

## 統計的前編集のための対訳コーパスからの 学習データの自動生成

山本 祐司<sup>†1</sup> 南條 浩輝 吉見 毅彦

機械翻訳システムを用いて自然性が高いテキストを翻訳すると、適当な結果が得られないことがある。逆に、自然性の低いテキストを機械翻訳システムに入力した方が好ましい翻訳ができることがある。このことから、自然な原文を機械翻訳しやすい文に前編集することで翻訳品質の向上が期待できる。そのため、本研究では、対訳コーパスを用いて統計的前編集モデルの自動学習を行う。具体的には、ある文とその対訳文を機械翻訳して得られる文のペアから、自然な文を翻訳しやすい文に変換するための学習データを自動作成し、自然性の低いテキスト、例えば直訳調の文に対してはうまく翻訳できることがある。統計的機械翻訳の枠組みに基づいて自動変換を行う方法について述べる。

### Automatic Generation of Training Data from Parallel Corpus for Statistical Pre-editing

YUJI YAMAMOTO,<sup>†1</sup> HIROAKI NANJO  
and TAKEHIKO YOSHIMI

Most of MT systems do not work well for natural and fluent texts. On the other hand, for easily machine translatable texts, MT systems work well. Therefore, for an improvement of translation quality, rewriting a fluent text to a machine translatable text before translation is promising. Based on the background, in this paper, we study a statistical pre-editing method. Specifically, first, we show an automatic generation of training data for statistical models from Japanese to English (J-E) parallel corpus, and then, describe a pre-editing method based on a statistical translation framework.

### 1. はじめに

現在の機械翻訳システムでは、自然性が高いテキストに対しては望ましい翻訳結果が得られないのに対して、自然性の低いテキスト、例えば直訳調の文に対してはうまく翻訳できることがある。すなわち、自然な原文を直訳調のような翻訳しやすい文に書き換える（前編集する）ことで翻訳品質の向上が期待できる。このような前編集の効果は、お互いに近いヨーロッパの言語間の翻訳においてよりも、日本語と英語などの大きく異なる言語間の翻訳において大きいと考えられる。また、使用人口が少ない言語などでは研究が十分に行われていないために機械翻訳自体の性能が低く、このような言語の翻訳においても前編集の効果は大きいと考えられる。機械翻訳の品質を向上させるためには、機械翻訳システム自体の性能向上を行うことが本質的であるが、そのためには翻訳対象の言語ペアについての豊富な知識および大量のデータが必要となる。前編集を考えた場合は、翻訳元の言語に関する知識やデータおよび機械翻訳システムの使用方法（入力インターフェース）を習得していれば十分であり、翻訳品質の向上が比較的得やすいという利点がある。

これまでに、翻訳品質の向上を目的として、自然な原文を翻訳しやすい文に前編集する研究は多く行われているが<sup>1)2)</sup>、それらは主にルールベースのものである。その上、変換規則獲得のための自然な文と翻訳しやすい文のペアを用意するコストが大きいという問題がある<sup>3)4)</sup>。これに対し、我々は統計的前編集およびそれに用いる統計モデルの学習データを対訳コーパスから自動生成する手法を提案してきた<sup>3)</sup>。具体的には、ある文とその対訳文を機械翻訳して得られるペアから、自然な文を翻訳しやすい文に変換するための学習データを自動生成する手法を提案し、その効果を確認した<sup>4)</sup>。ただし、統計モデルの学習のためのデータがそもそも少ないこと、自然な文と翻訳しやすい文のペアの選択基準について十分な検討が行われていないこと、実験において一つの機械翻訳システムしか使用されておらず提案手法の頑健性が明らかでないという問題があった。

このような背景に基づき、本論文では、これらの問題に対処を行う。具体的には、複数の機械翻訳システムを用いて実験を行い、学習データの問題と頑健性の問題に対応する。さらに、自然な文と翻訳しやすい文のペアのなかから、学習データとして適切なものを選択する基準についても検討を行う。

<sup>†1</sup> 龍谷大学理工学研究所

Graduate School of Science and Technology, Ryukoku University  
email:yyamamoto@nlp.i.ryukoku.ac.jp

