

学習型機械翻訳におけるデータスパースネスの問題解消に向けた省略可能情報に基づく部分対応学習とその有効性

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

対訳コーパスに基づく機械翻訳手法の研究が盛んに行われている。そこでは、翻訳例の原言語文と目的言語文間における句レベルの部分対応を利用することが有効となる。しかし、対訳コーパスに基づく機械翻訳手法においては、データスパースネスに起因する翻訳精度の低下が問題となる。解析的アプローチ [1][2] では、品詞情報や構文情報などの解析的な知識に基づき、翻訳例における句レベルの部分対応を決定するが、解析ツールに依存するため様々な言語への適用が困難となる。統計的アプローチ [3] では、大量の対訳コーパスに基づく統計情報を用いて、翻訳例における句レベルの部分対応を決定するが、データスパースな対訳コーパスにおいては、適用が困難となる。

我々は、データスパースな対訳コーパスからも、翻訳例における句レベルの部分対応を効率よく決定するための学習手法である、省略可能情報に基づく部分対応学習手法を提案している [4]。本手法では、翻訳例中の省略可能情報に基づき対訳規則を自動獲得し、対訳規則を用いて翻訳例における句レベルの部分対応を効率よく決定する。本稿では、この対訳規則を抽出ルールと呼ぶ。

本手法を適用する際には、品詞情報や構文情報などの解析的な知識を用いていない。そのため本手法は汎用性の高い学習手法である。我々は、本手法を学習型機械翻訳手法 [5] に適用することにより、その有効性を確認している [4]。本稿では、より詳細な性能評価と、統計的機械翻訳手法との比較について述べる。

2 省略可能情報に基づく部分対応学習

省略可能情報に基づく部分対応学習の概要を図 1 に示す。対応関係として (a room; 部屋) が既に獲得されている場合、その単語が全て含まれている翻訳例である (I'd like a quiet room.; 静か/な/部屋/を/お願い/し/ます。) に基づき抽出ルールの獲得処理を行う。英文においては、“a” と “room” に挟まれた部分である “quiet” を省略可能な部分と位置づける。また、日本文においては、“quiet” に対応する部分である “静か/な” を省略可能な部分と位置づける。

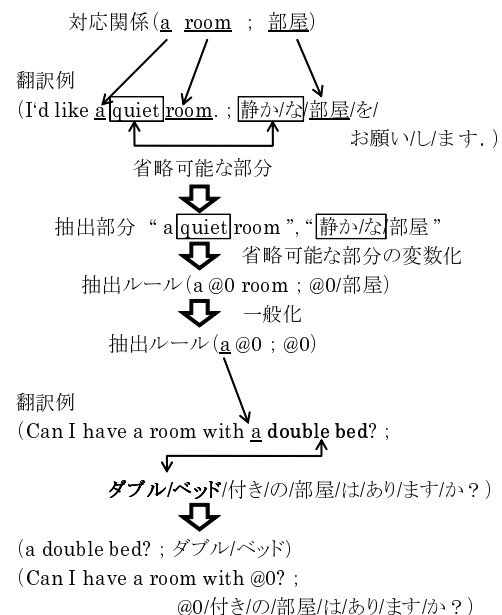


図 1: 省略可能情報に基づく部分対応学習の概要

次に、対応関係の部分と省略可能な部分で構成される “a quiet room” と “静か/な/部屋” において、省略可能な部分 “quiet” と “静か/な” を変数に置き換えることにより抽出ルール (a @0 room; @0/部屋) を獲得する。

この抽出ルールは翻訳例の原言語文中の “a” から “room” までの部分と、目的言語文中の “部屋” 及び “部屋” の左側に隣接する部分が句レベルで対応していることを表している。このようにして獲得された抽出ルールを、より一般化することで、さらに汎用的な抽出ルールを獲得する (a @0 room; @0/部屋) からは “room” と “部屋” を変数化することで、新たな抽出ルールとして (a @0; @0) を獲得する。この抽出ルールは原言語文中の “a” とその右側に隣接する部分で構成される句レベルの部分が、目的言語文中の句レベルの部分と対応することを表している。

続いて、抽出ルール (a @0; @0) を翻訳例 (Can I have a room with a double bed; ダブル/ベッド/付き/の/部屋/は/あります/か?) に適用することで、句レベルの部分対応 (a double bed; ダブル/ベッ

