

フレーズベース統計翻訳と対訳文一般化による翻訳規則の統合

寺島 涼[†] 越前谷 博^{††} 荒木 健治[†]

Ryo Terashima Hiroshi Echizen-ya Kenji Araki

[†] 北海道大学大学院情報科学研究科

Graduate School of Information Science and Technology, Hokkaido University

^{††} 北海学園大学工学部

Faculty of Engineering, Hokkai-Gakuen University

1 はじめに

近年、統計的機械翻訳手法においては、フレーズベース統計翻訳 [1][2][3] の研究が主流となっている。このフレーズベース統計翻訳では、フレーズごとに翻訳を行うため、翻訳文を生成するという点においては非常に頑健性が高くなる。しかし、文の構造を保持してはいないため、不自然な翻訳文を生成することがある。これは、フレーズベース統計翻訳が言語モデルとして N-gram モデルを使用しているためと考えられる。N-gram モデルは、局所的な言語モデルであるため、文全体の構造を保持するには不十分である。

一方、対訳文一般化による翻訳規則を用いた翻訳手法 [4][5] では、対訳文を部分的に変数と置き換えることで、文の構造を保持した翻訳規則を獲得する。翻訳規則は文の構造を有しているため、それぞれの翻訳規則における汎用性は高くないが、正しい翻訳規則においては、流暢な翻訳文を生成する可能性が高くなると考えられる。

そこで本稿では、フレーズベース統計翻訳を基に、対訳文一般化による翻訳規則を組み合わせた方式を提案する。フレーズベース統計翻訳において、翻訳対象文に対して類似した対訳文が学習データの中に存在する場合、その対訳文を部分的に一般化した翻訳規則を獲得し利用することで、自然な訳文を生成する可能性が高くなると考えられる。

性能評価実験の結果、フレーズベース統計翻訳に対して、対訳文を一般化した翻訳規則を組み合わせることにより、翻訳精度が向上することが明らかとなった。

2 対訳文一般化による翻訳規則

2.1 システム概要

Moses[1], GIZA++[2], SRILM (SRI Language Modeling Toolkit) [3] を用いて、フレーズベース統計翻訳システムを構築する。次に、対訳文一般化による翻訳規則を用いた処理を組み合わせる。本処理は、翻訳対象文に対し文の構造が類似した対訳文が存在した場合にのみ行われ、それ以外の場合は通常のフレーズベース統計翻訳システムの翻訳結果を利用する。また、対訳文一般化による翻訳規則の獲得と利用は、統計情報に基づいて行うため、解析的な知識やツールを必要としない。

入力文: Hello, this is Tanaka speaking .



(1) 類似する対訳文の選択

入力文: Hello, this is Tanaka speaking .対訳文 (Hello, this is Richard speaking . ;

こちらは/は/リチャード/です/.)



差異部分: Richard

(2) 対訳文の一般化

	こ	ち	は	リ	チャ	ード	で	す	.
Richard									

■: GIZA++により日英・英日両方向でアラインメントされた箇所



部分対応: (Richard; リチャード)

翻訳規則: (Hello, this is @0 speaking . ;

こちらは/は/@0/です/.)

(3) 翻訳規則の適用

(Hello, this is Tanaka speaking . ;

こちらは/は/Tanaka/です/.)

(4) 未翻訳部分の翻訳

翻訳結果: こちらは田中です。

図 1: 翻訳処理の具体例

2.2 処理過程

対訳文一般化による翻訳規則を利用した翻訳処理の過程を図 1 に基づき述べる。図 1 では、原言語として英語、目的言語として日本語を用いている。

1. 類似する対訳文の選択

入力文に対して原言語文が類似した対訳文を選択する。現状の本システムでは、より類似性の高い対訳文のみを対象とするために、入力文と対訳文の原言語文において差異部分が 1 箇所の場合、それらを類似する文と位置づけている。その際、差異部分は複数の単語で構成されていても良いが、単語数による制約