## 2. 機械翻訳品質向上のための前編集

### 2.1 ルールベースの前編集

これまでに、翻訳品質向上を目的として、自然な原文を翻訳しやすい文に自動変換する前編集の研究は多く行われているが、それらは主にルールベースのものである。また、変換規則獲得のために、自然な文と翻訳しやすい文を用意する必要があり、その作成のコストが大きいという問題もある。

例えば、山口らは前編集前後の文から前編集規則を獲得する方法を提案している<sup>1)</sup>。ここでは人間が実際に前編集を行う必要がある。さらに、その際作業には、どのように前編集を行えば機械翻訳がより正しく翻訳できるかについての経験や知識が必要とされる。阿辺川らは翻訳経験の浅い翻訳者が翻訳した文である下訳と、ベテランの翻訳者が下訳を修正した後の文である修正訳を対応づけて変換ルールの獲得を目指している<sup>2)</sup>。この学習データの作成コストは大きい上に、これらを使って獲得した変換モデルが、実際の機械翻訳の前編集にとって適切かは保証されない。

このように従来のルールベースの前編集には、翻訳に関する深い経験や豊富な知識が必要であることや、変換ルールの適切性が証されないといった問題点がある。

### 2.2 統計的前編集

これに対し、本研究では特別な知識を必要としない統計的な前編集手法、すなわちデータから変換ルールを学習し、それをを用いて前編集する方法を考える。統計的前編集の枠組みに基づくと、前編集は、自然な原文  $S$  が与えられたときに、事後確率  $P(T|S)$  が最大となる機械翻訳しやすい文  $\hat{T}$  を見つける問題として定式化できる<sup>5)</sup>。

$$\hat{T} = \operatorname{argmax}_T P(T|S) \quad (1)$$

ベイズの定理を用いると式 (1) は式 (2) に変形できる。

$$\hat{T} = \operatorname{argmax}_T \frac{P(S|T)P(T)}{P(S)} \quad (2)$$

ここで、分母  $P(S)$  は  $T$  に無関係であるので省略でき、式 (3) を得る。

$$\hat{T} = \operatorname{argmax}_T P(S|T)P(T) \quad (3)$$

この  $P(S|T)$  は異なる言語の文字列の対応スコア、すなわち翻訳スコアであり、これを与えるモデルを翻訳モデルとよぶ。 $P(T)$  は目的言語の文字列としての現われやすさ（言語ス

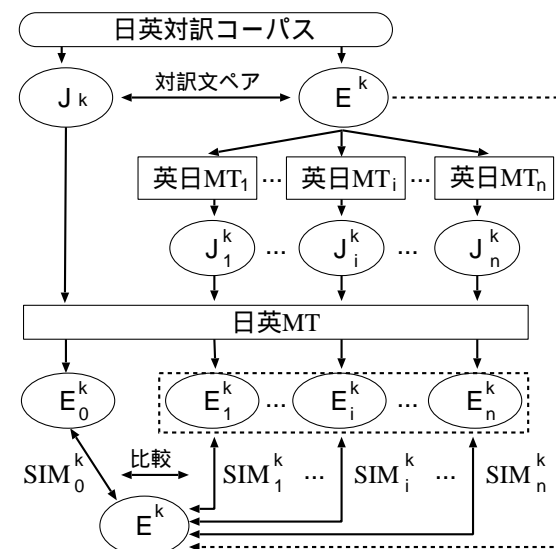


図1 前編集用学習データの作成手順

コア)を表しており、これを与えるモデルを言語モデルとよぶ。なお、本論文では、 $P(S|T)$  を与えるモデルを前編集変換モデルとよぶこととする。

本研究では、この枠組みに基づいた前編集（統計的前編集）を行う。統計的前編集は、前編集前の自然なテキストを  $S$  とし、前編集後のテキスト（翻訳しやすいテキスト）を  $T$  とし、 $P(S|T)$  と  $P(T)$  の積を最大とする  $\hat{T}$  を求める問題と説明できる。式 (3) からわかるように、統計的前編集のためには  $P(S|T)$  と  $P(T)$  を適切に学習することが重要となる。

