

対訳コーパスを用いた分野適応型翻訳機構

7Q-7

中岩浩巳 山田節夫 池原悟

NTT 情報通信網研究所

1. はじめに

近年、様々な機械翻訳システムが実用化されているが、全ての文章に対して十分な訳文品質を得ているシステムはない。機械翻訳では、翻訳対象分野で頻出する言語現象の訳文品質を向上させるために、システムの辞書やルールを追加修正する等の翻訳対象分野への適応作業が必要となる。

従来、この適応化作業は、翻訳システムの内部処理を良く理解したエキスパートがその経験を活かして適切であると思われるルールや辞書を人手により追加修正することによって行なわれていた。ユーザーが行なえる分野適応作業は、システムが未知語と認定した語の訳語を登録するレベルにとどまっていた。人手による分野適応のための費用的、時間的コストは高く、継続的で大量な翻訳需要が有る場合でなければ、機械翻訳システムの性能をフルに活かせるまで分野適応作業を行なうことは困難であった。

本稿では、翻訳対象分野の対訳コーパスを用いて効率的かつ効果的に分野適応作業を自動的に行なう手法の提案をする。

2. 分野適応作業

ある翻訳対象分野における訳文品質を向上させるためには、ブラッシュアップ、カスタマイズの2種類の訳文品質向上策をとる必要がある。図1に機械翻訳システムがカバーする言語現象の範囲と一般的表現、ある対象分野の表現の関係を示す。実用的機械翻訳システム（ルールベース型機械翻訳システム）では、特に翻訳対象文を限定して設計しているものでなければ、一般的言語表現の一部をカバー

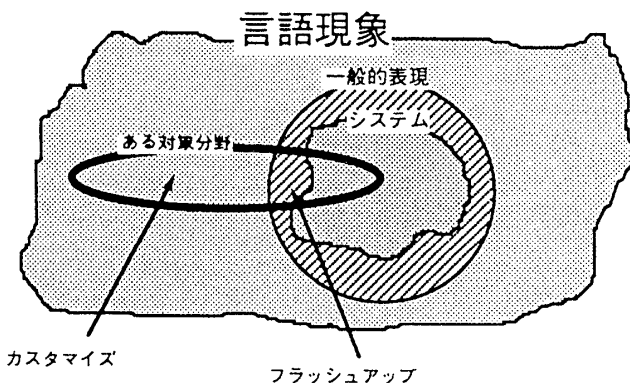


図1 機械翻訳システムのカバーする言語現象と分野適応

しているといえる。よって、システム開発者はこのシステムのカバー範囲を一般的表現全体に近づけるべく、機械翻訳システムが用いている辞書情報や文法などのルールを追加・修正する（ブラッシュアップ）。しかし、ある特定の対象分野の訳文品質を向上させる場合には、それがカバーする言語現象の範囲は、図1で示すとおり、一般的表現外のその分野固有の表現も用いられる。このような表現は、システムが常に使っている辞書やルールではなく、分野固有の用語を登録する辞書である専門用語辞書や、特殊な固有名詞等の翻訳対象文章のみで現われる語を登録するユーザー辞書を登録修正する（カスタマイズ）ことにより訳文品質が改善される。また、システム中に分野に応じて利用する文法などのルールを記述する方法も考えられる。

3. 分野適応作業の自動化

前章で述べたような分野適応作業の負荷を軽減するために、従来から様々な手法が提案されている。これらの手法は大きく次の2種類に分けられる。

その1つは、機械翻訳の際に行なわれる前編集、後編集作業の履歴を活用して、ルールや辞書を追加修正する方法である[1]。この手法は、翻訳システムがうまく解析・変換できなかった表現の訳文品質の改善が、蓄積された編集作業のノウハウを利用することによって行なわれる点で有効である。しかし、前編集、後編集は結局人手による作業となるため、負荷の軽減という観点では不十分である。

もう1つは、翻訳対象分野のコーパスを用いて訳文品質を自動的に向上させようという試みである。この手法としては、コーパスから翻訳対象分野向けのルールや辞書をルールベース型機械翻訳システムに追加するもの[2][3][4][5]、事例（用例）ベース型機械翻訳とルールベース型機械翻訳を混在させ事例ベース翻訳の枠組みでコーパスから翻訳対象分野向けの訳を得ようというもの[6][7]が提案されている。これらの手法は、ある程度の品質の辞書情報やルールが既に構築されている実用的ルールベース型機械翻訳システムに適用することを考えると、いくつかの問題がある。この翻訳システムで分野に応じた訳を生成させる場合、既存のルールや辞書情報を活かしつつ、その翻訳対象分野では不適切な訳を生成するルールや辞書情報のみについて追加、修正する必要がある。しかし、従来の手法では、分野適応のために用いるコーパスを翻訳システムの特性を意識して選択していないため、既存のルールや辞書に対して事例による悪影響が考慮できなかったり、この選択を事前に人間が行なう必要があるため人手による負荷の軽減が不十分である。また、分野適応を行なう場合には、対訳辞書情

報を修正するのみで解決するものから、文法を修正するものまで様々な問題が混在するが、これらの手法にはこのうちの一部の言語現象しか考慮していない。

以上のような、実用的ルールベース型機械翻訳システムにおける従来手法では解決できない分野適応の自動化のための必要条件をまとめると以下の2点になる。

- (1) 正しく訳せない事例を自動的に選択する
 - 翻訳システムの既存ルールを活かし、事例を有効に活用するため
- (2) 問題の性質に応じて事例を自動的に分類する
 - 最も効率のかつ効果的な方法で訳文品質の問題の解決するため

