

245 GA-ILMTにおける状態遷移に基づく翻訳ルールの階層化

越前谷博†

荒木健治† 桃内佳雄† 枅内香次†

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、インターネットに接する機会の急速な増加と共に、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、人手により文法規則を記述し、それらに基づき原文を解析することで翻訳を行う解析型機械翻訳手法を取り入れている[1]。しかし、この解析型機械翻訳手法では、多様な言語現象を完全に網羅できる文法規則を予め与えることの困難さが問題点となっている。この手法に対し、近年、文法規則に依存するのではなく、実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]が大きな注目を集めている。この手法においては、実用的な翻訳のために不可欠な知識をコーパスに付与することの労力や、コーパスそのものが大量に必要となることが問題点となっている。こうした背景において、これまでに我々は、翻訳例のみから翻訳ルールを獲得し、翻訳を行う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（以後、GA-ILMTと呼ぶ）[5]を提案してきた。しかし、これまでのGA-ILMTは、翻訳ルールの獲得及び、獲得した翻訳ルールの最大限の有効活用という点では十分なものとはなっていなかった。そこで、我々は、学習能力に基づく機械翻訳システムの構築という観点から正翻訳の生成過程に着目した。これは、正翻訳の生成過程が、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ためである。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。従って、我々は、正翻訳の状態遷移に基づいて、翻訳ルールの自動階層化する。更に、状態遷移より得られた結果を活用し、新たな良質な翻訳ルールを獲得する。本稿では、GA-ILMTにおいて、状態遷移を導入することの有効性を性能評価実験を通して述べる。

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、獲得された翻訳ルールを用いて翻訳結果を生成する。その翻訳結果に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、翻訳に使用した翻訳ルールに対して適応度を決定し、適応度の低い翻訳ルールを誤翻訳ルールと位置付け淘汰を行う。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。更に、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。

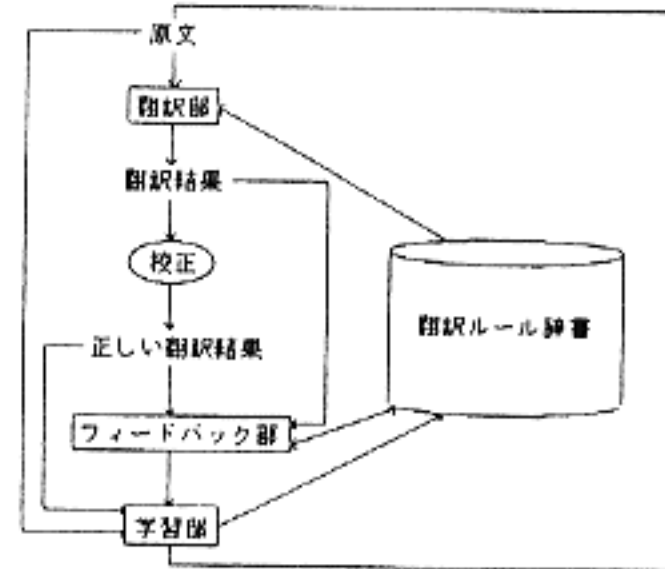


図1 GA-ILMTのシステム構成

3 状態遷移を用いた翻訳ルールの自動階層化と翻訳ルールの獲得

3.1 正翻訳の生成過程に基づく階層化

学習部では、正翻訳が導き出された過程を状態遷移として表し、翻訳ルールの自動階層化を行う。以下に処理過程を述べる。

(1) 正翻訳の生成過程における状態遷移を生成する。

状態遷移を生成する際に必要となる要素は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。

例として、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

3.2 確実性の高い翻訳ルールを用いた新たな翻訳ルールの獲得

状態遷移を導入することにより、個々の翻訳ルールの評価の精度が向上した。これまでは、翻訳ルールの組み合わせ結果に基づいた評価であったため、個々の翻訳ルールに着目した評価とはなっていなかった。それに対して、状態遷移に基づき翻訳ルールを評価することで、個々の翻訳ルールの組み合わせ過程に着目することが可能となり、翻訳ルールの評価の精度が向上した。そして、その評価結果を基に新たな翻訳ルールの生成を行う。以下にその具体例を示す。

(Makoto likes tea. ; 真はお茶が好きです。)
 (Makoto ; 真)
 ↓
 (@0 likes tea. ; @0 は お茶 が 好き です。)
 (tea ; お茶)
 ↓
 (@0 likes @1. ; @0 は @1 が 好き です。)

図3 翻訳ルールの獲得の具体例

図3では、(Makoto ; 真) と (tea ; お茶) が確実性の高い翻訳ルールとして、システムにより位置付けられた。そして、与えられた翻訳例にこれらが含まれる場合、その部分を抽象化することにより、新たな翻訳ルールが獲得される。

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Lesson1からLesson12までの翻訳例797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28.7%から36.0%に増加した。

4.2 考察

状態遷移を導入したことにより、正翻訳率は7.3ポイント向上した。これは、これまで正翻訳が存在しながら上位1位に選択されずに誤翻訳となっていたものが、上位1位に選択されるようになったためである。また、状態遷移により、翻訳ルールに対する評価の精度が向上したため、新たな良質な翻訳ルールの獲得が可能となり、正翻訳の向上をもたらしたと考えられる。こうした有効性は、新出語が少ない前半において特に見受けられた。Lesson1からLesson7までの翻訳例370組では、正翻訳率は35.4%から45.7%と10.3ポイントの向上であった。

5 おわりに

本稿では、GA-ILMTにおいて状態遷移を用いることの有効性について述べた。正翻訳率は7.3ポイント向上した。今後は、更なる学習能力の向上と共に、より実用的な学習型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] 野村浩郷(編): 言語処理と機械翻訳, 講談社, (1991).
- [2] 田中徳積(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).