## 3. 対訳コーパスからの前編集用学習データの選択

本章では、前編集変換モデルと言語モデルを得るために必要となる自然な文と翻訳しやすい文の対訳ペアの生成方法について述べる。ここでは、日英翻訳を対象として、文単位で対応づいた日英対訳コーパスと、日英および英日の双方向の翻訳システムを用いて  $P(S|T)$  と  $P(T)$  を学習するためのデータの自動獲得とそれを用いた統計的前編集について述べる。図1はこの様子を示したものである。前編集用学習データの獲得は、1) 翻訳しやすい文候補（図1の  $J_i^k$ ）を生成し、2) 自然な日本語文と翻訳しやすい日本語文のペア（図1の  $J^k$

と  $J_i^k$ ) から適切なものを選択して前編集用の学習データとする, という手順で行う. 3.1 節および 3.2 節では, これらの手順について詳しく述べる.

なお, 本手法は言語ペアに依存した手法ではなく, 原理的には, 言語ペアを入れ換えることでどのような言語ペアにも応用可能である.

### 3.1 翻訳しやすい文候補の生成

まず, 翻訳しやすい文候補の生成について述べる. これは具体的には以下の手順で行う.

- (1) 日英対訳コーパスの  $k$  番目 ( $k = 1 \dots N$ ) の日本語文  $J^k$  と, その対訳英文  $E^k$  のペアを用意する.
- (2)  $E^k$  を複数の翻訳システム  $MT_i$  ( $i = 1 \dots n$ ) で英日翻訳し, 日本語文  $J_i^k$  を得る.
- (3) 元の日本語文  $J^k$  と自動生成された日本語文  $J_i^k$  のペアを得る.

### 3.2 前編集変換モデルの学習データの選択

次に, 前節の方法で得られた  $J^k$  と  $J_i^k$  のペアの中から, 前編集変換モデルの学習データとして適切なものの選択方法について述べる. 前編集の目的は, 日英翻訳結果がより適切な英語となるようにすることであるため,  $J_i^k$  の日英翻訳結果  $E_i^k$  がもとの自然な日本語  $J^k$  の日英翻訳結果  $E_0^k$  よりも英語として適切でないもの場合は, そのような変換を学習すべきではない.

別の書き方をすると,  $E_i^k$  と参照訳  $E^k$  との類似度を  $SIM_i^k$  としたとき, 必ずしも  $SIM_i^k > SIM_0^k$  が成り立つわけではない. したがって,  $SIM_i^k \leq SIM_0^k$  となるような  $i$  ( $1 \leq i \leq n$ ) に対しては  $J^k$  を  $J_i^k$  に前編集するべきではなく, このような対応関係を学習データから除く必要がある. 具体的には, 以下の手順でデータを選択する.

- (1) 自然な日本語文  $J^k$  を日英翻訳し, 英文  $E_0^k$  を得る.
  - (2) 翻訳しやすい日本語文候補  $J_i^k$  を日英翻訳し, 英文  $E_i^k$  を得る.
  - (3) 参照訳  $E^k$  と  $E_i^k$  ( $i = 0 \dots n$ ) との類似度  $SIM_i^k$  を計算する.
  - (4)  $SIM_i^k > SIM_0^k$  を満たす  $i$  に対してのみ  $J^k$  と  $J_i^k$  のペアを前編集用の学習データとして選択する.
  - (5) 全ての  $i$  に対して  $SIM_i^k \leq SIM_0^k$  のときは,  $J^k$  と  $J^k$  のペアを学習データとする.
- こうして得られた自然な日本語文  $J^k$  と翻訳しやすい日本語文として選択された  $J_i^k$  のペア集合を前編集変換モデルの学習データとする.

なお, 統計的前編集に必要なもう一つのモデルである言語モデルについては, 翻訳しやすい日本語文として自動生成された  $J_i^k$  の集合から学習を行う.

