

機械翻訳の品質向上のための 対訳コーパスからの統計的前編集システムの自動構築

南條 浩輝^{1,a)} 山本 祐司² 吉見 毅彦¹

受付日 2011年11月7日, 採録日 2012年3月2日

概要: 機械翻訳 (MT) の品質向上のための翻訳前の書き換え (前編集) について述べる。これまでの前編集は主にルールに基づくものであるため, MT システムに依存し, かつシステム構築に労力を要した。これに対し, 本論文では, MT システムに依存しない統計的前編集システムの自動構築手法を提案する。具体的には, 対象とする MT システムに適した前編集システムを, 当該 MT システムを含む複数の MT システムおよび対訳コーパスから自動構築する手法を提案する。本手法の最も重要な特徴は, 前編集によって翻訳品質の向上が得られる学習データを自動生成できる点にある。ロイター日英記事の対応付けデータを用いて 4 種類の日英 MT システム用前編集システムを構築したところ, 3 種類の MT システムについて, 多くの文に対して翻訳品質を向上させる前編集システムが構築できた。提案手法である前編集システムの学習データの自動生成は, 3 種類の MT システムの前編集システム構築に効果的であったこと, および他の 1 種類の MT システムの前編集システム構築に悪影響がなかったことを示し, 提案手法の有効性を示した。

キーワード: 統計的前編集, 書き換え, 学習データ選択, 機械翻訳

Automatic Construction of Statistical Pre-editing System from Parallel Corpus for Improvement of Machine Translation Quality

HIROAKI NANJO^{1,a)} YUJI YAMAMOTO² TAKEHIKO YOSHIMI¹

Received: November 7, 2011, Accepted: March 2, 2012

Abstract: Rewriting before machine translation (MT), namely pre-editing, is addressed. Most of conventional pre-editings are MT dependent rule-based systems, and construction of such systems costs too much. In this paper, we propose an automatic construction method of MT independent statistical pre-editing systems. Specifically, for a specific MT system, statistical pre-editing system is trained with appropriate training data which are automatically generated with parallel corpus and some MT systems including the target MT system. Using Reuters Japanese-English parallel corpus, effective pre-editing systems were automatically constructed for several Japanese to English MT systems. We showed the effectiveness of the proposed method.

Keywords: statistical pre-editing, rewriting, training data selection, machine translation

1. はじめに

翻訳前の書き換え (前編集) について述べる。現在, さ

まざまな機械翻訳 (Machine Translation; 以下, MT) システムが実用化され, 多くの場面で使用されている。翻訳とは, 原言語の単語列を目的言語の単語列に変換する処理と定義でき, MT システムは単語列の変換モデルを持っている。しかし, 自然言語においては表現はほぼ無限に存在するため, どのような MT システムであっても適切に翻訳できない表現 (単語列) が存在する。このような表現は, 実際の MT システムの運用上で明らかになることが多いも

¹ 龍谷大学理工学部
Faculty of Science and Technology, Ryukoku University,
Otsu, Shiga 520-2194, Japan

² 龍谷大学理工学研究科
Graduate School of Science and Technology, Ryukoku University,
Otsu, Shiga 520-2194, Japan

a) nanjo@rins.ryukoku.ac.jp

の、実際に運用されている MT システムは非常に複雑であるため、現状の MT システムとの整合性を保ちながら新たな改良を加えることは容易ではない。さらに、実際の運用にあたっては MT システムの開発者が運用にかかわらないことも多いため、MT システムの翻訳モデルやアルゴリズムに関する深い知識を持たない運用者側やユーザ側でこの問題を解決する方法が強く望まれる。これらを解決する方法の 1 つとして、MT システム自体には手を加えずに、翻訳前に原言語表現を書き換える（前編集する）方法がある。前編集は、すでに運用されている MT システムをそのまま利用できる点、および MT システムにおいて翻訳可能な表現には影響を与えずに翻訳困難な表現のみに素早く対応できる点で、運用上の利点が多い。本論文では、このような前編集について述べる。

これまでに前編集に関する研究は行われているものの [1], [2], [3], それらは主にルールベースのものである。適切に前編集を行うためには対象とする MT システムの特性を調べて、前編集対象となる表現とそれと同じ意味の翻訳可能な表現を獲得することが必要である。これは、MT システムの翻訳モデルおよび翻訳アルゴリズムに関する知識がなければ効率的に行えない。商用の MT システムや WEB 上での MT サービスでは基本的に翻訳アルゴリズムが公開されておらず、実際に運用されている MT システム・サービスに対して適切な前編集システムを運用側やユーザ側で実現するのは困難であった。

この問題に対して、本論文では、MT に関する知識を必要としない、MT システム非依存の前編集システムの自動構築手法を提案する。具体的には、対象とする MT システムに適した前編集システムを、当該 MT システムを含む複数の MT システムおよび対訳コーパスから、自動的に構築する手法を提案する。本提案手法と同様に、対訳コーパスと MT システムを用いて前編集用の学習データを自動生成する試みはこれまでもみられる [4]。これは、目的とする方向とは逆方向の MT システムと既存の対訳データを用いて原言語文の書き換えデータを生成するものである。しかし、得られた学習用の書き換えデータが前編集にとって適切であるかの保証が与えられていないという問題がある。

この背景に基づき、本論文では、前編集システム自動構築の枠組みにおいて、新たに一般化された前編集に有用な書き換えデータの選別アルゴリズムを提案し、その有効性を示す。その際、これまで調査が行われていなかった統計的前編集システムにおけるモデル、具体的には統計的言語モデルのモデル化手法について調査を行い、その結果について述べる。さらに前編集システムの自動構築手法の MT システムおよび対訳コーパスの変更に対する頑健性の調査も行い、その結果についても述べる。

