

英語の主節と副詞節の順序を保持した和訳のための 前編集と後編集

柴田 誠 南條 浩輝 吉見 毅彦

龍谷大学 理工学部 情報メディア学科

e-mail: shibata@nlp.i.ryukoku.ac.jp

1 はじめに

従来の英日機械翻訳システム (以下, MT) では, 主節の後に副詞節がくる英文を日本語に翻訳すると, 節の順序が逆転することが多い. このとき, 例えば英文 (E1) のように, 主節中の名詞句を先行詞候補とする代名詞が副詞節中に存在する場合, 代名詞の訳語が先行詞の訳語より先に現れ, 代名詞と先行詞の照応関係が読み取りにくい文 (J1) のようになってしまう.

(E1) I bought a book because it was cheap.

(J1) それが安かったので, 私は 本を買った.

このような問題への対策として, 1) 先行詞の名詞句と代名詞を入れ替える方法と, 2) 主節と副詞節の順序を保持したまま翻訳する方法がある. 方法 1) は, 他表現への置き換えやゼロ代名詞化などの研究 [1][5] に基づいて, 英文 (E1) を次の文 (J1') のように翻訳することをめざしたものである.

(J1') 本が安かったので, 私は それを買った.

このような翻訳を実現するためには, 代名詞の先行詞を同定する照応解析を行う必要がある. しかし, 現状では照応解析の精度は十分ではないため, この方法は実用的とは言えない.

方法 2) については, 原文の節の順序を入れ替えないう和文生成を実現したという報告 [2] があり, この方法によると, 英文 (E1) は次の文 (J1'') のように翻訳される.

(J1'') 私は 本を買った. それが安かったから.

しかし, この先行研究の手法は特定の MT の日本語生成部として実現されたもので, 手法の詳細が公開されていないこともあり, 別の MT に適用できるかどうかは不明である. また, この手法で生成された複文の

和訳は文末が用言の終止形でないことがあるため, 意味は通じるが文体としては適切でないところがある.

本研究では, 方法 2) のように, 英文の主節と副詞節の順序を保持した和訳を, 前編集と後編集を行うことによって実現する. これにより, 文 (J1'') のような照応関係が理解しやすい和文を生成することができる.

(J1'') 私は本を買った. なぜなら, それは安かったからだ.

本研究では, 前編集と後編集を MT から独立させることで, 特定の MT に依存せずに複数の MT へ適用できるようにし, 先行研究 [2] よりも手法の汎用性を向上させる. 本稿では, このような方法において, 形態素解析や構文解析, 意味解析などを行わずに翻訳品質をどの程度改善できるかを検証する.

2 主節と副詞節の順序を保持した和訳のための自動編集

2.1 節の順序を保持するための処理の概要

提案手法の処理の流れを図 1 に示す. なお, 本研究では, 現在のところ, 処理対象を主節と副詞節の 2 つの節がこの順に出現する複文に限定し, さらに, 従属接続詞も before, because, when に限定している.

前編集 まず, 入力文に対して前・後編集を行うかどうかを判定する. 次に, 前・後編集を行うと判定された文を主節, 従属接続詞, 副詞節の 3 つに分割して出力する. 前・後編集を行わないと判定された文はそのまま出力する.

機械翻訳 前編集部から出力された主節と副詞節をそれぞれ既存の MT で翻訳する.

後編集 MT から出力された和文を読み込み, これの原文が編集対象ならば, 前編集で分割された主節

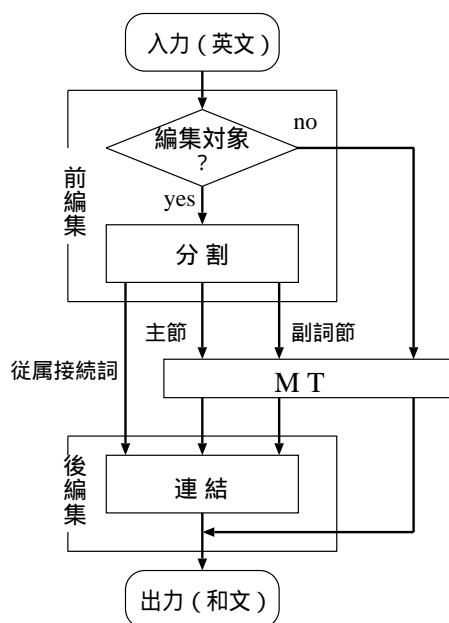


図 1: 提案手法の処理の流れ

表 1: 英文のデフォルト訳出形式

従属接続詞	訳出形式
before	[主節] . その後 , [副詞節] .
because	[主節] . なぜなら , [副詞節] からだ .
when	[主節] . そのとき , [副詞節] .

