}}{\text{全生成翻訳例数}} \times 100 \quad (3)$$

## 4 考察

### 4.1 本手法を適用した GA-ILMT の翻訳

実験結果より、本手法を適用した GA-ILMT の翻訳率が 6.6 ポイント向上していることが確認できる。また、表 3 より、本手法の適用により、生成される翻訳例の数が大きく減少し、さらに精度が向上している。従来の GA-ILMT では、生成される翻訳例の数が多し。正しい翻訳例の絶対数も多く、翻訳に適

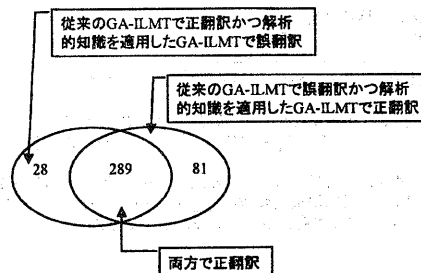


図 7: 翻訳数の比較

した翻訳ルールも多く抽出することが可能である。しかし、誤った翻訳例が更に多く生成される。誤った翻訳例は、翻訳に適さない翻訳ルールを抽出する大きな要因である。したがって、正しい翻訳結果が生成されたとしても、翻訳に適さない翻訳ルールで生成された誤った翻訳結果が多く発生するため、正しい翻訳結果が、上位の翻訳結果の候補となることが不可能となる。本手法を適用した GA-ILMT では、生成される翻訳例は少なく、生成される翻訳例の多様性は少ないが、誤った翻訳例の生成が減少するため精度が高い。したがって、従来の GA-ILMT のように誤った翻訳例の悪影響が小さいため、正翻訳率が向上したと考えられる。しかし、生成される翻訳例の絶対数が少ないため、翻訳に必要な翻訳ルールが得られず、翻訳が不可能となることがある。これは、遺伝的アルゴリズムに対して適用した解析的知識による制約によって起きる生成される翻訳例の多様性の損失である。この問題は、最適なしきい値を用いることにより、多様性の損失を最小限にし、翻訳率を最大にすることが可能であると考えられる。

図 7 に、従来の GA-ILMT と本手法を適用した GA-ILMT における正翻訳と誤翻訳の数とその相互関係を示す。この図は、従来の GA-ILMT で正翻訳となり、本手法を適用した GA-ILMT で誤翻訳となる数が 28 個、反対に、従来の GA-ILMT で誤翻訳で、本手法を適用した GA-ILMT で正翻訳となる数が 81 個、そして、両方の GA-ILMT で正翻訳となる数が 289 個であることを示している。図 8 は、従来の GA-ILMT で誤翻訳、本手法を適用した GA-ILMT で正翻訳となる例である。従来の GA-ILMT では、生成された翻訳結果の上位 10 番以内に正しい翻訳結果が出力されていない。一方、本手法を適用した GA-ILMT では、正しい翻訳結果が最上位の候補となっている。従来の GA-ILMT では、上位

入力文 :Does she have a good bike ?

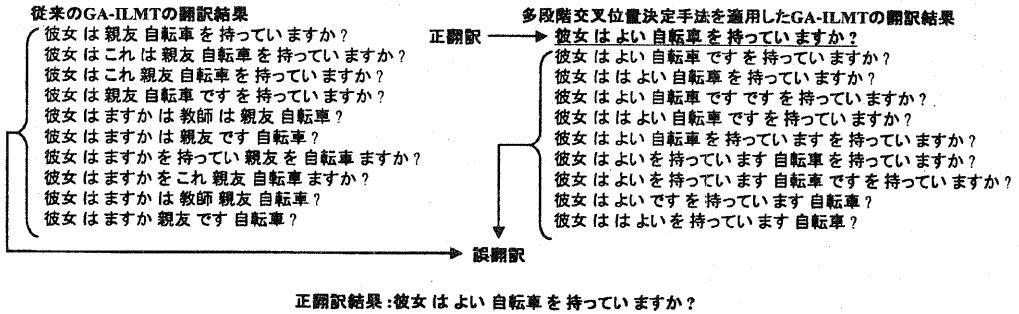


図 8: 従来の GA-ILMT で誤翻訳, 本手法を適用した GA-ILMT で正翻訳となる例

入力文 :You play baseball every Sunday .