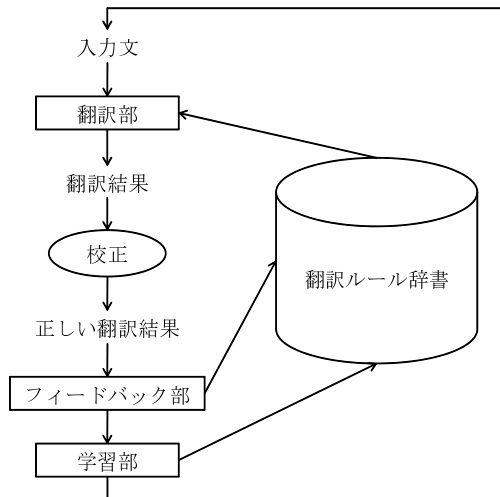


図 2: 学習型機械翻訳システムの概要

ド)を獲得する。さらに、翻訳例において (a double bed; ダブル/ベッド) を変数に置き換えた (Can I have a room with @0?; @0/付き/の/部屋/は/あり/ます/か?) も獲得することが可能となる。このように抽出ルールを適用することで、効率よくかつ正確に部分対応を決定することが可能となるため、データスパースネスに対しても有効であると考えられる。

3 学習型機械翻訳システムの概要

本手法を用いた学習型機械翻訳システムの概要を図 2 に示す。翻訳部では入力文に対し、翻訳ルール辞書中の翻訳ルールを用いて翻訳結果を生成する。生成された翻訳結果に誤りがある場合は、翻訳結果に対し校正を行う。

続いて、フィードバック部では、翻訳部で使用された翻訳ルールに対して評価を行う。学習部では、翻訳例対における共通部分・差異部分を多段階に抽出する帰納的学習と、さらに、2 で述べた省略可能情報に基づく部分対応学習を適用することで、翻訳ルールを獲得する。

本システムで扱う翻訳ルールは、部分翻訳ルールと文翻訳ルールに分類される。図 1 で獲得された (a double bed; ダブル/ベッド) のような、部分を示す対応関係を部分翻訳ルールと呼び (Can I have a room with @0?; @0/付き/の/部屋/は/あり/ます/か?) のように文の構造を持つ対応関係は文翻訳ルールと呼ぶ。学習部より獲得された翻訳ルールは、翻訳ルール辞書に登録される。

4 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翻訳例 1,710 組を使用した。そのうち 948 組の翻訳例を学習データとし、他の 762 組の翻訳例を評価データとした。学習データの英文、日本語の平均単語数はそれぞれ 6.7

表 1: 本手法を適用したシステムの実験結果

有効な翻訳率	内訳	
	未登録語なし	未登録語あり
55.91%(426)	73.47%(313)	26.53%(113)

表 2: 本手法を適用しないシステムの実験結果

有効な翻訳率	内訳	
	未登録語なし	未登録語あり
50.79%(387)	73.39%(284)	26.61%(103)

と 7.8 である。また評価データの英文、日本語の平均単語数はそれぞれ 6.4 と 7.4 である。

実験には学習型機械翻訳システムにおいて、本手法を適用したシステムと、適用しないシステムの 2 つを用いて比較を行った。実験条件は、初期状態として翻訳ルール辞書を空の状態にし、1 文入力されるごとに翻訳と学習を繰り返し行った。

4.1 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正誤評価と自動評価を行った。なお、システムが複数の翻訳結果を生成した場合は、最も優先順位の高い翻訳結果 1 つのみを評価対象とした。

4.2 人手による正誤評価

人手による正誤評価では、翻訳結果を有効な翻訳結果と無効な翻訳結果に分類した。有効な翻訳結果とは、以下の 2 つである。

- (1) 未登録語を含まない正しい翻訳結果
- (2) 未登録語を含む正しい翻訳結果

未登録語を含む正しい翻訳結果とは、未登録語が名詞、形容詞、及び名詞句に相当するもので、これらが未登録語の場合は、訳語を与えることにより未登録語を含まない正しい翻訳結果が容易に得られる翻訳結果である。また、無効な翻訳結果とは以下の 3 つである。

- (1) 未登録語を含まない誤った翻訳結果
- (2) 未登録語を含む誤った翻訳結果
- (3) 翻訳不能

翻訳不能とは、入力文に対し適用可能な文翻訳ルールが存在しない場合である。なお、翻訳結果における未登録語部分は原言語の単語で出力され、翻訳不能においては原文がそのまま出力される。

