

GA-ILMT における状態遷移を用いた

3N-1

翻訳ルールの階層化とその有効性

越前谷博† 荒木健治†† 桃内佳雄† 枋内香次††
 北海学園大学工学部† 北海道大学大学院工学研究科††

1 はじめに

近年、インターネットの普及に伴い異言語で記述された情報に接する機会が増大していると共に、実用的な機械翻訳システムに対するニーズが高まっている。現在、商用化されている機械翻訳システムの多くは、人手で文法規則を与え、それらを用いて文を解析することにより翻訳を行う解析型機械翻訳手法 [1] を取り入れている。しかし、この手法においては、全ての言語現象を網羅するような文法規則を予め与えることは非常に困難である。解析型機械翻訳手法に対し、近年、実例や用例を模倣利用することで翻訳を行う実例型機械翻訳手法 [2, 3, 4] の研究が注目を集めている。この手法においては、大量のコーパスが不可欠であり、それと共に、コーパスに様々な情報を付与する際の膨大な労力が問題点となっている。こうした機械翻訳手法の現状において、我々は、これまでに翻訳例のみから翻訳ルールを自動的に獲得し、それらを用いて翻訳を行う学習型機械翻訳手法として、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 (GA-ILMT) [5] を提案し、その有効性を確認してきた。しかし、獲得した翻訳ルールを最大限に活用するという点においては、十分なものではなかった。本稿では、翻訳ルールを最大限に活用した翻訳を目的に行った、翻訳結果の生成過程の状態遷移を用いた翻訳ルールの自動階層化と、その有効性を確認するために行った翻訳実験の結果について述べる。

2 GA-ILMT の概要

GA-ILMT に基づき構築した英日機械翻訳システムの処理過程を図 1 に示す。原文が入力されると、翻訳部において、それまでに獲得された翻訳ルールを用いて翻訳結果を生成する。誤りが含まれている場合には、人手により校正を行う。フィードバック部では、正しい翻訳結果より、使用した翻訳ルールに対して適応度を決定し、誤翻訳ルールに対する淘汰を行う。そして、学習部では、原文とその訳文の対である翻訳例に対し、遺伝的アルゴリズムの基本操作を施すことにより新たな翻訳例を自動的に生成する。また、翻訳例に対して帰納的学習を行い、多様な翻訳ルールを獲得する。

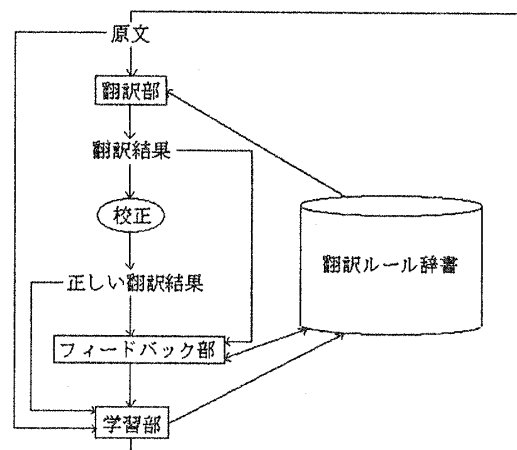


図 1 GA-ILMT のシステム構成

3 状態遷移を用いた翻訳ルールの自動階層化

3.1 基本的な考え方

これまでの GA-ILMT では、獲得された翻訳ルールは全てフラットな状態のままであり、翻訳を行う際に、正翻訳の生成過程に着目したものとはなっていなかった。そこで、我々は、獲得した翻訳ルールを階層化し、どのような翻訳ルールを使用していくことで、正翻訳が導き出されていくのかをシステム自身が認識する機構を取り入れた。階層化された結果、より汎用性が高く、確実性の高い翻訳ルールほど、上位階層に位置付けられる。また、階層化を行う際には、解析的な知識に基づき行うのではなく、翻訳例のみを用いたシステム自身の持つ学習能力により行う。その具体的な方法としては、正翻訳の生成過程の状態遷移を用いる。これは、正翻訳の生成過程というものが、正翻訳を導き出すための学習のメカニズムを模倣していると考えられるからである。そして、その正翻訳の生成過程は、翻訳ルールの状態遷移として表すことができる。したがって、翻訳ルールの自動階層化は、正翻訳の生成過程の状態遷移を用いて行う。

3.2 処理過程

状態遷移を用いた翻訳ルールの自動階層化は、図 1 のシステム構成における学習部で行う。また、翻訳部においても、翻訳処理を行う際に、同様に状態遷移を用いた翻訳ルールの自動階層化を導入している。以下に処理過程を述べる。

Layered Process for Translation Rules using
 State Transition in GA-ILMT and Its Effectiveness
 Hiroshi Echizenya†, Kenji Araki††,
 Yoshio Momouchi† and Koji Tochinal††
 †Faculty of Engineering, Hokkai-Gakuen University
 ††Graduate School of Engineering, Hokkaido University
 e-mail: echi@eli.hokkai-s-u.ac.jp

