適応型機械翻訳手法の概要

4B－7

荒木健治† 栃内香次††
北海学園大学工学部† 北海道大学工学部††

1．はじめに

機械翻訳システムは，その精度や翻訳品質の問題から広く一般に使われていないというのが現状である．しかし，この問題の原因の大半は現状の機械翻訳手法の多くが文脈情報や背景知識をその膨大さから完全には持つことも利用することもできないということが挙げられる．しかし，機械翻訳手法の現状のレベルでも対象分野を限定すると処理速度が早く高い精度でしかも良質な翻訳品質を得ることができる可能性がある．一方，対象分野を限定すると汎用性が低下し，少し対象分野から外れると全く翻訳できない．この問題に対して我々は，実例から翻訳ルールを帰納的に学習する機械翻訳システムを用いて予め種々の場面の翻訳ルールのセットを用意し，入力文と種々の場面の例文との類似性を用いて最適な翻訳ルールセットを自動的に選択することにより高い精度と汎用性を合わせ持つ機械翻訳手法の開発を試みている．このような手法を我々は適応型機械翻訳手法と呼んでいる．本稿では，本手法の概要とその有効性を確認するために行なった実験結果について述べる．

2．概要

本手法は，学習機能によりゆるやかに適応する従来の学習型機械翻訳手法と異なり，予め学習によって獲得された翻訳ルールを各場面別のセットにして蓄えておき，入力文の類似性を用いて場面を判断し，使用する翻訳ルールのセットを決定する．このことにより迅速な適応機能が実現できる．また，対応する場面が決定された後も学習機能により入力文が翻訳されるたび毎にその場面内の細かい範囲での適応が行なわれる．

本手法に基づく実験システムを作成する上で，用いられた学習型機械翻訳手法は，我々が従来より提案している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（GA-ILMT）[1] である．本手法を用いた理由は，事前に解析的な知識を一切与える必要がなく，完全に翻訳例のみより翻訳ルールを獲得できる手法であるためである．また，類似性の判定には，解析的な知識を多段階に用いることにより文章の類似性を判定する多段階類似性判定手法を用いている．なお，本稿における実験では英日の機械翻訳を対象として実験を行なった．

3．GA－ILMT

GA-ILMT の処理過程を図1に示す．

図1で英文が入力されると翻訳部で，それまでに獲得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．このようにして出力された翻訳結果に誤りが存在する場合には，人手により校正する．その校正済みの翻訳結果と翻訳結果を用いて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なう．次に学習部においては，与えられた原文とその日本語訳文からなる翻訳例に対して選択交配と突然変異を行なうことで多様な翻訳ルールを作成し，以後の翻訳で利用する．

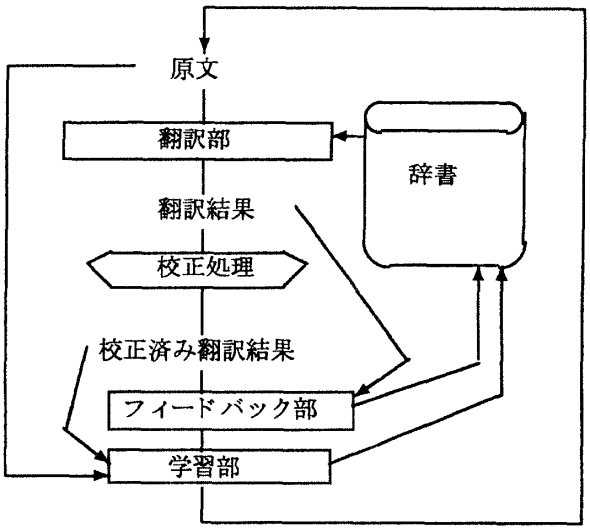

図 1: 処理過程

4．多段階類似性判定手法

本手法では入力された文章に対して，その文章の所属する分野を判定し，利用するルールセットを決定する．その際に類似性を判定するのが，多段階類似性判定手法である．多段階類似性判定手法では，類似性を対応関係が決定できるものが多さにより決定している．その対応関係決定の方法は，以下のように行なわれる．

（1）出現位置が同じで字面が一致
（2）出現位置が異なり字面が一致
（3）原形が一致
（4）同一の単語の訳語として存在
（5）上位概念が一致
（6）一語で決定済みの対応関係に挟まれている．
（7）品詞が一致

より上位の段階で決定された語は，それより以下の段階では処理対象とならない．したがって，本手法では確実性の高いものより順に処理を進めている．各段階で種々の知識を用いている．（3）の原形の検索には，

Outline of Adaptable Machine Translation Method
Kenji Araki† and Koji Tochinai††
Fac. of Engineering, Hokkai-Gakuen University†
Fac. of Engineering, Hokkaido University††

Princeton University で開発された WordNet, University of London で電子化された The oxford advanced learners dictionary of current English を用いている．また，（４）では英和辞書として [2] と [3, 4] の巻末に集められた Word list を用いた．（５）では，WordNet を用いている．（７）では，形態素解析を行なうのに，Massachusetts Institute of Technology と University of Pennsylvania で Eric Brill によって開発された tagger を用いている．また，以下の式（１）を用いてその類似性を評価する．

$$V = (2.0 \times a1 + 1.8 \times a2 + 1.6 \times a3 + 1.0 \times a4 + 0.8 \times a5 + 0.4 \times a6 + 0.2 \times a7 - 0.5 \times a8) \times \alpha \quad (1)$$