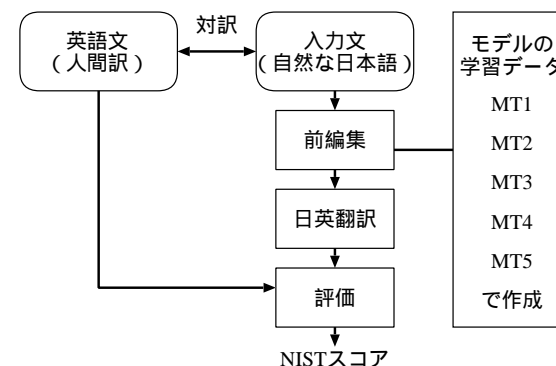


図 2 前編集実験の概観

## 4. 実験と結果

### 4.1 実験条件

学習データにはロイター日英対訳記事コーパス<sup>6)</sup>の日英対訳 31580 文対を用いた. 前編集用学習データの生成のための英日翻訳システムの数  $n$  は 5 とした. データ生成時の候補の英訳と参照訳との類似度  $SIM_i^k$  の算出には自動評価尺度 NIST<sup>7)</sup>を用いた.

統計的前編集システムは, 以下のデコーダとモデルで構成した.

- デコーダ: Moses SMT Decoder<sup>8)</sup>
- 言語モデル: IRST LM Toolkit<sup>9)</sup> を用いて学習した単語 5-gram モデル
- 前編集変換モデル: GIZA++<sup>10)</sup> を用いて学習したフレーズ変換モデルと reordering モデル

評価データにはロイター日英対訳記事コーパス中の日本語文 1000 文を用いた. これらの日本語文は学習データに含まれてはいない\*1. 最終的な日英翻訳結果の英文品質の評価には NIST を用いた. この実験手順は図 2 に示されている.

### 4.2 前編集用学習データ候補の生成に関する検討

はじめに, 前編集用学習データ候補の生成に関して検討を行った. 具体的には, 図 1 の英日翻訳システムとして, 種々の翻訳システムを用いて学習データを生成し, その効果を比較

\*1 対訳の英文 1000 文のうち, 8 文に関しては学習データと同一の文が評価データに含まれている

表 1 前編集の効果

学習データ作成に用いた 英日翻訳システム	英文品質 (NIST)	NIST の向上/低下 <sup>*1</sup> がみられた文の割合
前編集なし (ベースライン)	3.57	
MT1	<b>3.63</b>	50.5%/47.8%
MT2	3.36	42.0%/56.5%
MT3	3.31	40.3%/57.6%
MT4	3.60	49.3%/48.6%
MT5	3.55	48.1%/51.2%
MT1 ~ MT5 全て	<b>3.63</b>	<b>51.3%/48.0%</b>

\*NIST の変化がみられなかった文もあるため、合計は 100%にならない。

した。我々の以前の研究<sup>4)</sup>では、図 1 の英日翻訳システムを一つとしており、学習データを多く得にくいという問題があった。また、英日翻訳システムに対する頑健性も明らかでなかった。本章では、これらの問題について考察を行う。

市販の 5 つ英日翻訳システム (MT1, MT2, ..., MT5) のそれぞれで翻訳しやすい文候補を生成し、3.2 節で説明した手順にしたがって、その候補の中の適切なデータのみを学習データとして前編集変換モデルと言語モデルを学習した。これらのモデルを用いて前編集を行った上で、日英翻訳した結果を表 1 に示す。

前編集せずに日英翻訳した場合 (ベースライン) の英文品質 (NIST) は 3.57 であった。MT1 で前編集用学習データを生成・学習し、そのモデルを用いて前編集を行った上で日英翻訳を実行した場合は、英文品質 (NIST) の向上が見られた (3.57→3.63)。なお、改善した文の割合は 50.5%であった。同様に MT4 で前編集用学習データを生成した場合も、ベースラインの結果よりも高い英文品質が得られた。これに対して、上記以外の 3 つの日英翻訳システム単体で学習データを生成した場合は、平均としての英文品質の向上を確認できなかった。特に MT2 および MT3 では英文品質性能が大きく低下している。MT2 では、英日翻訳において接続詞や代名詞が可能性のある全ての単語の列挙として出力される (例: 「～と (彼・それ) らが言いました。」「ビルに入っていた (とき・から・につれて・ように) 」) ため、出力される日本文の構文解析が難しくなる上に、原文の英文に含まれる以上の情報を含む傾向にあった。MT3 では英単語がそのまま日本語文中に出力される割合が高いという特徴がみられた。その他、多くの英日翻訳システムにおいて、関係詞節が日本語文中に括弧くくりで埋め込まれて訳出されており (例: 「小麦 (それはおよそ 20 パーセントの完全穀物出力を占める) の大半は」)、前編集における語順入れ替えのモデルの学習に悪影響を及