を与えている．許容する差異部分の個数や，単語数による制約については，今後，更に検討を続ける予定である．図 1 では，“Hello, this is Tanaka speaking.” が入力された場合，差異部分が “Richard” のみとなる対訳文 (Hello, this is Richard speaking.; こちら/は/リチャード/です/) を選択する．

2. 対訳文の一般化

差異部分を構成する全ての単語に対し対応する単語を，選択された対訳文の目的言語文の中から探索する．その際には，GIZA++により日英・英日の両方向でアラインメントされた単語を対応する単語と位置づける．図 1 では，差異部分を構成する単語 “Richard” と日英・英日の両方向でアラインメントされた単語 “リチャード” を対応する単語として選択する．更に部分対応 (Richard; リチャード) を一般化することで，翻訳規則 (Hello, this is @0 speaking.; こちら/は/@0/です/) を得る．

3. 翻訳規則の適用

獲得した翻訳規則を入力文に適用する．その結果，翻訳規則の変数部以外の翻訳が完了する．図 1 では，入力文 “Hello, this is Tanaka speaking.” に対して翻訳規則 (Hello, this is @0 speaking.; こちら/は/@0/です/) を適用することで “こちら/は/Tanaka/です/.” を生成する．

4. 未翻訳部分の翻訳

翻訳規則を適用することで生成された，翻訳途中の文の未翻訳部分に対して，フレーズベース統計翻訳システムを用いて翻訳を行う．図 1 では，最終的な翻訳結果として “こちらは田中です/.” を出力する．

上記の処理により翻訳規則が獲得されなかった場合は，フレーズベース統計翻訳システムの翻訳結果をそのまま利用する．性能評価実験では，評価データに対する全ての翻訳結果の内，44.7%において，フレーズベース統計翻訳システムの翻訳結果がそのまま利用された．

3 性能評価実験

実験データには『ひとり歩きの英語自遊自在』(日本交通公社出版事業局 1996) などの旅行用英会話テキスト 10 冊に掲載されている英日の対訳文 1,710 組を使用した．そのうち 1,368 組の対訳文を学習データとし，他の 342 組の対訳文を評価データとした．

フレーズベース統計翻訳システムの構築には Moses, GIZA++, SRILM を用いた．このように構築した機械翻訳システムをベースラインとし，更に，対訳文一般化による翻訳規則を取り入れた機械翻訳システムを提案システムとする．

3.1 評価方法

システムが生成した翻訳結果に対して人手による一対比較の評価と自動評価を行った．

3.2 一対比較による評価

人手による一対比較の評価では，ベースラインの翻訳結果と提案システムの翻訳結果を 1 文ずつ比較し，以下の 3 つに分類した．

- (1) 提案システムの方が良い翻訳である
- (2) ベースラインの方が良い翻訳である
- (3) 両システムの翻訳が同程度である

提案システムのベースラインに対する相対評価の値を以下の式 (1) で表す．

$$\text{相対評価の値} = \frac{\text{改善文数} - \text{改悪文数}}{\text{評価データの文数}} \quad (1)$$

式 (1) における改善文数と改悪文数はそれぞれ，提案システムの方が良い翻訳であった文数と，ベースラインの方が良い翻訳であった文数を示している．

3.3 自動評価

自動評価システムには，BLEU[6]，IMPACT[7]，METEOR[8] を用いた．これらの自動評価はスコアが高いほど良質な翻訳となり，スコアは 0.0～1.0 の値を取る．参照訳は翻訳結果 1 文に対して 4 つ用いている．

3.4 実験結果

表 1 に人手による一対比較の結果を示す．評価データ 342 文の内，提案手法により改善された文数は 28，逆に改悪された文数は 14 となった．その結果，ベースラインに対する相対評価は +4.09% となった．また，表 2 に自動評価の結果を示す．BLEU，IMPACT，METEOR において，提案システムのスコアがベースラインのスコアを 0.008～0.016 上回った．この結果は，対訳文一般化による翻訳規則を用いることで，より翻訳品質の高い文が増加したためである．

表 1: 一対比較の実験結果

改善された文の数	28
改悪された文の数	14
ベースラインに対する相対評価	+4.09%

表 2: 自動評価の実験結果

	BLEU	IMPACT	METEOR
提案システム	0.331	0.580	0.600
ベースライン	0.322	0.572	0.584

表 3: 翻訳結果の具体例

入力文	I'd like a room with shower.
提案システム	シャワー付きの部屋をお願いします.
ベースライン	シャワーの部屋をお願いします.

