

統計的機械翻訳におけるフレーズ対応最適化を用いた 翻訳候補のリランキング

越川 満* 内山 将夫† 梅谷 俊治‡ 松井 知己§ 山本 幹雄*

*筑波大学大学院 システム情報工学研究科

†情報通信研究機構 MASTAR プロジェクト

‡大阪大学大学院 情報科学研究科

§中央大学 理工学部 情報工学科

1 はじめに

統計的機械翻訳におけるデコーダはモデルが与える確率を最大とする目的言語文を探索するが、探索空間が膨大であるため何段階かの簡化により近似的探索を行っている。フレーズベースモデル [1] では同一の翻訳文であってもフレーズ対応の違いによって確率値が異なるため、各候補文に対するフレーズ対応のうち厳密に確率最大の対応を求め、その確率を用いて翻訳候補を比較すべきであるが、このような厳密な比較は行われていない。一方で翻訳結果として採択されなかった第 2 位以下の候補には、より翻訳精度の高いものが含まれることが知られている [2]。そこで、これらの候補に対して確率を厳密に最大化するフレーズ区切り・対応を求め直すことで翻訳候補の順位付けを改善し、翻訳精度を向上させることができると思われる。

対訳文に対して与えられたフレーズテーブルから確率最大すなわち最適なフレーズ対応 (phrase alignment) を求める手法としては、整数計画法を用いた手法 [3] が提案されている。しかし、この手法では現在広く用いられているリオーダーリングモデルが考慮されていない。本稿では、リオーダーリングモデルを組み込み可能な新たなフレーズアライナ (phrase aligner) の定式化を提案し、これを統計的機械翻訳システムの翻訳候補のリランキングに応用することで翻訳精度が向上することを実験的に示す。

2 フレーズアライナ

2.1 統計的機械翻訳

統計的機械翻訳システムは、入力文 \mathbf{f} に対する翻訳候補としてあらゆる目的言語文 \mathbf{e} を考慮し、最も確率の高い目的言語文 $\hat{\mathbf{e}}$ を翻訳結果として出力する。すなわち、以下の式に基づいて翻訳候補 $\hat{\mathbf{e}}$ は決定される [1]。

$$\begin{aligned} \hat{\mathbf{e}} &= \arg \max_{\mathbf{e}} \sum_{\mathbf{c}} P(\mathbf{f}, \mathbf{c}|\mathbf{e})P(\mathbf{e}) \\ &\approx \arg_{\mathbf{e}} \max_{\mathbf{e}} \max_{\mathbf{c}} P(\mathbf{f}, \mathbf{c}|\mathbf{e})P(\mathbf{e}) \end{aligned} \quad (1)$$

式 (1) において、 $P(\mathbf{e})$ は言語モデルを、 \mathbf{c} は \mathbf{f} 、 \mathbf{e} 間での単語あるいはフレーズの対訳関係を表す。また $\arg_{\mathbf{e}} \max_{\mathbf{e}} \max_{\mathbf{c}}$ はデコーダと呼ばれ、与えられた \mathbf{f} の翻訳として最も確率の高い \mathbf{e} を探索する。フレーズ翻訳モデル [1] では、翻訳の最小単位として連続する単語列であるフレーズを用い、以下の近似をおく。

$$P(\mathbf{f}, \mathbf{c}|\mathbf{e}) \approx \left(\prod_{i=1}^I P(\bar{f}_i|\bar{e}_{c_i}) \right) \cdot P(\mathbf{c}|\mathbf{e}) \quad (2)$$

ここで、 \bar{f} 、 \bar{e} は各々原言語フレーズ、目的言語フレーズを、 $\mathbf{c} = c_1, c_2, \dots, c_I$ は原言語-目的言語間でのフレーズ対応関係を表す。 c_i は原言語側で i 番目のフレーズ \bar{f}_i が対応する目的言語側フレーズの番号である。すなわち \bar{f}_i は目的言語側で c_i 番目のフレーズ \bar{e}_{c_i} に対応する。 $P(\bar{f}_i|\bar{e}_{c_i})$ はフレーズ翻訳確率、 $P(\mathbf{c}|\mathbf{e})$ はリオーダーリングモデルとそれぞれ呼ばれる。リオーダーリングモデルは翻訳前後での語順変化に対する確率を与えるモデルであり、現在は目的言語側で隣り合うフレーズ同士の原言語側における位置関係を確率モデル化したものが広く用いられている。