表 2 自分自身への変換を前編集用学習データとして選択する効果

学習データ作成に用いた 英日翻訳システム	自分自身への変換	学習データ数 (文対)	英文品質 (NIST)
MT1 ~ MT5 全て	なし	143613	3.51
	あり	144078	3.63

ぼした可能性が考えられる。学習データを生成する翻訳システムの性質が統計的前編集に影響を与えることが確認できた。これらの対応については、たとえば学習データから複雑すぎる文を除くことが考えられる。ただし、どの英日翻訳機を使って前編集モデルの学習データを生成しても、文ごとにみた場合、40%から 50%の文で改善がみられることがわかった。

MT1 から MT5 の 5 つ全ての英日翻訳システムを用いて前編集用学習データを生成・学習し、そのモデルを用いて前編集を行った上で日英翻訳を実行した場合に、もっとも大きい英文品質 (NIST) の向上および改善した文の割合 (51.3%) が得られた。これは、多数の英日翻訳システムを用いることで、より多くの学習データを生成できたため統計モデルの信頼性が高まり、頑健性が増したためと考えられる。なお、学習データとして選択できた文ペアの総数は 144078 である。元の日英対訳コーパスの対訳ペア数は 31580 であり、そのほぼ 5 (英日機械翻訳の数) 倍のデータを学習データとして使用できることがわかった。

#### 4.3 適切な前編集用学習データ選択に関する検討

次に適切な前編集用学習データの選択手法の検討を行った。

##### 4.3.1 自分自身への変換データの選択

我々がこれまでに提案してきた適切な前編集用学習データの選択<sup>4)</sup>では、3.2 節で説明した手順の手順 (5) を行っていないかった。これにより、前編集システムは、自分自身に変換すること、すなわち前編集を行わないことを正しく学習できず、前編集が不要である文に対しても何らかの変換を行ってしまう可能性が高かった。本実験では、この問題に対して 3.2 節の手順 (5) を用いて対処を行い、その効果を示す。

自分自身への変換データの選択を行って学習データを生成した場合と、行わずに生成した場合 (従来手法) の結果を表 2 に示す。得られた自分自身への変換ペア数は 465 であった。それ以外の変換に関する学習データ数 143613 に比べると少ないが、この 465 文対を学習データとして加える効果が大きいことがわかる。

##### 4.3.2 英文品質の向上が大きいペアの選択

3.2 節の手順 (4) および (5) で前編集用の学習データを選択する際、翻訳しやすい日本

表 3 英文品質の向上が大きい学習データの選択の効果

学習データ作成に用いた 日英翻訳システム	重み $\alpha$	英文品質 (NIST)	学習データ数 (文対)
MT1 ~ MT5 全て	0	3.63	144078
	1	3.64	125197
	2	<b>3.67</b>	98384
	3	3.65	71792

表 4 絶対的品質の高い英文を生成できる学習データ選択の効果

学習データ作成に用いた 英日翻訳システム	重み $\alpha$	閾値 $\beta$	英文品質 (NIST)	学習データ数 (文対)
MT1 ~ MT5 全て	0	0	3.63	144078
		0	3.64	125197
	1	1	3.61	125113
		2	3.62	124598
		3	3.60	122248
		0	<b>3.67</b>	98384
	2	1	3.64	98304
		2	<b>3.67</b>	97833
		3	3.65	96250
	3	0	3.65	71792
		1	3.64	71694
		2	3.64	70838
		3	3.66	68353