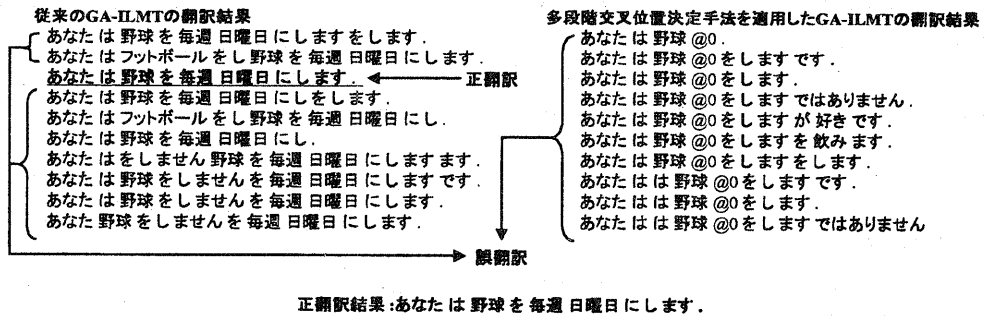


図 9: 従来の GA-ILMT で正翻訳, 本手法を適用した GA-ILMT で誤翻訳となる例

10 番目よりも下位の翻訳結果に正しい翻訳結果が出力されている。これは、従来の GA-ILMT において誤った翻訳結果を生成する翻訳ルールを抽出する翻訳例が生成されるため、正しい翻訳結果が上位 10 番以内に入ることができないためである。また、本手法を適用した GA-ILMT では、解析的知識を用いた制約により、そのような翻訳ルールが抽出される翻訳例が生成されないため、正しい翻訳結果が最上位の候補になる。この例からも、本手法による誤った翻訳例に対する制約の効果が確認できる。そして、図 9 は、従来の GA-ILMT で正翻訳、本手法を適用した GA-ILMT で誤翻訳となる例である。この例では、本手法を適用した GA-ILMT では、この入力文の正しい翻訳結果を生成するための翻訳ルールを抽出する翻訳例が生成されない。そのため、正しい翻訳結果が生成できないのである。一方、従来の GA-ILMT では、正しい翻訳結果の生成に必要な翻訳例が生成されるため、正しい翻訳結果が生成可

能である。したがって、この図から解析的知識を用いた制約による多様性の損失を確認できる。このように、本手法を適用した GA-ILMT において 81 文が正翻訳となったのは、本手法の適用による誤った翻訳例の生成に対する制約の効果であり、また、誤翻訳となった 28 文は、制約による多様性の損失の影響が原因であると考えられる。

## 4.2 生成される翻訳例への制約の有効性

実験結果より、解析的知識を用いた制約が翻訳例の生成及び翻訳率に影響することが確認された。しかし、実験結果からもわかるように翻訳率に対する影響は、翻訳例の生成ほどではない。これは、多段階交叉位置決定手法が学習部における翻訳例の生成に対してのみ適用され、翻訳部やフィードバック部などの処理に対して、本手法のような知識を導入した処理を全く適用していないためである。更に、翻訳例の生成における制約が、学習部を通して翻訳

