

単文化処理による主語補完を適用した翻訳品質の向上について

高橋 彩子

岸 義樹

茨城大学大学院理工学研究科

茨城大学工学部情報工学科

1 はじめに

これまで、多くの機械翻訳ソフトが研究開発され、我々の情報処理の効率を上げる位置付けとして認知されるに至っている。しかし実際の利用の際、入力文の前編集や出力に対する後編集作業がユーザにとって大きな負担となっている。そこで本研究では、日英機械翻訳における日本語原文の整形（英語に訳しやすくするための整形）に単文化処理を利用した主語省略補完を適用することで、日英機械翻訳の効率を上げることが目的として評価実験を行った。

2 数字的裏付

2.1 文整形の有効性

機械翻訳への入力原文に対する適切な整形が翻訳の精度向上に及ぼす効果と、その整形方法に関する評価実験について報告する。

過去の研究報告（[1][2]）において、日英翻訳の訳質を向上させるために必要とされる日本語原文側での修正項目として、以下の4点が挙げられている。

- ・長文は短文に分割してから翻訳する。
- ・省略が起こっている場合は、補完してから翻訳する。
- ・漢字での表記が自然である言葉がひらがなで表記されていた場合は、漢字に変換してから翻訳する。
- ・日本語特有の慣用句・言い回しを持つ文は、簡潔な表現に変換してから翻訳する。

上記の4項目について、5種類の機械翻訳システムと実際の文章（平成13年度労働経済白書）を用いて、人手で文の整形を行い、翻訳精度を

検証する実験を行った。結果を表1に示す。

表1：4種類の文に対する整形前・後の翻訳結果

		MT1	MT2	MT3	MT4	MT5
長文分割	分割前誤訳率	40.0%	70.0%	60.0%	90.0%	60.0%
	分割前平均誤訳率	64.0%				
	分割後誤訳率	30.0%	30.0%	40.0%	90.0%	50.0%
	分割後平均誤訳率	48.0%				
省略補完	補完前誤訳率	80.0%	40.0%	60.0%	80.0%	90.0%
	補完前平均誤訳率	70.0%				
	補完後誤訳率	40.0%	20.0%	60.0%	60.0%	40.0%
	補完後平均誤訳率	44.0%				
漢字表記	変換前誤訳率	40.0%	60.0%	30.0%	50.0%	60.0%
	変換前平均誤訳率	48.0%				
	変換後誤訳率	20.0%	20.0%	0.0%	20.0%	40.0%
	変換後平均誤訳率	20.0%				
慣用句変換	変換前誤訳率	60.0%	10.0%	70.0%	60.0%	30.0%
	変換前平均誤訳率	46.0%				
	変換後誤訳率	30.0%	10.0%	40.0%	30.0%	10.0%
	変換後平均誤訳率	24.0%				
整形前平均誤訳率		57.0%				
整形後平均誤訳率		34.0%				

この実験結果から、現在の翻訳システムでも、入力原文に対する整形は有効であると言える。特に、省略が起こっている文に対する誤訳率が高く、省略補完処理の有効性が期待できる。

2.2 主語補完の有効性

省略補完の有効性を踏まえ、主語補完の効果

Improvement of translation quality by subject supplementation using sentence simplification

Ayako Takahashi, Yoshiki Kishi
Ibaraki University
4-12-1 Nakanarusawa, Hitachi, Ibaraki, 316-8511, JAPAN

について述べる。

どの種類の省略補完が訳質向上に資するかを調べるために、省略された語句を人手によって補完した後、機械翻訳を行い、訳質の変化を検証した。具体的には、複数の文章（一文章あたりの文数は30~50文のものとする）を機械翻訳にかけ、その中で、省略補完によって訳質が改善したものを、以下の4種類に分類した。各分類の後に書かれている数字は、省略補完によって訳質が改善する文数に対する割合である。

- (1) 省略を補完するためのヒントが同一文中に出現していない場合で
 - (1-a) 省略語が主語であり、それを補完した際に訳質が改善した文：20.4%
 - (1-b) 省略語は主語以外のもので、それを補完した際に訳質が改善した文：21.0%
- (2) 省略を補完するためのヒントが同一文中に出現している場合で
 - (2-a) 省略語が主語であり、それを補完した際に訳質が改善した文：23.0%
 - (2-b) 省略語は主語以外のもので、それを補完した際に訳質が改善した文：50.1%

これらの結果から、主語の補完が機械翻訳の訳質向上に効果的であることがわかる。

3 単文化処理の利用

3.1 単文化

単文化とは、一文中に述語が複数存在している文に対して、それらの述語の主語を特定し、一つの文に述語一つである文に分割する処理のことである。本研究では、この単文化処理のプログラムを作成し、省略補完に利用した。

例)

私は学校に行き、研究に励みます。
主語 述語1 述語2

私は学校に行きます。私は研究に励みます。
主語 述語 主語 述語

上記の例からわかるように、単文化処理は主語の補完処理を内包している。

3.2 実験

この実験ではまず、日英機械翻訳において、日本語原文での主語の省略に対応した翻訳がで

きない文に単文化処理を施す。その結果、検出される省略された主語を補完することで、訳質改善効果を検証した。

文数50の文章を実験対象とした際、省略補完によって訳質の改善が想定される文は19文存在し、その内、本システムで検出できる省略補完を行うことで完全な英訳を得ることができるケースが5文あった。また、完全な英訳ではないが、省略補完を行わない場合よりも訳質の改善が見られるケースが4文あった。

例)

「雇用者は増加したものの、自営業や家族従業者が大きく減少したため、これらを含む就業者全体の増加は遅れ、また弱いものとなった。」

誤訳の原因となる省略された主語を補完

「雇用者は増加したものの、自営業や家族従業者が大きく減少したため、これらを含む就業者全体の増加は遅れ、また増加は弱いものとなった。」

再度機械翻訳

「Though the number of employers increased, an increase of the entire employed including these was delayed and an increase became weak because the independent enterprise and the family employed person decreased greatly.」(正訳)

4 おわりに

省略補完が日英機械翻訳の訳質を向上させること、及び、その中でも主語の補完が有効であることを確認した。これらの結果に基づき、単文化処理を利用した主語補完を日本語原文に施すことによる、日英機械翻訳の訳質改善効果について実験と検証を行い、その効果を確認した。今後はより多くの文章に対する実験を進め、単文化処理を組み込んだ効率的な日英機械翻訳前処理システムの構築を検討していきたい。

参考文献

- [1] 白井・池原・河岡・中村「日英機械翻訳における原文自動書き換え型翻訳方式とその効果」情報処理学会論文誌 Vol.36 No.1 pp12-21 Jan 1995
- [2] 木村・野村・平川「日英機械翻訳における日本語分割処理について」情報処理学会 自然言語処理研究会報告書 96-8, pp57-64 1993
- [3] 中原・杉浦・岸「意味構造抽出によるテキストマイニングと知識利用」FIT 2002 E-15