ここで，αは１００を比較する二文に出現する全単語数で割った値であり，$a1, a2, ..., a7$ はそれぞれ上記の（１）～（７）の各段階で対応関係が決定された組の数である．また，$a8$ は対応関係の決定できなかった語の数である．したがって，字面上全く同じ文では類似度は１００となる．

５．性能評価実験

５．１ 実験方法

旅行用英会話文を用いて実験を行なった．実験に用いた資料は [5, 6, 7, 8, 9, 10, 11, 12, 13, 14] から場面を機内と空港に限定し，そこで使われる訳文付きの会話文を機内での会話６１３例，空港での会話１０３９例を取り上げたものである．データの入力順は上述した順とし，始めから順に１番から１０番までの番号を付けて呼ぶこととする．学習型機械翻訳システムとしてはGA-ILMT とその比較のために A 社，B 社の製品 A，B を用いている．類似性の判定には多段階類似性判定法を用いている．

正誤の判定方法は，正しい訳文と字面上で完全に一致するか極めて近いもののみを正しい翻訳結果としている．表現が異なっても意味的に近ければ正しいとする評価方法もあるが，現在の機械翻訳が全く英語を知らない人も使うということを考慮するとこのような厳しい基準で評価すべきであると考えている．実験の手順は図1である．また，GA-ILMT の辞書の初期状態は全く空の状態より行なった．

５．２ 実験結果と考察

実験の結果，総合での正翻訳率は，機内での会話では GA-ILMT，A. B の順にそれぞれ 27.1%, 19.1%, 28.1% であったが，７番目のデータ以降の総合では，それぞれ 53.1%, 18.9%, 31.1%となった．また，同様に空港での会話では，総合が 30.7%, 25.9%, 30.1%であったが，７番目のデータ以降総合では 49.6%, 24.9%, 24.1%となった．すなわち，GA-ILMT が A, B に比べて最初は低かった精度が次第に上昇し，７番目のデータ以降は常に一番高い精度であった．このことは，学習機能により適応が進んだことを示している．また，翻訳の質という点でも GA-ILMT は良質な結果であった．例えば，機内のので会話で"Beef, please"を GA-ILMT は"牛肉料理をお願いします．"と訳しているが，A は"牛肉をどうか．"B は"牛肉をお願いします．"となった．機内という場面では，"Beef"は"牛肉料理"という意味なので，GA-ILMT が学習機能により文脈を捉えた翻訳を行なっていることがわかる．しかし，これは場面を限定した場合の結果であり，汎用性は当然 A,B の方が高い．

そこで，多段階類似性判定手法を用いて入力会話文の最適な場面を決定する実験を行なった．１から６番目のデータをサンプルデータとし適応が進んだ７番目のデータの場面を推定する実験を行なった．実験は，７番目のデータの各文がサンプル中で最も高い類似度を持つものの合計をそのサンプルとテキストの類似度とする方法で行なった．実験の結果，機内での会話のサンプルと機内での会話の７番目のデータ４４文の類似度は３９２３，空港での会話のサンプルと機内での会話の７番目のデータ４４文の類似度は２５２７で，この２つのサンプルでは正しく場面を推定できることが確認できた．

６．おわりに

GA-ILMT 及び多段階類似性判定手法を用いた適応型機械翻訳手法の概要及びその有効性を確認するために行なった実験結果について述べた．今後はさらに種々の場面のデータを用いて実験進める予定である．

参考文献

[1] 越前谷 他：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、情処論文誌，Vol.37, No,8, pp.1565-1579(1996).

[2] 久保：電脳宝船 英和・和英 電策辞典，技術評論社(1995).

[3] 長谷川 他：ONE WORLD English Cource 1 NEW EDITION, 教育出版 (1991).

[4] 太田 他：NEW HORIZON English Course 1, 東京書籍 (1991).

[5] 荒木 他：旅行英会話ポケットブック，日本文芸社(1995).

[6] 旅行会話研究会：海外旅行英会話，実業之日本社(1980).

[7] Gilbert：ケントのトラベル英会話，実業之日本社(1995).

[8] 石川 他：ひとり旅これで十分 英会話，実業之日本社 (1995).

[9] 前川：アメリカを自由に歩く 旅の米会話，池田書店 (1994).

[10] Read：困った時のトラベル英会話入門，日本文芸社 (1995).

[11] ブックメーカー：海外旅行かんたん英会話，池田書店 (1996).

[12] 甲斐：ひとり歩きの英語自遊自在，日本交通公社出版事業局 (1991).

[13] 地球の歩き方編集室：旅の会話集２ 米語／英語，ダイヤモンド・ビック社 (1993).

[14] 斉藤：六ヵ国語会話１ pocket interpreter，日本交通公社出版事業局 (1960).