4.3 自動評価

自動評価システムには、BLEU, IMPACT, METEOR, ROUGE-L, NIST, mWER, mPER を用いた。BLEU, IMPACT, METEOR, ROUGE-L, NIST はスコアが高いほど良質な翻訳となり、また

表 3: 様々な自動評価によるスコア

	BLEU	IMPACT	METEOR	ROUGE-L	NIST	mWER	mPER
本手法を適用したシステム	0.355	0.549	0.611	0.560	6.17	0.555	0.460
本手法を適用しないシステム	0.329	0.519	0.602	0.525	5.79	0.585	0.507

mWER, mPER はスコアが低いほど良質な翻訳となる．参照訳は翻訳結果 1 文に対して 4 つ用いている．

4.4 実験結果

表 1, 表 2 に本手法を適用したシステムと適用しないシステムの有効な翻訳率を示す．本手法を適用することで，有効な翻訳率は 5.12 ポイント増加した．また，表 3 に様々な自動評価によるスコアを示す．今回使用した全ての自動評価において，本手法を適用することでスコアが向上した．

4.5 考察

4.5.1 本手法による学習能力の向上について

実験結果より，本手法を適用することで，無効な翻訳結果から有効な翻訳結果となったものが 44 文存在した．表 4 に 44 文の内訳を示す．

また，本手法を適用することで，有効な翻訳結果から無効な翻訳結果となったものが 5 文存在した．表 5 に 5 文の内訳を示す．

結果として，本手法を適用することにより有効な翻訳結果は 39 文増加した．これは本手法を適用することで，良質な翻訳ルールが増大したためである．

さらに，本手法を適用することで，出現頻度の低い部分対応に対する獲得能力が向上した．実験において，本手法を適用したシステムが獲得した部分対応の数は 997 であった．そのうち，出現頻度が 1 のみで獲得されたのは 658(66.0%)，出現頻度が 2 で獲得されたのは 194(19.5%)，さらに出現頻度が 3 以上で獲得されたのは 145(14.5%) であった．獲得された部分対応の半数以上が，出現頻度 1 のみで獲得されており，本手法の学習能力の高さを示す結果となった．

表 4: 本手法の適用による有効性

	文数
無効(未登録無)	5
有効(未登録無)	
無効(未登録有)	11
有効(未登録無)	
無効(未登録有)	14
有効(未登録有)	
無効(翻訳不可)	3
有効(未登録無)	
無効(翻訳不可)	11
有効(未登録有)	

出現頻度の低い部分対応を獲得することで，より効率的な翻訳ルールの獲得が可能となる．この結果は，本手法がデータスパースネスの問題に対して有効であることを示している．

4.5.2 統計翻訳との比較

学習の観点より翻訳ルールを自動獲得することで翻訳を行う本システムとはアプローチは異なるが，本システムと同様に対訳コーパスのみに基づき翻訳を行う統計翻訳との比較について述べる．統計翻訳システムの構築には GIZA++[3]，SRILM (SRI Language Modeling Toolkit) [6]，Moses[7] を用いた．実験は本システムと可能な限り同じ条件になるように，以下の手順で行った．

- (1) 本システムで用いた学習データと同様の翻訳例 948 組を学習データとする．
- (2) 本システムで用いた評価データと同様の翻訳例 762 組から 1 組を順に選択し，その原言語文を入力文として翻訳結果を得る．
- (3) (2) で選択した 1 組の翻訳例を学習データに加えたものを新たな学習データとする．そして選択された翻訳例は評価データから除外する．
- (4) (2) から (4) を評価データ中に翻訳例が存在しなくなるまで繰り返す．

統計翻訳の実験結果を表 6 に示す．本システムと比較した場合，表 1 より，本システムの有効な翻訳率が 7.22 ポイント上回った．また，その内訳に着目すると，未登録語を含まない正しい翻訳結果の個数は近い値であるが，未登録語を含む正しい翻訳結果の個数は，本システムが大きく上回った．

表 7 に，本システムと統計翻訳システムが生成した翻訳結果の具体例を示す．入力文 “Where can I catch the bus?” に対し，統計翻訳システムは，未登録語を含む誤った翻訳結果 “はどこです catch バスはありますか?” を生成した．一方，本システムでは，文翻

表 5: 本手法の適用による誤り

	文数
有効(未登録無)	2
無効(未登録無)	
有効(未登録有)	2
無効(未登録無)	
有効(未登録有)	1
無効(未登録有)	