本論文の構成は以下のとおりである。2 章では関連研究について述べる。3 章で前編集用の学習データの自動獲得

手法について述べ、4 章で前編集について述べる。5 章と 6 章ではそれぞれ実験システムと結果について述べ、7 章で結論を述べる。

2. 関連研究

本章では、提案手法に関連する先行研究について述べ、提案手法の新規性および利点を明確にする。

提案手法は、前編集用の学習データを自動獲得し、自動獲得された学習データから統計的機械翻訳の枠組みに基づく前編集システムを構築するものである。そのため、MT システムに関する知識が不要かつ MT システム非依存の手法である。これに対し、人手でルールを記述しそれを用いるルールベースの前編集がある。山口らは前編集前後の文から書き換えルールを獲得する方法を提案している [2]。ここでは人間が実際に前編集を行い、どのように前編集を行えば翻訳品質が向上するかを調査する必要がある。適切な書き換えルールを得るためには、MT システムに関する経験および知識を持った作業者を必要とするため、書き換えルールを大量に獲得するのは容易ではない。さらに、MT システムごとの翻訳モデルの違いにより書き換え対象となる単語列が異なるため、各 MT に特化した作業者の育成が必要であり、コストが大きいという問題がある。Xu らは意味解析を行ったうえで前編集する手法を提案している [3]。前編集はルールベースで行われており、基本的に語順の入れ替えが行われている。英中翻訳を指向した前編集手法であり、汎用性は高くない。提案手法は、ルールベースの前編集システムの構築と比べて、汎用性が高いという利点がある。

提案手法は、対象とする MT システムの前編集にとって有用なデータのみを学習データとして自動獲得するものである。翻訳精度を向上させるために書き換えデータを獲得しようとした研究には、阿辺川らの研究 [5] や Zhao らの研究 [6], Sun ら [4] の研究がみられる。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修正操作（本論文での書き換えに相当）を獲得している [5]。人手作業を要求するうえに、下訳-修正訳の修正操作の前編集ルール獲得用データとしての適切性は保証されていない。Zhao らは、言い換え (paraphrase) を自動生成する研究を行っている [6]。MT システムによって言い換え候補を自動生成している点において、提案手法と関連がある。言い換えは原文の意味が保持されるように行われるが、前編集を目的としておらず、得られる言い換えが前編集にとって適切かは保証されない。Sun ら [4] の手法は、提案手法と同様に、対訳コーパスと MT システムを用いて前編集用の学習データを自動生成するものである。しかし、得られた学習用の書き換えデータが前編集にとって適切かの保証が与えられていないという問題がある。さらに、英中翻訳

において前置詞を含む文の翻訳が難しいことに着目した手法であるため、学習データ・評価データは前置詞を含む文に限定されており、汎用性に問題がある。これに対し、提案手法は、前編集システム自動構築の枠組みにおいて前編集に有用な書き換えデータのみを選別するものであり、この点において新規性を有する。その際、獲得したデータから統計的前編集システムを構築する際の統計的言語モデルのモデル化手法についての調査、および前編集システム自動構築手法の MT システムおよび対訳コーパスの変更に対する頑健性についての調査も行ったのでその結果について述べる。これらの調査はこれまでに行われておらず、この調査で得られる知見は新しい。

3. 対訳コーパスからの前編集用学習データの生成と選択

本章では、ある MT システム用の前編集のための学習データ、すなわち前編集に適した書き換えデータを自動的に生成する提案手法について述べる。具体的には、対象とする MT システム (Source-Target MT システム；以下 S-T MT) と文単位で対応づいた対訳コーパス、および目的言語から原言語方向の複数の MT システム (Target-Source MT システム；以下 T-S MT) を用いて、当該 S-T MT の前編集のための学習データを自動生成する方法について述べる。本手法は言語ペアに依存しておらず、対訳コーパスと MT システムがあればどのような言語ペアにも応用可能である。

図 1 に学習データの自動生成方法を示す。まず、各対訳文ペアの目的言語文を複数の T-S MT を用いて原言語の文に翻訳して書き換え候補 (図 1 の S_i^k) とし、次にオリジナルの原言語文と得られたそれぞれの書き換え候補のペア (図 1 の S^k と S_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。3.1 節で書き換え候補の生成方法について、3.2 節で適切なデータの選択方法について詳しく述べる。

3.1 書き換え候補の生成

まず、書き換え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) 対訳コーパスの k 番目の原言語の文 S^k と、その対訳の目的言語の文 T^k のペアを用意する。
- (2) T^k を複数の目的言語-原言語方向の MT システム T-S MT $_i$ ($i = 1 \dots n$) で翻訳して複数の原言語の文 S_i^k を生成し、書き換え候補とする。

関連研究 [4] での書き換え候補の生成は手順 (2) での T-S MT 数を 1 とした方法と見なすことができ、提案手法はこれを包含する一般化された手法と位置づけられる。

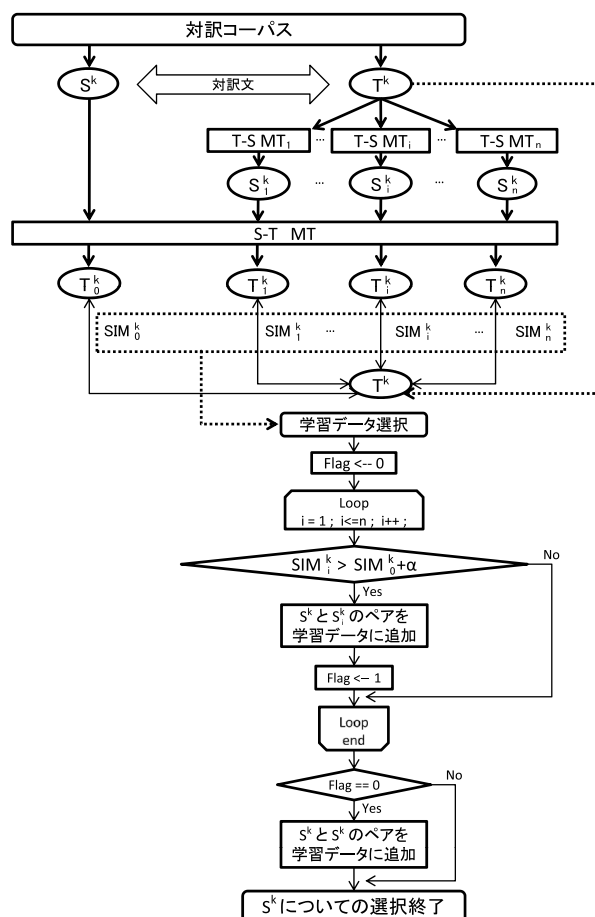


図 1 前編集用学習データの自動生成

Fig. 1 Automatic generation of training data for pre-editing system.