語候補の  $J_i^k$  を日英翻訳して得られた英文  $E_i^k$  が、自然な日本語文  $J^k$  を日英翻訳して得られた英文  $E_0^k$  より少しでも優れていた場合に、学習データとして選択していた。このことは、品質向上が期待できない変換を多数学習することを示しており、変換候補の数が増加するために統計的前編集時の探索効率の低下や探索誤り増大の原因となると考えられる。

この問題に対し、本研究では英文品質の向上が大きいペアのみを前編集モデルの学習データとして選択することを考える。具体的には 3.2 節の手順 (4) および (5) で述べた類似度の比較の際に、自然な日本語文  $J^k$  を日英翻訳して得られる英文の評価値  $SIM_0^k$  に正の重みを与える。具体的には、それぞれの比較式を以下で置き換える。

$$\text{手順 (4): } SIM_i^k > SIM_0^k \rightarrow SIM_i^k > SIM_0^k + \alpha$$

$$\text{手順 (5): } SIM_i^k \leq SIM_0^k \rightarrow SIM_i^k \leq SIM_0^k + \alpha$$

重み  $\alpha$  を調整し、英文品質の向上が大きいデータのみを用いて前編集モデルを学習した。  $\alpha$  を 0, 1, 2, 3 とした時の結果を表 3 に示す。なお、 $\alpha = 0$  の場合は、前節までの実験と同じであることを意味する。重みを大きくすることで、不要な学習データを減らすことができ、英文品質 (NIST) の向上が得られることがわかる。ただし大きな重みを用いると学習データの絶対数が減少するため、前編集モデルの推定精度が十分でなくなると考えられる。表 3 には重みが 1, 2, 3 の場合の結果が示されており、実際に重みが 2 の場合の英文品質が最も高く、それより大きい重みを用いた場合は英文品質が低下することがわかる。

#### 4.3.3 絶対的品質の高い英文を生成できる学習データの選択

前節で述べたとおり、3.2 節の手順 (4) および手順 (5) の学習データの選択においては、二つの英文を比べて、より品質の高い英文を与える元の日本語文を変換ターゲットとして選択していた。ただし、ターゲットとして選択される日本語文が絶対的に品質の高い英文に訳されることは保証されていない。すなわち、極端に低い品質の英文に訳される日本語文であったとしても、相対的に品質の高い英文を生み出すものであれば変換ターゲットとして学

習される。このことは、統計的前編集時に悪影響を及ぼす可能性があるため、本研究では 3.2 節の手順 (4) と (5) の比較式をそれぞれ以下で置き換えて、絶対的品質の高い英文を生成できる学習データを選択することを考える。

$$\text{手順 (4): } SIM_i^k > SIM_0^k \rightarrow SIM_i^k > SIM_0^k + \alpha \text{ かつ } SIM_i^k > \beta$$

$$\text{手順 (5): } SIM_i^k \leq SIM_0^k \rightarrow SIM_i^k \leq SIM_0^k + \alpha \text{ かつ } SIM_0^k > \beta$$

本実験では、 $\beta$  を 0, 1, 2, 3 とした。結果を表 4 に示す。特に効果は確認できなかった。

#### 4.4 機械翻訳システムの変更に對する提案手法の頑健性

我々の提案している統計的前編集は、その後段に位置する機械翻訳システムに依存している。本節では、この後段の機械翻訳システムに種々のシステムを用い、提案手法の頑健性について調査した結果について述べる。図 1 および図 2 の日英翻訳システムを 4 種類用意し、それぞれについて実験を行った。結果を表 5 (「前編集あり」の欄) に示す。

MT\_A 以外では効果がほとんど見られないことがわかった。詳細に分析したところ、MT\_A 以外の日英翻訳システムは、英訳が困難な場合には元の日本語を出力する仕様となっており、前編集によって解析困難な日本語が生成された場合は英訳されず、英文品質が著しく低下していることがわかった。このように、前編集した後の日英翻訳結果が明らかに精度が低いと容易にわかる場合には、それを出力せずに、前編集なしの日英翻訳結果を出力すればよい。実際に、前編集後の日英翻訳結果に英文に使用されない文字が含まれていた場合に