表 8: 統計翻訳に対する自動評価によるスコア

	BLEU	IMPACT	METEOR	ROUGE-L	NIST	mWER	mPER
統計翻訳	0.419	0.615	0.669	0.638	7.33	0.493	0.315

表 6: 統計翻訳の実験結果

有効な翻訳率	内訳	
	未登録語なし	未登録語あり
48.69%(371)	83.02%(308)	16.98%(63)

表 7: 翻訳結果の具体例

入力文	Where can I catch the bus?
本手法を適用したシステム	バス乗り場はどこですか？
統計翻訳	はどこです catch バスはありますか？

訳ルール (Where can I catch @0?; @0/乗り場/は/どこ/です/か?) と部分翻訳ルール (the bus; バス) を適用することにより, 良質な翻訳結果 “バス乗り場はどこですか?” を生成した. 文翻訳ルール (Where can I catch @0?; @0/乗り場/は/どこ/です/か?) は, 翻訳例 (Where can I catch a taxi?; タクシー/の/乗り場/は/どこ/です/か?) に対して, 抽出ルール (a @0; @0) を適用し, 部分対応 (a taxi; タクシー) を決定することで獲得されている. このように, 本手法により獲得される抽出ルールを適用することで, より正確な部分対応が決定可能となり, 良質な翻訳ルールが効率よく獲得されるようになった.

表 8 に統計翻訳に対する自動評価の結果を示す. 本システムに対する自動評価である表 3 と比較すると, 統計翻訳システムが高いスコアを得ている. 自動評価は正誤評価とは異なり, 誤った翻訳結果においても, 部分的に正しい翻訳が存在すれば, スコアに加算される. 自動評価において, 統計翻訳手法が本システムより高いスコアを得た大きな要因としては, 本システムでは入力文に対し適用可能な文翻訳ルールが存在しない場合は翻訳不能となることが挙げられる.

学習に基づき対訳規則を自動獲得する本システムは, 大規模対訳コーパスを前提とした統計翻訳とはアプローチが異なるため, 今回の実験結果のみで優劣を一概に述べることはできない. しかし, 今回の実験結果は本システムが統計翻訳システムと比較して, 遜色のない学習能力及び翻訳精度を有することを示していると考えられる.

5 おわりに

本稿では, 我々が提案する省略可能情報に基づく部分対応学習の性能評価について述べた. 本手法を

学習型機械翻訳手法に適用することで, 部分対応を獲得する能力が向上し, 良質な翻訳ルールをより効率よく獲得することが可能となった. その結果, 性能評価実験では, 人手による正誤評価, 自動評価双方において, 本手法を適用したシステムは本手法を適用しないシステムに対して向上した. 統計翻訳との比較においては, アプローチが異なることから一概に優劣を述べることはできないが, 本システムは統計翻訳に比べて遜色のない学習能力及び翻訳精度を有していると考えられる.

今後は, 省略可能情報に基づく部分対応学習の精度向上による, より高度な学習能力の実現を目指す. また, 他の言語を用いた実験や, 他のタスクへの応用を検討している.

謝辞 本研究の一部は, 平成 19 年度財団法人北海道科学技術総合振興センター「研究開発助成事業」の補助金によって行われた.

参考文献

- [1] 荒牧英治, 黒橋禎夫, 佐藤理史, 渡辺日出雄: 用例ベース翻訳のための対訳文の句アライメント, 自然言語処理, Vol.10, No.5, pp.75-92 (2003).
- [2] 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).
- [3] Och, F.J. and Ney, H.: A Systematic Comparison of Various Statistical Alignment Models, Computational Linguistics, vol.29, No.1, pp.19-51 (2003).
- [4] 寺島涼, 越前谷博, 荒木健治: 学習型機械翻訳手法における省略可能性を用いた翻訳ルールの自動獲得とその有効性, 情報処理学会研究報告, Vol.2008, No.4, pp.127-134 (2008).
- [5] 越前谷博, 荒木健治, 桃内佳雄, 柄内香次: 翻訳例に基づく再帰チェーンリンク型学習による機械翻訳手法, 電子情報通信学会論文誌 D, Vol.J85-D-II, No.12, pp.1840-1852 (2002).
- [6] Andreas Stolcke: SRILM - an Extensible Language Modeling Toolkit, Proceedings of the International Conference on Statistical Language Processing, pp.901-904 (2002).
- [7] 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, ACL'07 demonstration session, pp.177-180 (2007).