最近、(1)に関して、翻訳パターンを自動抽出の際に、機械翻訳の結果と正しい翻訳結果を見比べて差分を見つけ、現在の翻訳システムにとって有効な翻訳パターンを抽出する方法が提案されているが[8]、実用的機械翻訳システムへの適用を考えると、実現容易性の面で問題がある。

4. 分野適応型翻訳機構

本章では、前章の条件を満たす分野適応型翻訳機構について述べる。ここで提案する機構は、大きく2つの部分に分かれる。1つは、原言語と目的言語の対訳コーパスをもとに、原言語の文を翻訳システムで翻訳した訳(機械訳)とコーパスの模範訳(理想訳)を比較して、問題事例(正しく訳せない事例)を抽出し、効率的な解決法別に事例を分類する部分である。この処理の実現には、理想訳と機械

訳の構文木を比較するための距離を定義し、その距離に応じて事例を分類する手法が考えられる[9]。もう1つは、分類された問題事例を分析し、分野適応のための辞書やルールとして登録する処理である。これは、正しい訳が生成出来ない原因を分析し、翻訳対象分野向けの効果的なルールや辞書情報を自動登録修正する処理である。この2種類の処理を繰り返し、翻訳対象分野向けのルールや辞書を蓄積していくことで、翻訳対象分野の訳文品質が向上する。

この分野適応型翻訳機構を、我々が開発中の日英翻訳システムALT-J/E[10]に適用したのが図2である。本システムは、単語レベルで変換するための辞書である単語辞書、用言を中心とした単文構造を変換するための辞書であるパターン対辞書、複文、重文、埋め込み文などより複雑な構造を変換するためのルールを持っているため、上記3種類に対応した分野適応型辞書・ルールを用意し、解決法別に分野適応化を行なう。この図の流れで、分野適応型辞書情報・ルールを抽出し、再度これら辞書・ルールを用いて翻訳を繰り返すことにより、最適なルールや辞書が抽出され、訳文の変化が定常状態になった時点で、この分野適応化の学習サイクルは終了する。

5. おわりに

本稿では、原言語と目的言語の対訳コーパスを用いた実用的機械翻訳システムの分野適応を自動的に行なう手法を提案した。現在、単語レベルの分野適応化処理の実現を行なっており、実験によって翻訳不良表現の検出とそれに対する対訳辞書作成が自動化できる見通しを得た。今後は、句や文のレベルでも本手法の有効性を検証していく。

<参考文献>

- [1] M.Miura, M.Hirata and N.Hoshino: Learning Mechanism in Machine Translation System "PIVOT", Proc of COLING-92, pp. 693-699, 1992
- [2] 野美山: 目的言語の知識を用いた訳語選択とその学習性, 自然言語処理研究会, NL86-8, 1991
- [3] S.Doï and K.Muraki: Translation of Ambiguity Resolution based on Text Corpora of Source and Target Languages, Proc of COLING-92, pp.525-531, 1992
- [4] H.Kaji, Y.Kida and Y.Morimoto: Learning Translation Templates from Bilingual Text, Proc of COLING-92, pp.672-678, 1992
- [5] T.Utsuro, Y.Matsumoto and M.Nagao: Lexical Knowledge Acquisition from Bilingual Corpora, Proc of COLING-92, pp-581-587, 1992
- [6] 古瀬, 隅田, 飯田: 変換手動型機械翻訳の実現手法, 自然言語処理研究会, NL-80-8, 1990
- [7] H.Watanabe: A Similarity-Driven Transfer System, Proc of COLING-92, pp.770-776, 1992
- [8] H.Watanabe: A Method for Extracting Translation Patterns from Translation Examples, Proc of TMI-93, pp-292-301, 1993
- [9] 山田, 中岩, 小倉, 池原: 分野適応型翻訳機構のための翻訳不良表現の自動抽出, 情報処理学会第48回全国大会, 7Q-8, 1994
- [10] 池原, 宮崎, 白井, 林: 言語における話者の認識と多段翻訳方式, 情報処理学会論文誌, Vol.28, No.12, pp.1269-1279, 1978

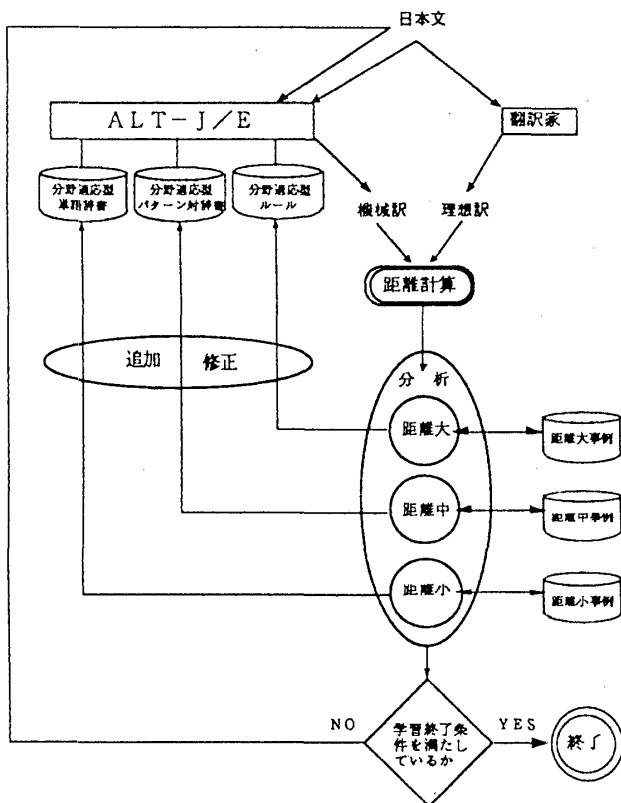


図2 分野適応型翻訳機構の実現例