3.2 学習データの選択アルゴリズム

次に、前節の方法で得られた S^k と S_i^k のペアの中から、対象とする S-T MT の前編集用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、より適切な翻訳結果 (目的言語の文) を得ることである。ある S-T MT による前編集後の文 S_i^k の翻訳結果 T_i^k が、オリジナル文 S^k の翻訳結果 T_0^k よりも不適切である場合は、そのような $S^k \rightarrow S_i^k$ の書き換えを当該 S-T MT のための前編集で行うべきではない。別の書き方をすると、 T_i^k と参照訳 T^k との類似度を SIM_i^k としたとき、 $SIM_i^k \leq SIM_0^k$ となるような i ($1 \leq i \leq n$) に対しては S^k を S_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。このような背景に基づき、我々は次の方法を提案する。

- (1) 原言語のオリジナル文 S^k を翻訳し、目的言語文 T_0^k を得る。
- (2) 参照訳 T^k と T_0^k との類似度 SIM_0^k を計算する。
- (3) 書き換え候補 S_i^k ($i = 1 \dots n$) を翻訳し、目的言語文 T_i^k を得る。
- (4) 参照訳 T^k と T_i^k との類似度 SIM_i^k を計算する。
- (5) $SIM_i^k > SIM_0^k$ を満たす i に対してのみ S^k と S_i^k のペア

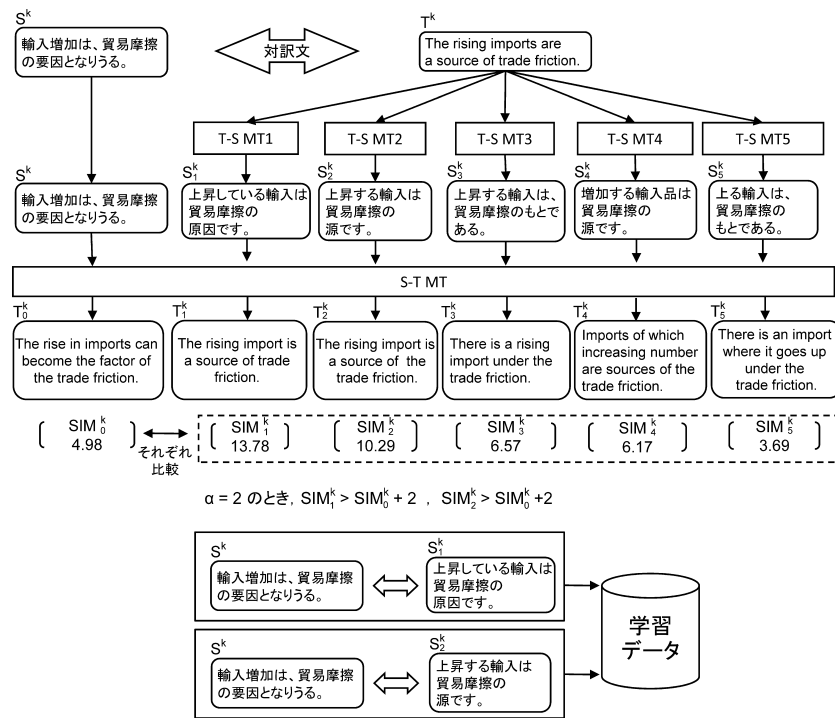


図 2 日英翻訳の前編集のための学習データの選択例

Fig. 2 Example of training data selection for pre-editing system of J-E translation.

アを前編集用の学習データとして選択する。

関連研究 [4] では、このような選択 (手順 (2), (4) および (5)) は考慮されておらず、上で述べた不適切な書き換えデータも前編集用学習データに加えてしまうという問題がある。すなわち、前編集用学習データの適切性が保証されないという問題がある。これに対し提案アルゴリズムでは、翻訳性能が向上する書き換えデータのみを学習データとして選択できる。ただしこの選択アルゴリズムでも、手順 (5) における選択基準の検討が不十分である。具体的には、当該 S-T MT による前編集後の文 S_i^k の翻訳結果 T_i^k が、オリジナル文 S^k の翻訳結果 T_0^k に比べて少しでも優れている (参照訳との類似度が高い) 場合、すなわち品質向上がほとんど期待できない書き換えデータである場合でも学習データとして選択する点に問題がある。品質向上がほとんど期待できない書き換えデータを学習データに含めることは、書き換え候補数の増大による前編集時の処理効率の低下や、書き換えの競合候補数の増大にともなう前編集誤りの増加の原因となると考えられる。

この問題に対し、本論文では翻訳品質の向上が大きい書き換えデータのみを前編集システムの学習データとして選択するアルゴリズムも提案する。具体的には手順 (5) で述べた類似度の比較に基づく選択時に、オリジナル文 S^k の翻訳結果 T_0^k の類似度 SIM_0^k に正の重み α を与える。すなわち手順 (5) を以下の手順 (5') で置き換える。

(5') $SIM_i^k > SIM_0^k + \alpha$ を満たす i に対してのみ S^k と S_i^k のペアを前編集用の学習データとして選択する。