表 5 機械翻訳システムの変更にに対する提案手法の頑健性

	日英機械翻訳システム			
	MT_A	MT_B	MT_C	MT_D
前編集なし	3.57	3.49	<b>3.65</b>	3.57
前編集あり	<b>3.63</b>	3.50	3.58	3.54
前編集+前編集失敗判定	<b>3.63</b>	<b>3.55</b>	3.63	<b>3.66</b>

表中の数字は最終的に得られた英文の品質 (NIST スコア)

は前編集が失敗したとみなし、前編集なしの日英翻訳結果を出力するシステムを構築した。結果は表 5 (「前編集+前編集失敗判定」の欄) に示されている。MT\_C 以外で前編集の効果がみられ、日英翻訳システムの変更に頑健であることがわかった。

## 5. 前編集結果の実例

実際におこなわれた前編集と、その日英翻訳結果の実例を図 3 から図 6 に示す。図 3 と図 4 は英文品質 (NIST) が向上した例であり、図 5 と図 6 は英文品質が低下した例である。

図 5 の例では、前編集前の日本語文中の「金価格」が前編集により「金の価格」となり、その結果、英訳において“the gold price”ではなく“the price of money”と翻訳されている。図 6 の例では、自然な日本語文の「話題」が「言いました」と誤って前編集されたため、“a topic”となるべきところが“called”と翻訳されてしまった。今後は、このような前編集がうまく働かない文について詳細な分析を行い、提案手法の改善を行っていく予定である。

## 6. 結 論

翻訳品質の向上を目的として、自然な原文を翻訳しやすい文に自動前編集するシステムを自動構築する方法について述べた。対訳コーパスと機械翻訳システムから前編集用学習データを自動生成する手法を提案した。学習データ候補の生成に用いる機械翻訳システムによらず、半数程度の文で前編集による改善が得られた。また、前編集の後段に位置する機械翻訳システムの変更に頑健であることを示した。

## 参 考 文 献

- 1) 山口昌也, 乾伸雄, 小谷善行, 西村恕彦. 前編集結果を利用した前編集自動化規則の獲得. 情報処理学会論文誌, Vol. 39, No. 1, pp.17-28, 1998.
- 2) 阿辺川武, 影浦峽. 下訳と修正訳を用いた訳文修正パターンの発見. 言語処理学会 第 13 回年次大会発表論文集, pp.919-922, 2007.
- 3) 岡田真也, 南條浩輝, 吉見毅彦. 機械翻訳における前編集規則の自動獲得. 言語処理学会第 15 回年次大会 (NLP2009) 講演論文集, P1-28, 2009.
- 4) 南條浩輝, 吉見毅彦, 岡田真也. 機械翻訳のための統計的手法に基づく前編集. 情報処理学会研究報告, 2009-SLP-76-1, 2009.
- 5) Brown.P.F, Della Pietra.V.J, Della Pietra.S.A, and Mercer.R.L. The mathematics of statistical machine translation: Parameter estimation. In *Computational Linguistics*, Vol. 19, No.2 pp.263-311, 1993.
- 6) M.Utiyama and H.Isahara. Reliable measures for aligning japanese-english news articles and sentences. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics(ACL)*, pp. 72-79, 2003.
- 7) G.Doddington. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In *Proceedings of the 2nd Human Language Technologies Conference(HLT)*, pp. 128-132, 2002.
- 8) H.Hoang and P.Koehn. Design of the mooses decoder for statistical machine translation. In *Software Engineering, Testing, and Quality Assurance for Natural Language Processing*, pp. 58-65, 2008.
- 9) M.Federico and M.Cettolo. Efficient handling of n-gram language models for statistical machine translation. In *Proceedings of the Second Workshop on Statistical Machine Translation*, pp. 88-95, 2007.
- 10) FranzJosef Och and Hermann Ney. A systematic comparison of various statistical alignment models. In *Computational Linguistics*, Vol. 29, No. 1, pp. 19-51, 2003.