3.5 考察

表 3 に、提案システムとベースラインが生成した翻訳結果の具体例を示す。これは、提案システムがベースラインより質の高い翻訳を行った例である。入力文 “I'd like a room with shower.” に対し、提案システムは、対訳文 (I'd like a room with a balcony.; バルコニー/付き/の/部屋/を/お願い/し/ます/。) を選択した。次に、選択された対訳文において、部分対応 (a balcony; バルコニー) を一般化することにより、翻訳規則 (I'd like a room with @0.; @0/付き/の/部屋/を/お願い/し/ます/。) を獲得した。最後に、“shower” の訳語として “シャワー” を得て、翻訳結果として “シャワー付きの部屋をお願いします。” を生成した。この翻訳結果は、ベースラインが生成した翻訳結果 “シャワーの部屋をお願いします。” に対し、より自然で正しい良質な翻訳結果となっている。

改善された 14 の文については、誤った翻訳規則の獲得と部分翻訳の失敗が主な原因であった。現在のシステムは、対訳文一般化による翻訳規則を用いて翻訳結果を生成した場合、その翻訳結果を優先して利用している。しかし、本システムの基となるフレーズベース統計翻訳システムが、より質の高い翻訳を行った場合は、そちらの翻訳結果を優先するべきである。よって今後は、それぞれの翻訳処理が生成した翻訳結果から、より良質な翻訳結果を選択することが重要となる。これは今後の課題である。

4 おわりに

本研究では、フレーズベース統計翻訳を基に、対訳文一般化による翻訳規則を組み合わせることで翻訳精度の向上を図った。性能評価実験の結果、一対比較による評価では、ベースラインに対して 4.09% の対訳文について改善され、自動評価 BLEU, IMPACT, METEOR においてもスコアが 0.008–0.016 向上した。これらの結果より、フレーズベース統計翻訳に対訳文一般化による翻訳規則を組み合わせることで翻訳精度が向上することを確認した。

今後は、翻訳処理における最適な翻訳結果選択の方法の検討と、大規模なデータを用いた性能評価実験を行う予定である。

参考文献

- [1] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin and E. Herbst: Moses: Open Source Toolkit for Statistical Machine Translation, Proceedings of ACL 2007, pp.177-180 (2007).
- [2] Och, F.J. and Ney, H.: A Systematic Comparison of Various Statistical Alignment Models, Computational Linguistics, vol.29, No.1, pp.19-51 (2003).
- [3] Andreas Stolcke: SRILM - an Extensible Language Modeling Toolkit, 7th International Conference on Spoken Language Processing, pp.901-904 (2002).
- [4] 寺島 涼, 越前谷 博, 荒木 健治: 対訳コーパスに基づく学習型機械翻訳における省略可能情報を用いた部分対応学習の有効性, 電子情報通信学会論文誌 D, Vol.J93-D, No.3, pp.377-388 (2010).
- [5] Manabu Sasayama, Fuji Ren, Shingo Kuroiwa: Automatic Super-Function Extraction for Translation of Spoken Dialogue, Natural Language Processing and Knowledge Engineering 2007, pp.141-148 (2007).
- [6] Kishore Papineni, Salim Roukos, Todd Ward, and Wei Jing Zhu: BLEU: a Method for Automatic Evaluation of Machine Translation, Proceedings of ACL 2002, pp.311-318 (2002).
- [7] Hiroshi Echizen-ya, Kenji Araki: Automatic Evaluation of Machine Translation based on Recursive Acquisition of an Intuitive Common Parts Continuum, Proceedings of the Eleventh Machine Translation Summit (MT SUMMIT XI), pp.151-158 (2007).
- [8] Satanjeev Banerjee, Alon Lavie: METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments, Proceedings of the ACL 2005 Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization, pp.65-72 (2005).