このアルゴリズムで選択される学習データでは、前編集

を行わないことを正しくモデル化することが困難である。すなわち、この修正した学習データの選択アルゴリズムには、 S^k を S^k 自身に書き換える学習データを獲得できないという問題が残っている。前編集時に、前編集が不要である文に対しても何らかの前編集を適用し、翻訳品質の低下を招く可能性がある。この問題に対し、手順 (1) から (5') に以下の手順 (6) を追加することを提案する。

(6) すべての i に対して $SIM_i^k \leq SIM_0^k + \alpha$ のときは、 S^k と S^k のペアを学習データとする。

この処理により、書き換えないことが望ましい表現、すなわちそのままの形が当該 S-T MT にとって翻訳しやすい表現も学習データに追加できる。

実際に日英翻訳の前編集を対象として学習データを選択する例を図 2 に示す。ここでは選択時重み $\alpha = 2$ としている。オリジナルの文の翻訳結果 T_0^k と参照訳 T^k の類似度 $SIM_0^k = 4.98$ である。前編集後の文 S_i^k の翻訳結果 T_i^k と参照訳 T^k の類似度はそれぞれ $SIM_1^k = 13.78$, $SIM_2^k = 10.29$, $SIM_3^k = 6.57$, $SIM_4^k = 6.17$, $SIM_5^k = 3.69$ であり、 $SIM_1^k > SIM_0^k + 2$, $SIM_2^k > SIM_0^k + 2$ であるので、 S^k と S_1^k のペア、 S^k と S_2^k のペアが前編集用の学習データとして選択される。

4. 機械翻訳品質向上のための前編集

4.1 前編集の機械翻訳への組み込み

本節では、機械翻訳品質向上を目的とした、任意の MT システムへの前編集の組み込み方法について述べる。図 3 に組み込みの様子を示す。MT システムの前段に前編集部

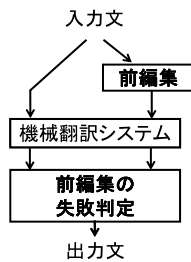


図 3 統計的前編集を組み込んだ機械翻訳システム

Fig. 3 Incorporation of statistical pre-editing system into MT system.

をおき、MT システムへの入力、オリジナルの文とそれを前編集した文とする。前編集した文の翻訳後、前編集の失敗判定を行い、失敗と判断した場合はオリジナル文の翻訳結果を、そうでなければ前編集後の文の翻訳結果を出力する。ここで前編集失敗とは、オリジナルの文の翻訳結果の品質に比べて、前編集した文の翻訳結果の品質が低くなることを指す。このような失敗判定は必ずしも必要ではないが、適切な失敗判定が行えれば前編集による悪影響を抑えることができ、全体として翻訳品質の向上が得られる。実際には、与えられた 2 文のうちどちらの翻訳品質が高いかを決定することは難しいため、前編集後の文の翻訳結果が明らかに不自然な場合を失敗と見なせばよい。その失敗判定基準例としては、「前編集後の文の翻訳結果に原言語の文字が含まれている場合は失敗」があげられる。本論文では、提案する前編集システムをこの枠組みと失敗判定基準を用いて MT システムに組み込み、評価を行った。

4.2 統計的前編集

3 章では、前編集用の適切な学習データの自動生成方法について述べた。本節では、この自動獲得した学習データを用いて、翻訳に関する知識を必要としない前編集システムの自動構築について述べる。具体的には、自動獲得した学習データを用いた統計的機械翻訳の枠組みに基づく前編集（統計的前編集）を行う。

統計的機械翻訳には、対訳コーパスから学習される翻訳に関するモデルと、目的言語のテキストから学習される言語的制約に関するモデルが必要である。対訳コーパスとしてオリジナルの文と前編集後の文のペアを、目的言語のテキストとして前編集後の文集合を与えることで、統計的前編集システムが学習できる。本論文では、オリジナルの文と前編集後の文のペアから学習されるモデルを前編集変換モデルとよぶこととする。また、目的言語のテキストから学習される言語的制約に関するモデルを言語モデルとよぶこととする。

前編集変換モデルの学習には 3.2 節の選択アルゴリズムで得られる書き換えデータを、言語モデルの学習には得られる書き換えデータのうちの書き換えターゲット側を用い

ることができる。これにより、前編集システムを自動で構築できる。

5. 実験システムの構成

5.1 学習データと評価データおよび評価基準

本論文で提案する前編集の自動構築手法は言語ペアに依存しない手法である。今回は、日英翻訳において提案手法の有効性を示す。

まず日英翻訳の前編集のための学習データの自動生成に用いる、日英対訳コーパスおよび日英 MT システムと英日 MT システムについて述べる。日英対訳コーパスには、ロイター日英記事対応付けデータ（以下、ロイター対訳コーパス）[7] 中の対訳文を用いた。ロイター対訳コーパスでは、英日間の自動対応付けが行われており、複文に対する対応付け精度が低い。本実験では、このような対応の信頼が低いものをデータから除いた。具体的には、日本語文にかぎっこ“[” または “]” を含む対訳を使用しないこととした。さらに、ロイター対訳コーパスでは、いくつかの対訳ペアでは英語文は異なるものの日本語文が同じものがある。これらが含まれていると、それらの日本語文に特化された学習が行われる可能性があること、評価データを抽出する際に学習データに含まれるものが抽出される可能性があることから、本実験ではこれらをデータから除いた。さらに評価用のデータ 1,000 文対を除き、残りの総数 31,580 文対を学習データとして用いた。日英 MT、英日 MT とも、商用の MT とした。なお英日 MT の数 n は 5 とした。実験に用いた日英 MT と英日 MT の一覧は表 7 に示されている。

次に評価データについて述べる。評価データには先ほど述べたロイター対訳コーパス中の日本語文 1,000 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。

最後に評価基準について述べる。翻訳結果の評価尺度にはさまざまなものがある。提案手法は 3.2 節でのデータ選択時にある評価尺度に基づいて評価値（類似度）を設定し、その類似度が向上することが保証されたデータを学習データとするものである。人間が関与する必要がある評価値を用いるとデータ選択時に自動化が行えないため、提案手法では自動評価尺度が望ましい。したがって本実験では、人手評価と相関が強い [8] とされる NIST スコア [9] を評価値（類似度）として用いる。すなわち、最終的に翻訳された英文の評価にも、3.2 節でのデータ選択時に用いる類似度 SIM にも、NIST スコアを用いる。なお、本提案手法の枠組みは用いる自動評価尺度に依存しないものである。

*1 対訳の英文 1,000 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

Table 1 Effects of pre-editing.