部に影響するため、直接的に翻訳に影響しないためである。しかし、解析的知識による制約を用いることにより、不要な翻訳例の生成を抑えながら、翻訳率を向上させることが可能である。これは、従来のGA-ILMTと本手法を適用したGA-ILMTの翻訳例の生成数が5,834個から912個に減少しながらも、翻訳率が向上していることから明らかである。また、これらの翻訳例から獲得される翻訳ルールも減少するため、翻訳辞書の肥大化を防ぐことも可能である。GA-ILMTは、適用されている遺伝的アルゴリズムにより、学習した翻訳例から多様な新たな翻訳例を生成する。しかし、翻訳例の学習を続けるならば、生成する翻訳例も無限に増殖し、辞書の肥大化や誤翻訳ルールの増加による翻訳率の低下を招く。したがって、翻訳例の生成に対して本手法のような制約を加え、制限された交叉を行うことにより、このような問題は解決する。本手法は、GA-ILMTの学習能力の向上に有効であると考えられる。これらから、GA-ILMTへの解析的知識を用いた制約は、有効であると考えられる。

## 5 おわりに

本論文では、多段階交叉位置決定手法による翻訳例生成に対する制約を提案した。そして、その手法に基づく実験システムを用いた性能評価実験から、GA-ILMTに対する解析的知識を用いた制約の有効性を確認した結果について述べた。多段階交叉位置決定手法を適用したとき、翻訳率が向上することから、機械翻訳システムとしてのGA-ILMTの性能を解析的知識による制約が向上させていることが確認できる。したがって、学習型機械翻訳に適用された遺伝的アルゴリズムへの解析的知識の導入による制約は有効である。また、自然言語処理分野における遺伝的アルゴリズムの利用では、本手法のような解析的知識による制約が重要であると考えられる。

しかし、実験結果から解析的知識による制約を適用したときの翻訳率の向上の効果は、それほど大きくはない。これは、学習部における翻訳例の生成という限定した部分に解析的知識による制約を用いているためである。今後は、このような限定された部分に対してのみ解析的知識の導入を行うのではなく、直接翻訳にかかわる翻訳部や淘汰処理を行うフィードバック部に対して解析的知識の導入と知識による制約の適用を更に進めていく予定

である。そして、機械翻訳の発展のために、学習型機械翻訳手法と解析型機械翻訳手法の柔軟な融合を果たし、学習と解析の能力を兼ね備えた機械翻訳システムについて研究を進めていく予定である。

**謝辞** なお、本研究の一部は文部省科学研究費(No.09878070, No.10680367)及び北海学園大学ハイテク・リサーチ・センター研究費による補助のもとに行われた。

## 参考文献

- [1] 長尾真：機械翻訳サミット，オーム社(1989)。
- [2] 越前谷博，荒木健治，桃内佳雄，柄内香次：実例に基づく帰納的学習による遺伝的アルゴリズムの適用とその有効性，情報処理学会論文誌，Vol.2, No.8, pp.1565-1579, (1996)。
- [3] 北野宏明：遺伝的アルゴリズム，産業図書(1993)。
- [4] 工藤晃一，荒木健治，桃内佳雄，柄内香次：多段階交叉位置決定手法における新翻訳例の生成，言語処理学会第3回年次大会発表論文集，p589-592(1997)。
- [5] Kudo, K., Araki, K., Mmouchi, Y. and Tochinnai, K.: Multi-Stage Decision Method for Production of New Translation Examples, Proceedings of the IASTED International Conference Artificial Intelligence and Soft Computing, pp.125-128 (1997)。
- [6] Eric Brill, A CORPUS-BASED APPROACH TO LANGUAGE LEARNING, (1993)。
- [7] 荒木健治，柄内香次：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-II, No.7, pp.1213-1221, (1992)。
- [8] 久保正治：英和・和英電策辞典 gene, 技術評論社，東京, (1995)。
- [9] 教科書ガイド教育出版ワンワールド，日本教材，東京, (1991)。
- [10] 工藤晃一，荒木健治，桃内佳雄，柄内香次：多段階類似性判定手法におけるしきい値の最適化，情報処理北海道シンポジウム'98講演論文集，pp.67-68 (1998)。