

単文化処理における連体修飾語の一部省略を適用した 日英翻訳向上へのアプローチ

馬場先 智美 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

今まで多くの機械翻訳ソフトが、我々の情報処理の効率を上げてきた。しかし実際の利用となると、入力文の前編集や出力文に対する後編集が必要となり、ユーザに多大な負担となっていた。そこで本研究では、日英翻訳の効率を上げることを目的とした方法を提案する。その方法は日英機械翻訳における日本語原文の整形において、まず、原文を主題・主語と述語の組合せからなる単文の集合に分解を試みる単文化処理を施し、この処理によって判明する省略された主題・主語を補う主語省略補完を行う([1])。同時に、この単文化処理において見いだされる連体修飾機能を有する単文について、連体修飾機能を含む単文中の一節に関して、その一部省略を適用することで日本語原文を再構成した後で翻訳を行う。提案方法を用いた翻訳実験を行い、翻訳訳質向上の可能性を検証した。

2 文整形作業と訳質の向上について

2.1 前処理としての文整形の有効性

機械翻訳への前処理として、過去の研究報告([2][3])において報告された効果的な文整形の方法を以下に示す。

- ・ 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る場合は、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。

以上の項目の他に、更に各種文章を翻訳した結果、連体修飾語の扱いで誤訳が多いので、長文分割後の処理として、連体修飾語の一部省略を導入した。その方法は、長文分割が行われた文より連体修飾語を節ごとに一番前から省略する支援ツールを実行し、その結果に機械翻訳を行う。

以上の 5 項目について、4 種類の機械翻訳システムと実際の文章(平成 11~13 年度版 労働経済白書、平成 13 年度版 情報通信白書、平成 13 年度版 中小企業白書)を用いて、人手で文整形を行

表 1: 5 種類の文に対する整形前・後の誤訳率

誤訳率		System 1	System 2	System 3	System 4
長文分割	分割前	56%	60%	55%	62%
	分割前平均	58%			
	分割後	33%	40%	37%	37%
	分割後平均	37%			
省略補完	補完前	47%	47%	36%	52%
	補完前平均	46%			
	補完後	52%	47%	57%	68%
	補完後平均	56%			
漢字表記	変換前	90%	90%	72%	72%
	変換前平均	81%			
	変換後	81%	72%	81%	81%
	変換後平均	79%			
慣用句変換	変換前	55%	44%	55%	55%
	変換前平均	52%			
	変換後	55%	77%	55%	44%
	変換後平均	58%			
長文連体分割修飾省略	省略前	43%	57%	50%	56%
	省略前平均	51%			
	省略後	39%	51%	40%	50%
	省略後平均	45%			
整形前平均		55%			
整形後平均		39%			

Approach to Improvement of Japanese-English Translation that Applying Partial Omission of Adnominals in Sentence Simplification

Satomi Babasaki, Yoshiki Kishi

Ibaraki University

4-12-1 Nakanarusawa, Hitachi, Ibaraki, 316-8511, JAPAN

い、翻訳精度を検証する実験を行った。複数の文章（一文章あたりの文^{*1}が30～50）を日英翻訳にかけた。結果を表1に示す。

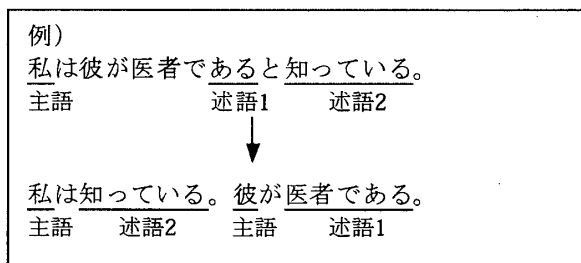
この実験結果から、現在の翻訳システムに入力文の整形が有効であると言える。特に長文分割後の連体修飾語の一部省略に対して誤訳率の減少が見られた。

これにより、入力文の文整形に長文分割後における連体修飾語一部省略の有効性が期待できる。

3 連体修飾語一部省略支援方法について

3.1 単文化処理の利用

単文化とは、複文を一文の中に述語が一つ存在するような文に分割する処理のことである。本研究では、連体修飾語の一部省略を行う前処理として、単文化処理のプログラムを利用した。



上記の例のように、単文化処理は主語補完も含んでいる。

3.2 支援ツールの作成と実験

作成した対話的支援ツールは以下の特徴を持つ。日英機械翻訳機において、日本語原文より正しい翻訳文が出力されない日本語文に単文化処理を施し、主語補完が必要ならば行う。さらに単文化処理を行っても正しい翻訳文が得られない文に対して、連体修飾語一部省略処理を行う。実験ではこのツールの結果得られる訳文に対して、訳質改善効果を検証した。

65文章（一文章あたりの文が30～50）を実験対象とし、連体修飾語一部省略によって訳質の改善が想定される文章は32存在し、その内システムによって連体修飾語一部省略を行うことで分かり易い英訳を得ることができる場合が9文章あった。また、分かり易い英訳ではないものの、連体修飾語一部省略を行わない場合より、行った場合に訳質が改善された文章が20あった。

例)

「IEEEが11の賞(注)について1991年から2000年の10年間の受賞者数をみると」

↓ 機械翻訳

「When IEEE watches the number of the prize winners of ten years of 2000 about 11 prizes (a note) from 1991」(誤訳)

↓ 連体修飾語を省略する(11の、2000年の)

「IEEEが賞(注)について1991年から10年間の受賞者数をみると」

↓ 再度機械翻訳

「When IEEE watches the number of the prize winners of ten years about a prize (a note) from 1991」(正訳)

今回は、例文の斜体部分を省略した。まず、「11の」を省略したが、訳文が改善されなかったため、次の「2000年の」も省略することで、分かり易い英訳を得ることができた。

4 おわりに

手作業による連体修飾語一部省略が日英機械翻訳システムの訳質を向上させることを確認した。これらの結果に基づき、日英機械翻訳の前処理として、日本語原文に対する単文化処理を利用した主語補完を伴う、連体修飾語一部省略支援方法を提案し、その有効性を実験によって、検証し、確認した。今後はより多くの文章に対する実験を行い、単文化処理を組み込んだ効率的な日英機械翻訳支援方法を検討していきたい。

参考文献

- [1] 高橋、岸：「単文化処理による主語補完を適用した翻訳訳質の向上について」,情報処理学会第69回全国大会講演論文集,3Q-7,pp.2-389~2-390(2007)
- [2] 白井、池原、河岡、中村：「日英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,pp.12-21,(1995)
- [3] 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,pp.57-64(1993)

*1 ここで言う文とは、ほぼ文節のことである。