- (1) 正翻訳の生成過程における状態遷移を生成する。

まず、状態遷移を生成する際に必要となる要素を以下のように定義する。

- ①状態 (翻訳ルール) の集合 K
- ②入力 (翻訳ルール) の集合 $I: I \subset K$
- ③最終状態 (正翻訳例及び抽象化されていない翻訳ルール) の集合 $F: F \subset K$
- ④初期状態の集合 $S: S=K-F$
- ⑤状態遷移関数 δ : ある状態 q_i にあるとき入力 a が与えられ、状態 q_j に遷移する場合、これを

$$\delta[q_i, a] = q_j$$

と表し、 δ を状態遷移関数とする。

- ⑥状態遷移の集合 T

ここで、状態遷移とは、最終状態に到達するまでの状態 (翻訳ルール) の組である。

- (2) 生成された状態遷移において、個々の状態遷移が包含関係にないものを選択する。
- (3) 選択された状態遷移において最終状態から初期状態へ移行している翻訳ルールを下位階層から上位階層へと位置付ける。

例えば、原文 "She is my sister." に対する正翻訳 "彼女は私の姉です。" を生成する際の状態遷移を示す。初期状態の翻訳ルールとして、正翻訳に使用された翻訳ルールの中から q_4 : (@0 is my sister. ; @0/は/私の/姉/です。) を任意に選択したとする。すると、最終状態の q_0 : (She is my sister. ; 彼女/は/私の/姉/です。) に到達するための状態遷移関数は $\delta[q_4, (She; 彼女)] = q_0$ となる。したがって、この場合の状態遷移 t_0 は、 $t_0 = \langle q_4, q_0 \rangle$ となる。このようにして得られた状態遷移の集合から、最も詳細に記述されているものを選択する。そして、その状態遷移を構成している初期状態の翻訳ルールを上位階層、最終状態の翻訳ルールを下位階層に対応付ける。図 2 に階層化された翻訳ルールの例を示す。

階層1

- q_0 : (She is my sister . ; 彼女は私の姉です。)
 q_1 : (She ; 彼女)
 q_2 : (sister ; 姉)
 q_3 : (my sister ; 私の姉)

階層2

- q_4 : (@0 is my sister . ; @0 は私の姉です。)
 q_5 : (She is my @0 . ; 彼女は私の @0 です。)
 q_9 : (my @0 ; 私の @0)

階層3

- q_6 : (@0 is my @1 . ; @0 は私の @1 です。)
 q_7 : (She is @0 . ; 彼女は @0 です。)

階層4

- q_8 : (@0 is @1 . ; @0 は @1 です。)

図 2 階層化された翻訳ルールの具体例

4 評価実験

4.1 実験結果

実験データには、中学 1 年生用教科書ガイド・ワンワールド [6] に掲載されている翻訳例 549 文を使用した。また、実験を行う際の辞書の初期状態は空から始めた。そして、翻訳結果に対する評価は、生成された翻訳結果の上位 1 位のみを対象とし、意味的に正しいものを正翻訳とした。実験の結果、本手法を導入することにより正翻訳率は 32.1% から 38.1% に増加した。

4.2 考察

複数の翻訳結果が生成され、その中に正翻訳が存在しているのにもかかわらず上位 1 位とならなかったものが、本手法の導入により、より多くの正翻訳が上位 1 位に選択されるようになった。原文 1 文に対して生成された複数の翻訳結果の中に正翻訳が存在しているものだけを対象とし、その正翻訳が上位 1 位に選択されているかどうかを調べると、本手法の導入前では 59.9% であったものが、本手法の導入後では 71.1% に増加していた。これは、本手法の導入により、システムが正翻訳の生成過程を考慮した翻訳処理が可能になったためである。

5 おわりに

本稿では、状態遷移を用いた翻訳ルールの自動階層化とその有効性について述べた。正翻訳率は 6.0 ポイント向上し、また、正翻訳の選択の精度は 11.2 ポイント向上した。今後は、より実用的な機械翻訳システムの構築に向けての研究を行う予定である。

謝辞

本研究の一部は、文部省科学研究費補助金 (第 10680367 号, 第 09878070 号) 及び北海道大学ハイテク・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] 野村浩郷 (編): 言語処理と機械翻訳, 講談社, (1991).
- [2] 田中穂積 (監): 自然言語処理—基礎と応用—コロナ社, (1999)
- [3] 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).
- [4] 北村美穂子, 松本裕治: 対訳コーパスを利用した翻訳規則の自動獲得, 情報処理学会論文誌, Vol.37, No.6, pp.1030-1040 (1996).
- [5] 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, 情報処理学会論文誌, Vol.37, No.8 (1996).
- [6] 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).