	言語モデル			学習データ数 (文数)
	3-gram	4-gram	5-gram	
学習データ選択なし	3.64	3.59	3.56	157,900
学習データ選択あり ($S^k \rightarrow S^k$ 変換なし)	3.66	3.60	3.55	143,613
学習データ選択あり ($S^k \rightarrow S^k$ 変換あり)	3.66	3.61	3.65	144,074

表中の数字は最終的に得られた英文の翻訳品質 (NIST スコア)

5.2 統計的前編集システム

5.2.1 前編集変換モデル

3.2 節で述べたデータ自動作成アルゴリズムにより前編集学習データを生成し、前編集変換モデルを学習した。本実験では、オリジナルの日本語文 S^k と書き換えターゲットの日本語文 S_i^k をそれぞれ形態素解析して、その形態素を単位として、前編集変換モデルを学習した。具体的には、Moses SMT Decoder (2008-7-11 version) [10] 付属の train-factored-phrase-model.perl (デフォルト設定) と GIZA++-v2 [11] を用いて、フレーズ翻訳モデルと reordering モデルを学習した。その際、形態素解析には ChaSen version 2.2.1+ipadic-2.4.1 [12], [13] を使用し、形態素の表層表現のみを使用した。

5.2.2 言語モデル

言語モデルは、自動作成された書き換えデータのうちの書き換えターゲットの日本語文 S_i^k のみから学習した。ChaSen version 2.2.1+ipadic-2.4.1 を用いて形態素解析を行い、形態素単位の N-gram 言語モデルを学習した。その際、IRST LM Toolkit [14] を使用し、 $N = 3, 4, 5$ として N-gram 言語モデルを学習した。ここでも形態素の表層表現のみを使用した。

5.2.3 デコーダ

統計的前編集では、前編集は、入力文に対して前編集変換モデルと言語モデルのスコアを算出し、それらを組み合わせたスコアが最大となる文を探索し見つける問題である。この探索部はデコーダとよばれ、本実験では Moses SMT Decoder (2008-7-11 version) [10] (デフォルト設定) を用いた。Moses には種々のデコーディングパラメータが存在するが、本実験ではそれらの変更は行わないこととした。すなわち、パラメータチューニングは行わず、train-factored-phrase-model.perl で前編集変換モデルを学習した際に出力されるパラメータ設定ファイル (デフォルト) をそのまま用いた。

5.3 前編集失敗判定

本実験では、前編集を図 3 の枠組みに組み込んで評価を行った。前編集失敗判定として今回は簡易な手法を用いた。具体的には、前編集後の文を日英翻訳した結果に、日本語の文字列が含まれている場合に失敗として判定した。

6. 評価実験

6.1 適切な前編集用学習データ選択の効果

商用の日英 MT システム (MT_A) を用いて 1,000 文を日英翻訳した。翻訳品質 (人手参照訳との NIST スコア) は 3.57 であった。次に、適切な前編集用学習データの選択手法の評価について述べる。

6.1.1 学習データの単純選択と自身への書き換えデータの選択

はじめに、3.1 節のアルゴリズムで生成した書き換え候補 $S^k \rightarrow S_i^k$ ($i = 1 \dots 5$, $k = 1 \dots 31,580$, 合計 157,900) をすべて用いて統計的前編集した結果について述べる。結果を表 1 (学習データ選択なしの欄) に示す。前編集システムでの言語モデルを 3-gram, 4-gram, 5-gram としたときのそれぞれの翻訳品質は、3.64, 3.59, 3.56 であり、いずれもベースラインよりも高い翻訳品質が得られた。統計翻訳 (本手法では統計的前編集) では、N-gram の N として 5 以上が用いられるのが一般的であるが、本実験では $N = 3$ のときの翻訳品質が最も高かった。これはデータ数の少なさによるものと考えられる。

次に、3.2 節の手順 (5') と (6) による選択の効果について述べる。その際、手順 (5') の選択時の重み $\alpha = 0$ とした (手順 (5) を用いたことに相当)。結果は表 1 の下段に示されている。手順 (5) のみで前編集データを選択した場合 ($S^k \rightarrow S^k$ 変換なし) の翻訳品質は、前編集システムでの言語モデルを 3-gram, 4-gram, 5-gram としたとき、それぞれ 3.66, 3.60, 3.55 であった。データ選択をしない場合からの改善は小さかった。これは、本実験では MT システムで自動生成したデータの多く (全体の 91%; 143,613 ペア) が前編集用の書き換えデータとして適切であったためと考えられる。学習データ生成時に自身への書き換えデータを選択する手法 (手順 (6) を用いる手法) では、461 の自身への書き換えデータを選択した。これはデータ全体の 0.3% 程度であるため、大きな効果はみられなかったものの、これを用いた前編集により最も高い翻訳品質を得ることができた。