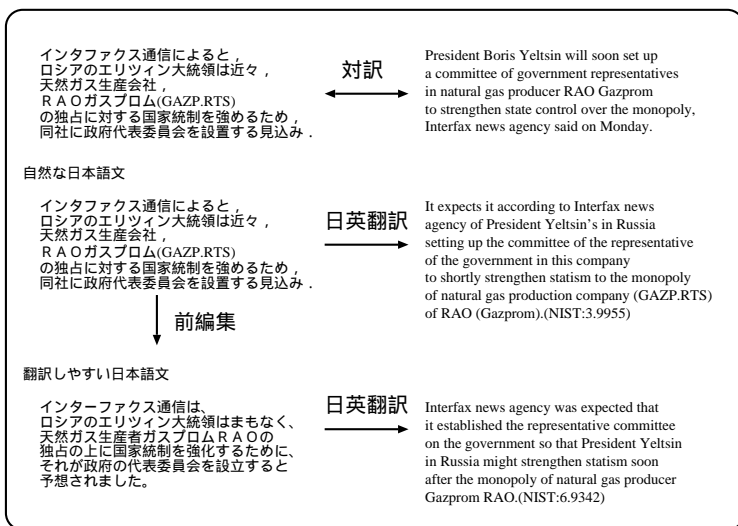


図 3 日英翻訳品質が向上した例 1

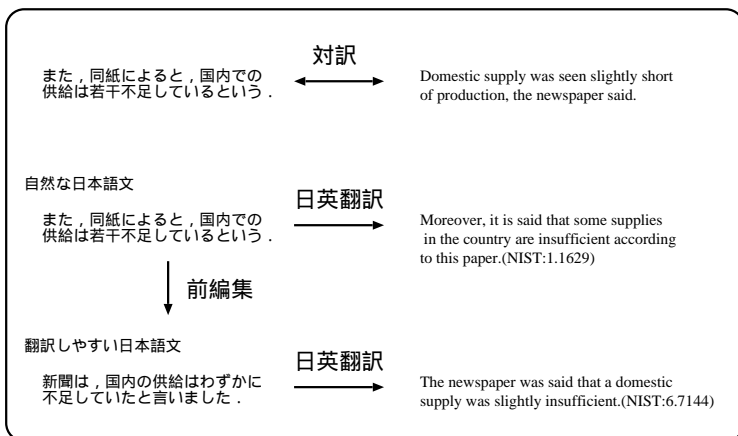


図 4 日英翻訳品質が向上した例 2

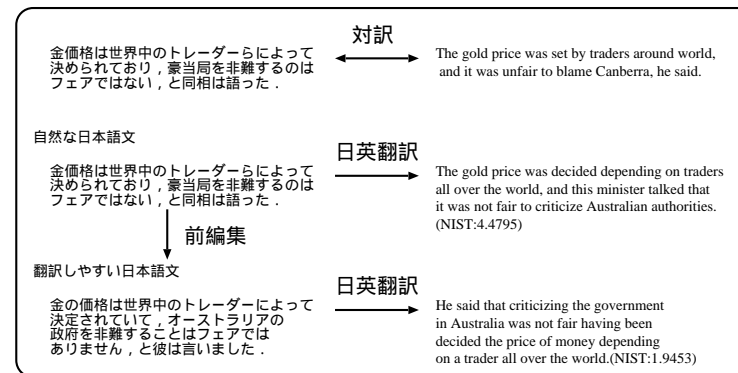


図 5 日英翻訳品質が低下した例 1

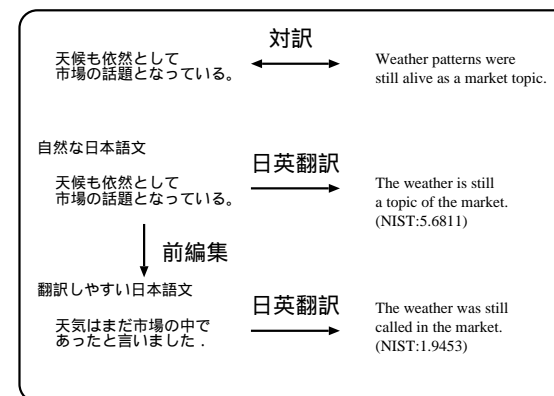


図 6 日英翻訳品質が低下した例 2