と副詞節を表 1 から表 4 の訳出形式に従って連結することにより、原文の節の順序に沿った和文を生成する。

前編集部での判定は、前・後編集によって翻訳品質が低下する可能性がある入力文に対しては、処理を行わないようにするためのものである。判定条件については、2.2.1 節で述べる。

後編集部で参照する表 1 の訳出形式に従って処理を行うと不自然になったり、英文とは意味が異なった和文となる場合がある。このような場合に対処するために、従属接続詞ごとに訳出形式を設定した。

2.2 従属接続詞ごとの訳出形式

2.2.1 before の訳出形式

英文 (E2) のように、before の直前に “two weeks” などの時間表現が存在する文に対して前・後編集を行うと文 (J2) のような和文が生成されてしまう。このため、現在のところ、before の直前に時間表現がする

表 2: before の個別訳出形式

従属接続詞	訳出形式
must ... before	[主節] . そうすれば , [副詞節] .
should ... before	[主節] . そうすれば , [副詞節] .

文は処理対象としないという判定条件を前編集部に設け、従来の MT による翻訳をそのまま出力する。

(E2) Taiwan’s Commission of National Corporations has rejected Chinese Petroleum Corp’s proposed floating price system two weeks before it was to take effect.

(J2) 台湾の国内企業の委員会は、2 週 中国の石油株式会社の提案された浮動価格システムを却下した。その後、それは効力を発する予定である。

英文 (E3) のように主節の助動詞が must や should である文は、主節の動作が副詞節の動作のための条件を表していると解釈できることがある。このような場合には、表 1 の訳出形式ではなく、表 2 の訳出形式に従う。これにより文 (J3’) のようなより適切な和文を生成することができる。

(E3) The executive director must obtain Congressional permission before he can vote.

(J3) 専務は米国議会の許可を得るべきである。その後、彼は投票できる。

(J3’) 専務は米国議会の許可を得るべきである。そうすれば、彼は投票できる。

2.2.2 because の訳出形式

because の直前に mainly や partly などの副詞が存在し、これが because 節を修飾しているような文 (E4) を前編集すると、この副詞が主節の動詞を修飾する形で文が分割されてしまう。これを翻訳し、後編集すると、文 (J4) のような和文になり、英文の意味と異なってしまふ。このような文に対しては、表 3 の訳出形式に従う。なお、表中の [主節] は、英文から mainly などの副詞を削除してから翻訳した主節である。これにより、原文と同じ意味を持つ文 (J4’) を生成することができる。

表 3: because の個別訳出形式

従属接続詞	訳出形式
mainly because	[主節] . 主な理由は, [副詞節] からだ .
partly because	[主節] . その理由の一つは, [副詞節] からだ .

(E4) Prices went up mainly because it followed Japanese prices.

(J4) 価格は 主に 上がった . なぜなら, それは日本の価格に続いたからだ .

(J4') 価格は上がった . 主な理由は, それは日本の価格に続いたからだ .

表 1 の訳語「からだ」は, 副詞節の末尾が「だろう」か「であろう」である場合, このまま接続すると文 (J5) のような不適切な和文となる . このため, 後編集では, 文 (J5') のように「だろう」「であろう」を「かもしれない」に書き換える .

(E5) We welcomed fair market competition because it would spur the republic to improve and stay ahead.

(J5) 我々は公正な市場競争を歓迎した . なぜなら, それは共和国が向上し, 有利な立場を保つのに拍車をかける だろう からだ .

(J5') 我々は公正な市場競争を歓迎した . なぜなら, それは共和国が向上し, 有利な立場を保つのに拍車をかける かもしれない からだ .

2.2.3 when の訳出形式

when の直前に particularly などの副詞が存在し, これが when 節を修飾している英文 (E6) のような文に対しては because の場合と同様に, 表 4 の訳出形式に従う . なお, 表中の [主節] は, 英文から particularly などの副詞を削除してから翻訳した主節である . これにより, 原文と異なる意味を持つ文 (J6) ではなく, 同じ意味を持つ文 (J6') を生成することができる .

(E6) Even monetary policy cannot be indifferent to this situation, particularly when it does not need to have any worries about inflation.

表 4: when の個別訳出形式

従属接続詞	訳出形式
particularly when	[主節] . 特にそのとき, [副詞節] .

(J6) 財政政策さえ 特に この状況に無関心であるはずがない . そのとき, それはインフレーションについての少しの心配も必要としない .

(J6') 財政政策さえこの状況に無関心であるはずがない . 特にそのとき, それはインフレーションについての少しの心配も必要としない .

3 実験と考察

3.1 実験方法

提案手法の有効性を検証するために, 提案手法を適用することによって翻訳品質がどの程度向上するかと, 複数の MT において翻訳品質が向上するか見る . 実験には, 3 つの市販 MT (MT-A, MT-B, MT-C) を用い, 提案手法を適用して得られた和文に対する翻訳品質評価値と適用しない場合の評価値を比較する . 翻訳品質の評価は, 流暢さと適切さについて 5 段階で, さらに代名詞の先行詞のわかりやすさについて 3 段階で著者 1 名が行った .