これらのデータ選択手法は悪影響は及ぼさず、効果が期待できることが分かった。データ生成に使用する MT システムの性能や対訳コーパスによっては、書き換えないこと

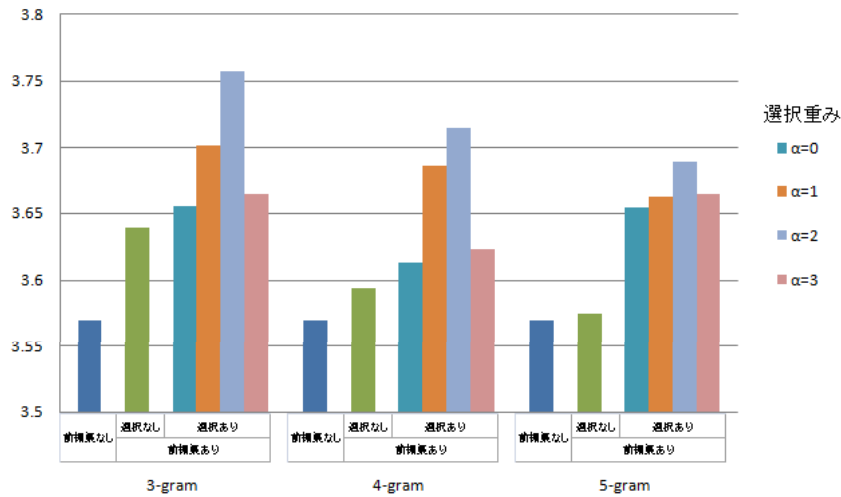


図 4 学習データ選択の効果 (言語モデルごと)

Fig. 4 Effects of training data selection.

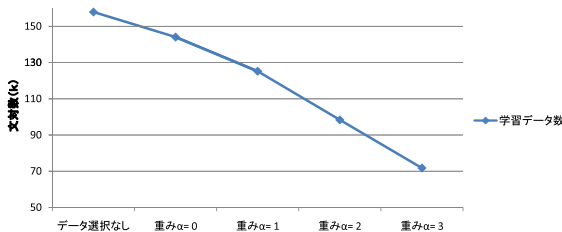


図 5 重み α を用いたときの学習データ数の変化

Fig. 5 Size of training data using each α.

が望ましいデータが多く生成される可能性もあり (本実験ではそのようなデータは 10% 未満), そのような場合には提案手法はより有効に働くと考えられる。

6.1.2 翻訳品質の向上が大きい書き換えデータの選択

次に, 3.2 節の手順 (5') および (6) での選択時の重み α を用いる, 翻訳品質の向上が大きい書き換えデータの選択アルゴリズムの効果について述べる. α を 0, 1, 2, 3 としたときの結果を図 4 に示す. その際の学習データ数を図 5 に示す. 比較のために, データ選択をしない場合と前編集をしない場合の結果も載せてある. なお, α = 0 の場合は, 前節の実験 (表 1 の最下段) と同じである. 大きい重み α を用いることで, 不要な学習データを減らすことができ, 翻訳品質 (NIST) の向上が得られることが分かる. ただし, 図 5 に示すように, 重みを大きくするに従って学習データ数が減少するため, 前編集用の統計モデルの推定精度が十分でなくなると考えられる. 実際に図 4 に示すように, 重みが 2 の場合に最も高い翻訳品質が得られており, それより大きい重みを用いた場合には翻訳品質が低下することが分かる.

本実験では α = 2, 言語モデルに単語 3-gram を用いる場合に最高の翻訳品質が得られ, 3.76 であった. 学習データ選択なしの場合に比べて本提案手法が優れているかについて符号検定を用いて検定した. すなわち, 提案手法によ

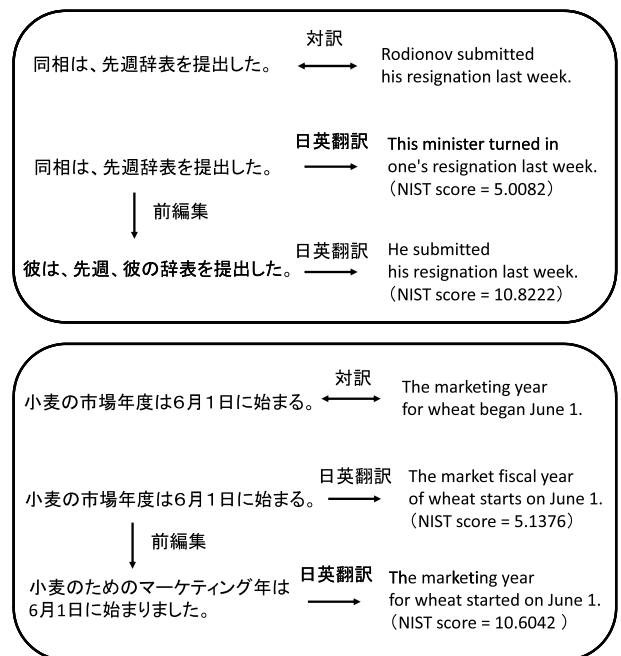


図 6 前編集事例 (成功例)

Fig. 6 Examples of appropriate pre-editing.

り翻訳品質が向上した文と低下した文の数を比較し, 向上する文が多いかを検定した. 検定の結果, 提案手法は有意水準 1% で有意であった.

なお, 言語モデルに 4-gram, 5-gram を用いた場合も同様に有意水準 1% で有意であった. これらのことは, 提案手法の有効性を示している.

図 6 に実際に前編集により翻訳品質が向上した例を示す. 「彼の」という通常の日本語では省略される代名詞を補完する前編集や, 「年度」を「年」に書き換える前編集が行われ, 前編集後の日本語が英語表現に近くなっており, 実際に英語に翻訳した際の品質が向上していることが分かる. 図 7 に前編集により翻訳品質が低下した例を示す.



図 7 前編集事例 (失敗例)

Fig. 7 Examples of inappropriate pre-editing.

