

## 翻訳自動評価法を用いた文アライメントの提案

安田 圭志<sup>†,‡</sup> 隅田 英一郎<sup>†,‡</sup>

<sup>†</sup> 情報通信研究機構音声言語グループ

<sup>‡</sup> ATR 音声言語コミュニケーション研究所

〒619-0288 「けいはんな学研都市」 光台二丁目 2 番地 2

E-mail: † {keiji.yasuda, eiichiro.sumita}@nict.go.jp

### 1. はじめに

コーパスベース翻訳技術の研究・開発において、対訳コーパスが不可欠である。特に、近年活発に研究・開発が行われている統計翻訳技術[1]においては、文単位で対応付けられた対訳コーパスが不可欠である。

従来より、コンパラブルコーパスから対訳文を自動抽出する文アライメントの研究[2,3]がさかんに行なわれているが、これらの手法は、対訳辞書や対訳シソーラスを必要とすることから、新たなタスクや言語対に対するポータビリティが低いという問題がある。

本研究ではこれらの問題を解決するため、統計翻訳技術と翻訳自動評価技術を応用して、文アライメントを行なう方法を提案する。

以下では、2 で提案手法について説明し、3 で実験結果について述べる。最後に 4 で論文を結ぶ。

### 2. 提案手法

図 1 に提案手法の処理の流れを示す。提案手法では、以下の 4 つの処理により、文アライメントを行なう。

1. 既存の対訳コーパスを用いて、統計翻訳システムの学習を行なう。(4からの繰り返し処理の場合は、4で得られたコーパスも学習に用いる。)
2. コンパラブルコーパスの各文を、1で得られた統計翻訳システムにより翻訳する。

3. 目的言語側のコンパラブルコーパスの各文と、目的言語へ機械翻訳された各文に対して、文単位の翻訳自動評価値を計算する。原言語側においても同様に自動評価値を計算する。

4. 3で得られた翻訳自動評価値をもとに文アライメントを行なう。ここでは、自動評価値が閾値よりも高くなれば、対訳文同士であるとみなす。

→1へ戻る。

提案手法では、1~4の処理を繰り返すことにより、翻訳システムの性能と、対訳コーパスの質とを同時に改善することが可能であると考えられる。しかしながら、本手法においては、いくつかの問題点がある。

第一の問題点は、計算量である。コンパラブルコーパスのサイズが大きくなると、ステップ 2 ならびにステップ 3 における計算量が非常に大きくなる。これらの問題点については、大型計算機[4]を用いることにより解決する。

第二の問題点は、BLEU[5]等に代表される翻訳自動評価値が文アライメントのスコアとして適当であるかという点である。この点については、3 に述べる実験において検証する。