従属接続詞ごとの個別訳出形式を検討するための開発用データとして, 本研究の現在の処理対象制限 (主節と before, because, when に導かれる副詞節の 2 つ節から成る複文) を満たす英文をロイターコーパス [3] から抽出した . また, テスト用データとして, 読売新聞 2000 データ集 CD [4] から処理対象制限を満たす英文 (従属接続詞が before, because, when である文それぞれ 42 文, 55 文, 66 文) を抽出した . 開発用データを用いて提案手法の改良を繰り返し行った後, テスト用データを用いて評価実験を行った .

3.2 結果と考察

各従属接続詞について, 提案手法を適用しない場合に比べて, 適用した場合に翻訳品質がどのように変化したかを表 5 に示す . 全体的に見ると, 適切さは 31% ~ 46% 程度の文でしか向上していないが, 流暢さは 39% ~ 67% 程度の文で向上している . また, 代名詞の先行詞のわかりやすさについては, 64% ~ 78% 程度

表 5: 提案手法適用による翻訳品質の変化

		MT-A			MT-B			MT-C		
		向上	同等	低下	向上	同等	低下	向上	同等	低下
before	流暢さ	59.5%	33.3%	7.1%	42.9%	38.1%	19.0%	66.7%	26.2%	7.1%
	適切さ	31.0%	59.5%	9.5%	31.0%	45.2%	23.8%	38.1%	67.6%	14.3%
	代名詞	64.3%	35.7%	0.0%	69.0%	31.0%	0.0%	78.6%	21.4%	0.0%
because	流暢さ	58.2%	34.5%	7.3%	52.7%	30.9%	16.4%	60.0%	32.7%	7.3%
	適切さ	36.4%	60.0%	3.6%	43.6%	52.7%	3.6%	45.5%	47.3%	7.3%
	代名詞	72.7%	27.3%	0.0%	78.2%	18.2%	3.6%	70.9%	29.1%	0.0%
when	流暢さ	47.0%	37.9%	15.2%	39.4%	42.4%	18.2%	50.0%	39.4%	10.6%
	適切さ	31.8%	50.0%	18.2%	31.8%	54.5%	13.6%	31.8%	54.5%	13.6%
	代名詞	72.7%	25.8%	1.5%	63.6%	27.3%	9.1%	77.3%	22.7%	0.0%

の文で向上している。各従属接続詞の文の間で比較すると、従属接続詞が before の文と because の文は、従属接続詞が when の文に比べて翻訳品質が向上した文が多い。

提案手法による翻訳品質改善に、使用した MT による違いがあるかどうかを見ると、従属接続詞が before の文の場合、MT-B において流暢さの向上率が他の 2 つ MT に比べて低いが、それ以外の場合においては、3 つの MT の間に大きな違いはないと考えられる。従って、提案手法はある程度の汎用性を持つ手法であるといえる。

提案手法を適用した場合の翻訳品質が、適用しなかった場合と比べて低下した原因として、MT によって英語の代名詞が日本語でゼロ代名詞化されていたために、提案手法を適用した場合よりも流暢さが高かったことが挙げられる。特にこのような翻訳は MT-B に多く見られた。このため、他の MT よりも翻訳品質が向上した文が少なかったと考えられる。また、従属接続詞が when の文では、流暢さと適切さの向上率が他の従属接続詞の文よりも低かった理由として、when の持つ意味が「時」のほかに複数あり、これらの訳し分けが不十分であったことが挙げられる。従属接続詞 when については、さらに訳出方式を改善する必要がある。

4 おわりに

主節の後に副詞節がくる英文の順序を保持した和文生成のための前・後編集を提案した。そして、この提案手法を従来の MT に適用した場合と適用しない場

合での翻訳品質を比較し、提案手法の有用性を検証した。その結果、代名詞の先行詞を理解しやすく、流暢さと適切さの高い和文の生成に効果があることを実証することができた。さらに、複数の MT で効果が見られたことから、提案手法は汎用性のある手法といえる。

参考文献

- [1] 荻野紫穂, 那須側哲哉. 表層的な文脈情報を用いた自然な文生成の試み. 情報処理学会第 52 回全国大会講演論文集 平成 8 年前期, No. 3, pp. 97–98, 1996.
- [2] 佐田いち子, 九津見毅, 日野ちなみ, 関谷正明. 英文語順に準拠した日本語生成. 言語処理学会第 4 回年次大会発表論文, pp. 670–672, 1998.
- [3] Masao Utiyama and Hitoshi Isahara. Reliable Measures for Aligning Japanese-English News Articles and Sentences. *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pp. 72–79, 2003.
- [4] 読売新聞社. 読売新聞 2000 データ集 disc.1, 2000. 日外アソシエーツ.
- [5] 吉見毅彦. 英日機械翻訳における代名詞翻訳の改良. 自然言語処理, Vol. 8, No. 3, pp. 87–106, 2001.