前編集後の日本語が構文解析できない文であったり、前編集で余計な語を挿入したため意味が異なる文になっていた例などがみられた。構文解析ができない文については、言語モデルを大量のデータで学習することで改善される余地があると考えられる。

6.2 対訳コーパスの変更に対する提案手法の頑健性

対訳コーパスの変更に対する提案手法の頑健性について調査するため、6.1 節と同じ実験 (MT_A を用いた実験) を学習データおよび評価データのみを変更して行った。具体的には、以下の学習用データと評価用データを用いた。

- 学習用データ：読売新聞日英コーパス 140,000 文対
- 評価データ：読売新聞日英コーパス 1,000 文対

学習データの選択は提案アルゴリズムで行い、6.1 節の実験で最も翻訳品質が高かったデータ選択時の重み $\alpha = 2$ および単語 3-gram 言語モデルを使用した。結果を表 2 に示す。読売新聞対訳コーパスを用いた実験でも改善が得られることが分かる。

表 3 に、データ選択時の重み α および単語 N-gram 言語モデルの次数を変更したとき ($N = 1, 2, 3$) の結果を示す。この対訳コーパスを用いた実験でも $\alpha = 2$, $N = 3$ のときに最も高い翻訳品質が得られており、6.1 節の実験で推定した α と N-gram 言語モデルの次数 N が機能していることが分かる。提案手法は対訳コーパスの変更に対しても翻訳品質を低下させることなく機能することが分かった。ただし、ロイター対訳コーパスを用いた実験に比べると改

表 2 学習用データと評価データの変更に対する提案手法の頑健性 (データ選択時の重み $\alpha = 2$, 前編集システムでは単語 3-gram 言語モデルを学習)

Table 2 Robustness of the proposed method on different corpora ($\alpha = 2$, 3-gram).

	読売新聞	ロイター
前編集なし	3.09	3.57
前編集あり	3.13	3.76

表中の数字は最終的に得られた英文の翻訳品質 (NIST スコア)

表 3 読売新聞日英コーパスを用いた前編集実験での前編集パラメータ (選択重み α , 言語モデルの次数) ごとの結果

Table 3 Pre-editing results on Yomiuri corpus experiment with several parameters.

選択重み α	言語モデル		
	3-gram	4-gram	5-gram
0	3.04	2.96	2.94
1	3.08	3.00	2.95
2	3.13	3.03	2.99
3	3.11	3.07	3.00

表中の数字は最終的に得られた英文の品質 (NIST スコア)

善は小さかった。これはロイター対訳コーパス中には全体的に似た文が多く、評価データと学習用データも比較的近かったためと考えられる。

6.3 機械翻訳システムの変更に対する提案手法の頑健性

提案手法は、ある MT システムのための統計的前編集を自動構築する、MT システム非依存の手法である。次に、この対象とする MT システムの変更に対する提案手法の頑健性を調べた。具体的には、ロイター対訳コーパスを用いた日英翻訳タスクにおいて 4 種類の日英 MT (詳細は表 7 に示されている) で評価を行った。ここでも、6.1 節の実験で最も高い翻訳品質が得られたデータ選択時の重み $\alpha = 2$ および単語 3-gram 言語モデルを使用した。

結果を表 4 および表 5 に示す。MT_A は、6.1 節の実験で用いたものと同一である。MT_B から MT_D の実験については、パラメータの設定、具体的には α の値および使用する言語モデルの種類の設定に関して、オープン環境である。前編集なしとの比較 (表 4) では 4 種類すべての日英 MT システムで前編集の効果 (平均 NIST の向上) がみられた。提案手法により前編集を行うことで品質の高い翻訳結果が得られたかについて符号検定を行ったところ、3 種類の MT システムにおいて有意である (有意水準 1%) ことが分かった。また、前編集用の学習データの選択によって品質の高い翻訳結果が得られたかについても符号検定を行ったところ (表 5)、3 種類の MT システムにおいて有意であることが分かった (有意水準 1% および 5%)。なお、MT_D については、前編集を行わない場合と提案手法による前編集を行った場合 (表 4)、およびデータ選択をし

表 4 機械翻訳システムの変更に対する提案手法の頑健性 (前編集なしとの比較)

Table 4 Robustness of the proposed method on other MT systems (Comparison of w/ and w/o pre-editing).

	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	3.65	3.57
前編集あり	3.76**	3.69**	3.77**	3.65

表中の数字は最終的に得られた英文の翻訳品質 (NIST スコア) 符号検定 **: $p < 0.01$, *: $p < 0.05$

表 5 機械翻訳システムの変更に対する提案手法の頑健性 (学習データ選択の効果)

Table 5 Robustness of the proposed method on other MT systems (Effects of training data selection).

前編集用学習データ	日英機械翻訳システム			
	MT_A	MT_B	MT_C	MT_D
選択なし	3.64	3.63	3.70	3.70
選択あり	3.76**	3.69*	3.77*	3.65

表中の数字は最終的に得られた英文の翻訳品質 (NIST スコア) 符号検定 **: $p < 0.01$, *: $p < 0.05$

表 6 ロイター日英コーパスを用いた前編集実験での前編集パラメータ (選択重み α , 言語モデルの次数) ごとの結果

Table 6 Pre-editing results on Reuters corpus experiment with several parameters.

MT_B			
選択重み α	言語モデル		
	3-gram	4-gram	5-gram
0	3.63	3.60	3.60
1	3.72	3.68	3.68
2	3.69	3.65	3.65
3	3.11	3.59	3.59
MT_C			
選択重み α	言語モデル		
	3-gram	4-gram	5-gram
0	3.72	3.66	3.66
1	3.76	3.72	3.69
2	3.77	3.67	3.65
3	3.71	3.66	3.70
MT_D			
選択重み α	言語モデル		
	3-gram	4-gram	5-gram
0	3.70	3.65	3.68
1	3.76	3.67	3.70
2	3.65	3.65	3.66
3	3.63	3.64	3.63

表中の数字は最終的に得られた英文の品質 (NIST スコア)

ない場合と提案手法で選択した場合 (表 5) の間には有意な差がない (悪影響がない) ことが分かった. 提案手法は, MT システムの変更にも頑健であることが分かった.