### 3. 実験

翻訳自動評価値の文アライメントへの応用について、有効性を確認するため、図 2 に示す実験を行なった。

1. 既存の対訳コーパスを用いて統計翻訳

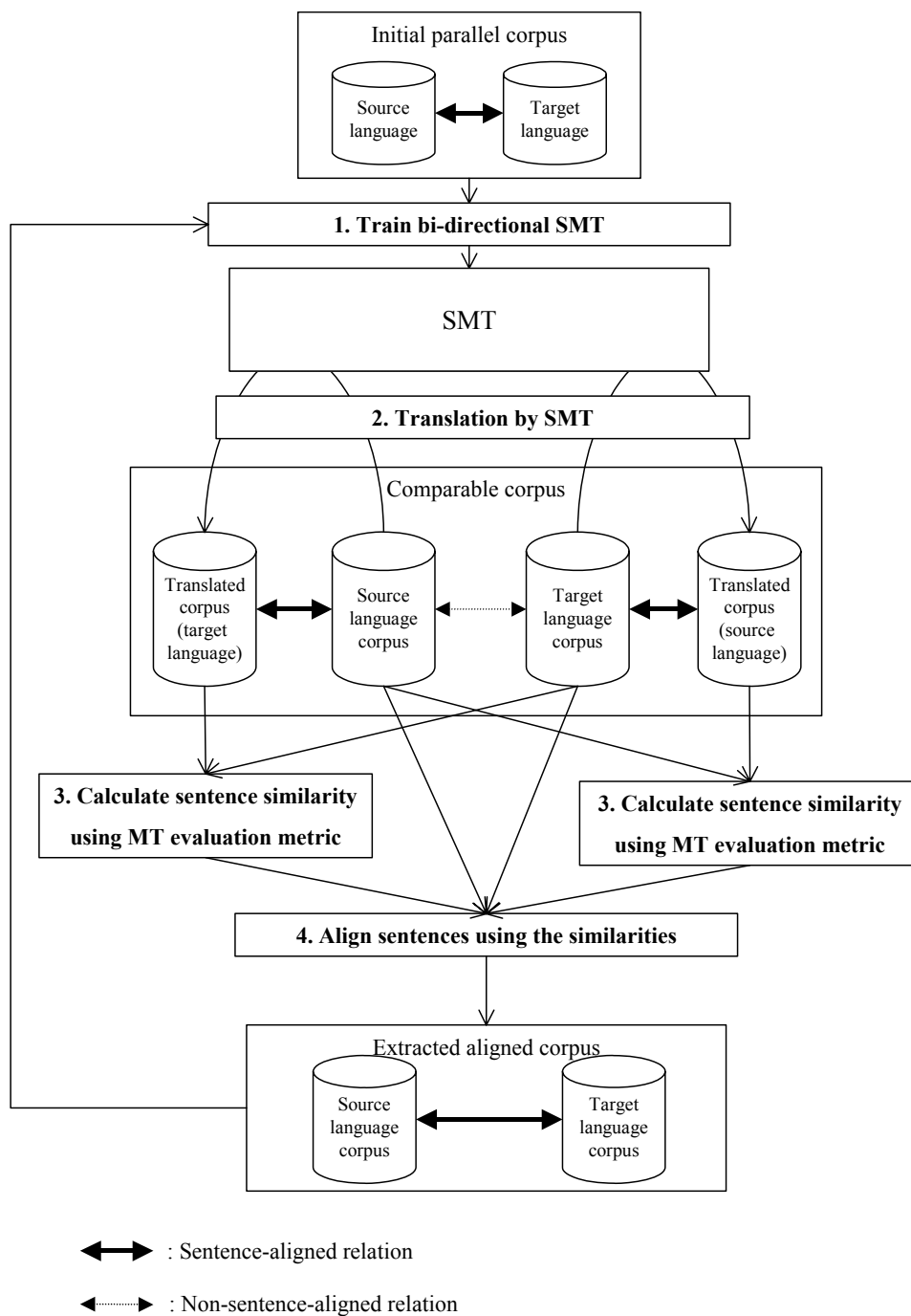


図1 提案手法の処理の流れ

- システムを学習。
2. 1 で得られた統計翻訳システムにより、1 の学習に用いた対訳コーパスを翻訳する。
  3. 既存の対訳コーパスにおいて対応付けられた文同士で、文単位の BLEU スコアを計算する。

4. 3 で得られた BLEU スコアを、文アライメントの信頼性基準として用い、データの取捨選択を行なう。
5. 4 で得られた対訳文のみを用いて、統計翻訳システムを再学習する。

翻訳自動評価値を用いた文アライメント性

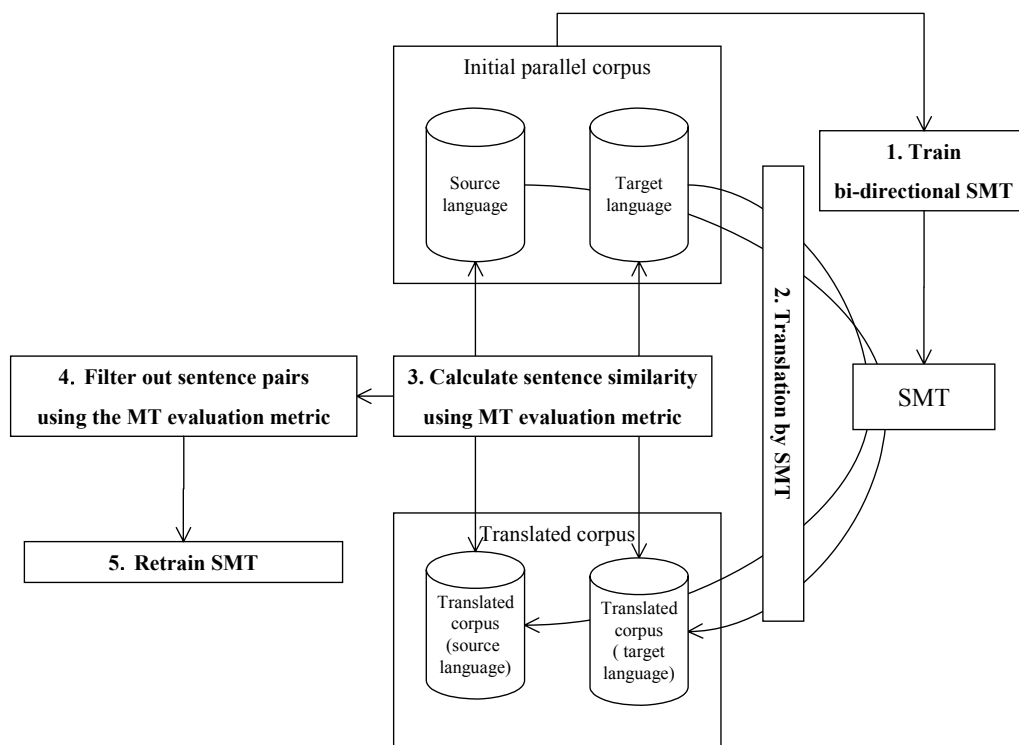


図 2 実験の処理の流れ

能を評価するため、1 で得られたシステムと、5 で得られたシステムとを比較する。もし、翻訳自動評価値を文アライメントに応用することが有効であれば、システム性能を劣化させることなく、学習コーパスのサイズを縮小されることが可能となる。

### 3.1 実験条件

実験においては、2 種類の対訳コーパスを用いた。

第一のコーパスは、日英特許コーパス[6]で、次に述べる手順により文アライメントを行なった。あらかじめ記事単位で対応づけられた日英の特許において、抄録部分のみを抽出する。日英のそれぞれの抄録における文数が同じであれば、各文が順番に翻訳されていると仮定し、それぞれの文を対応付ける。日英の抄録において、文数が異なる場合は、その抄録は採用しない。

日英特許コーパスを用いた実験では、国際特許分類 G6 に属する特許から上記の方法で抽出した 740,000 文を用いた。

第二のコーパスは、JENAAD コーパス[4].

150,000 文である。

実験において、翻訳モデルの学習には、pharaoh training toolkit を、言語モデルの学習には SRI language model tool kit を用いた。

### 3.2 実験結果

図 3 に日英特許コーパスを用いた実験の結果を示す。図 3 において、縦軸は BLEU スコアを表す。この BLEU スコアは、学習に用いられていないテストセット 500 文に対するテストセット単位の BLEU スコアである。横軸は文単位の BLEU スコアをもとに取捨選択した学習セットのサイズを表す。図中の破線は、全てのデータ(740,000 文対)を学習に用いた結果である。

図 3 を見ると、300,000 文程度、学習セットから削除しても、全てのデータを用いた場合と同等の性能が得られていることが分かる。

表 1 は、JENAAD コーパスを用いた実験の結果である。表 1 に示された BLEU スコアは、図 3 の縦軸と同様、500 文からなるテストセットに対するスコアである。

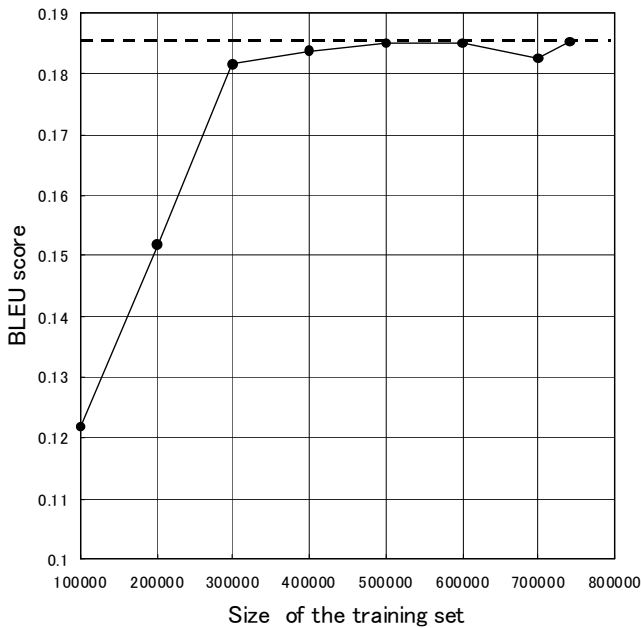


図3 日英特許コーパスを用いた  
実験結果

表1のBaselineは、従来法[4]により得られる文アライメントスコアを用いて学習文の取捨選択を行なった結果である。表1を見ると、BLEUスコアを用いて学習文50,000文を選択した場合では、Baselineよりもシステム性能が劣るものの、100,000文を選択した場合では、BLEUスコアを用いて学習文を選択した方が、高い翻訳性能が得られている。

これらのことにより、翻訳自動評価値を用いた文アライメントは有効であり、従来法と同等の性能が得られることが示唆された。

#### 4. まとめ

統計翻訳技術と翻訳自動評価法を応用した文アライメント法を提案し、翻訳自動評価法を文アライメント指標として適切であるかを検証する実験を行なった。

JENAADコーパスを用いた実験の結果、BLEUを用いた文アライメントは、対訳ソース等言語資源を必要とする従来法と同等の程度の性能が得られることが明らかとなった。

表1 JENAADコーパスを用いた実験結果

Method	# of training sentence pairs	BLEU score
Proposed method	50000	0.1019
Proposed method	100000	0.1090
Baseline	50000	0.1069
Baseline	100000	0.1057
Baseline	150000	0.1092

#### 文 献

- [1] Brown, P. F., Cocke, J., Della Pietra, S. A., Della Pietra, V. J., Jelinek, F., Lafferty, J. D., Mercer, R. L., Roossin, P. S. "A statistical approach to machine translation," Computational Linguistics Vol. 16, Issue 2 (June 1990) MIT Press Cambridge, MA, USA, pp. 79 – 85, 1990.
- [2] Masao Utiyama and Hitoshi Isahara, "Reliable Measures for Aligning Japanese-English News Articles and Sentences," Proc. of ACL-2003, pp. 72--79.
- [3] Xiaoyi Ma, "Champollion: A Robust Parallel Text Sentence Aligner," Proc. of LREC 2006.
- [4] <http://pr.fujitsu.com/en/news/2002/08/22.html>
- [5] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu, "Bleu: a Method for Automatic Evaluation of Machine Translation," Proc. of the ACL-2002, pp. 311—318.
- [6] Masao Utiyama and Hitoshi Isahara, "A Japanese-English Patent Parallel Corpus," MT-Summi XI.