また, 表 6 に, MT_B から MT_D の実験でのデータ選

表 7 用いた MT システムの詳細

Table 7 Specification of MT systems.

学習データ候補生成のために用いた 5 つの英日 MT システム

ATLAS V12.0 L10
MED-TranSer 2008 Professional Version7.0
The 翻訳プロフェッショナル V11
PC-TranSer 翻訳 Studio 2009 Version16.00
コリヤ英和! 一発翻訳 2009 R01

評価に用いた 4 つの日英 MT システム

MT_A	ATLAS V12.0 L10
MT_B	MED-TranSer 2008 Professional Version7.0
MT_C	The 翻訳プロフェッショナル V11
MT_D	PC-TranSer 翻訳 Studio 2009 Version16.00

択時の重み α および単語 N-gram 言語モデルの次数 N ごとの結果を示す (MT_A に対する結果は 6.1 節で述べており, 図 4 に示されている). いずれの MT システムに対しても $N = 3$ の場合に最も高い翻訳品質が得られ, α は 1 または 2 の場合に最も高い翻訳品質が得られている. 6.1 節の実験 (MT_A での実験) では, $N = 3$, $\alpha = 2$ が最適パラメータであり, 今回用いたすべての MT システムで最適パラメータがほぼ同じであることが分かった. なお, α の値については 1 と 2 の間の値が最適な可能性もある.

7. おわりに

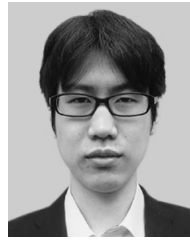
翻訳品質の向上を目的とした翻訳前の書き換えのための前編集システムを自動構築する方法の提案を行った. 具体的には, 対訳コーパスと機械翻訳システムから適切な前編集用学習データを自動生成し, 前編集システムを自動構築する手法を提案した. 提案する適切なデータ選択アルゴリズムの有効性を示し, 提案手法が MT システムの変更にも頑健であることを示した.

今後は, 本手法の他の言語ペアへの適用や, 前編集失敗判定の高精度化, 複数の前編集システムの構築とそれらの組合せ利用について研究を行いたい.

参考文献

- [1] 白井 諭, 池原 悟, 河岡 司, 中村行宏: 日英機械翻訳における原文自動書き替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).
- [2] 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).
- [3] Xu, Y. and Seneff, S.: Two-Stage Translation: A Combined Linguistic and Statistical Machine Translation Framework, *Proc. 8th Conference of the Association for Machine Translation in the Americas*, pp.222–231 (2008).
- [4] Sun, Y., O'Brien, S., O'Hagan, M. and Hollowood, F.: A Novel Statistical Pre-Processing Model for Rule-Based Machine Translation System, *Proc. 14th Annual Conference of the European Association for Machine Transla-*

- tion (2010).
- [5] 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が発見, 言語処理学会第 13 回年次大会講演論文集, pp.919–922 (2007).
 - [6] Zhao, S., Wang, H., Lan, X. and Liu, T.: Leveraging Multiple MT Engines for Paraphrase Generation, *Proc. 23rd International Conference on Computational Linguistics*, pp.1326–1334 (2010).
 - [7] Utiyama, M. and Isahara, H.: Reliable Measures for Aligning Japanese-English News Articles and Sentences, *Proc. 41st Annual Meeting of the Association for Computational Linguistics*, pp.72–79 (2003).
 - [8] Zhang, Y., Vogel, S. and Waibel, A.: Interpreting BLEU/NIST scores: How much improvement do we need to have a better system, *Proc. Language Resources and Evaluation*, pp.2051–2054 (2004).
 - [9] Doddington, G.: Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics, *Proc. 2nd International Conference on Human Language Technology Research*, pp.138–145 (2002).
 - [10] Hoang, H. and Koehn, P.: Design of the Moses Decoder for Statistical Machine Translation, *Software Engineering, Testing, and Quality Assurance for Natural Language Processing*, pp.58–65 (2008).
 - [11] Och, F. and Ney, H.: A Systematic Comparison of Various Statistical Alignment Models, *Computational Linguistics*, Vol.29, No.1, pp.19–51 (2003).
 - [12] 松本裕治: 形態素解析システム「茶筌」(〈特集〉使いやすくなった自然言語処理のフリーソフト: 知っておきたいツールの中身), 情報処理, Vol.41, No.11, pp.1208–1214 (2000).
 - [13] 松本裕治, 北内 啓, 山下達雄, 平野善隆, 松田 寛, 高岡一馬, 浅原正幸: 日本語形態素解析システム『茶筌』version 2.2.1 使用説明書, 奈良先端科学技術大学院大学松本研究室 (2000).
 - [14] Federico, M. and Cettolo, M.: Efficient Handling of N-gram Language Models for Statistical Machine Translation, *Proc. 2nd Workshop on Statistical Machine Translation*, pp.88–95 (2007).



山本 祐司

2010 年龍谷大学理工学部情報メディア学科卒業。2012 年同大学院修士課程修了。



吉見 毅彦 (正会員)

1987 年電気通信大学大学院計算機科学専攻修士課程修了。1999 年神戸大学大学院自然科学研究科博士課程修了。(財)計量計画研究所(非常勤), シャープ(株)を経て, 2003 年より龍谷大学理工学部勤務。2004 年より 2008 年まで(独)情報通信研究機構専攻研究員を兼任。



南條 浩輝 (正会員)

1999 年京都大学工学部情報学科卒業。2001 年同大学院情報学研究科修士課程修了。2004 年同大学院情報学研究科博士後期課程修了。同年龍谷大学理工学部助手。2007 年同助教。音声言語処理, 特に音声認識・理解に関する研究に従事。日本音響学会, 電子情報通信学会, IEEE, ISCA 各会員。2008 年度日本音響学会粟屋潔学術奨励賞受賞。