2.2 フレーズ対応問題

対訳文およびフレーズ対集合が与えられたとき、対訳文の各単語を一度ずつ被覆するフレーズ対の組合せをその対訳文に対するフレーズ対応という [3]。フレーズ対応問題とは、対訳文 $\langle \mathbf{f}, \mathbf{e} \rangle$ 、およびフレーズ対とその翻訳確率 $\langle \bar{f}, \bar{e}, P(\bar{f}|\bar{e}) \rangle$ の集合であるフレーズテーブルが与えられたとき、確率最大すなわち最適なフレーズ対応を求める問題であり、これを実現するシステムをフレーズアライナと呼ぶ [3]。フレーズアライナは以下の式で定義される。

$$\langle \hat{f}_1^I, \hat{e}_1^I, \hat{\mathbf{c}} \rangle = \arg \max_{\substack{\bar{e}_1^I = \mathbf{e}, \\ \bar{f}_1^I = \mathbf{f}, \mathbf{c} = \mathbf{c}_1^I}} P(\bar{f}_1^I|\bar{e}_1^I, \mathbf{c}) P(\bar{e}_1^I, \mathbf{c}|\mathbf{e}) \quad (3)$$

ここで、 $\mathbf{c}_1^I = c_1, c_2, \dots, c_I$ であり \bar{f}_1^I 、 \bar{e}_1^I についても同様に定義される。式 (3) の右辺第一項および第二項は、それぞれ式 (2) のものと対応している。

2.3 整数計画問題としての定式化

フレーズテーブル中のフレーズ対 $\langle \bar{f}_k, \bar{e}_{c_k} \rangle$ に対してその使用の有無を表す 2 値変数 $x_k \in \{0, 1\}$ 、および各フレーズ対が原言語文で被覆する単語位置を 1、それ以外を 0 で表す 2 値行列 F を導入する。なお、 F は行が原言語文に含まれる単語で、列が各フレーズ対でインデクシングされている。目的言語側についても F と同様に行列 E を定義する。

対訳文 $\langle \mathbf{f} = f_1, f_2, f_3, f_4, \mathbf{e} = e_1, e_2, e_3 \rangle$ に対して図 1 のようなフレーズ対が適用可能な場合を考える。図 1 において四角い枠がフレーズを表し、フレーズ同士を結ぶ線がフレーズ対応である。このとき、 F, E は式 (4) のようになる。例えば、フレーズ対 1 は原言語側で f_1, f_2 を被覆するため、行列 F の 1 行 1 列および 2 行 1 列成分が 1 となっている。

$$F = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, E = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (4)$$

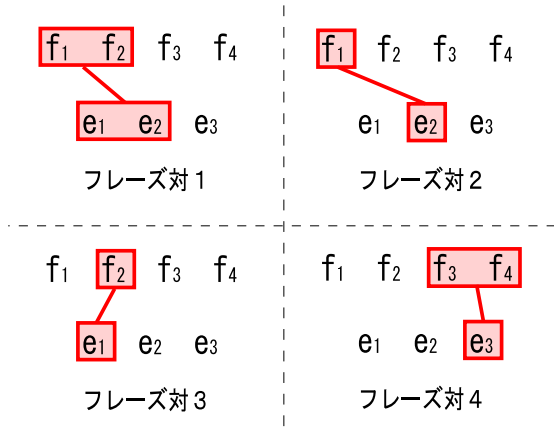


図1 フレーズ対集合

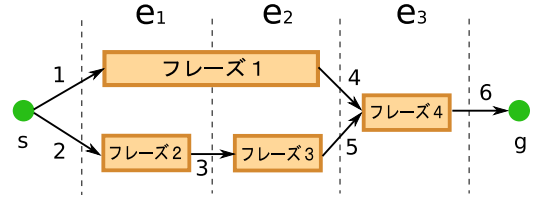


図2 有向グラフ上に表したフレーズ同士の関係

簡単化のため、式 (3) の右辺第二項すなわちリオーダーリング確率を 1 と置き、フレーズ翻訳確率のみを考慮すると仮定すれば、フレーズ対応問題は次のように表される [3]。

$$\begin{cases} \text{maximize} & \sum_{k \in K} x_k \log p_k \\ \text{subject to} & F\mathbf{x} = \mathbf{1}, \\ & E\mathbf{x} = \mathbf{1}, \\ & x_k \in \{0, 1\} \quad (\forall k \in K). \end{cases} \quad (5)$$

ここで p_k は $P(\bar{f}_k | \bar{e}_{c_k})$ の略記であり、 K は対訳文に適用可能なフレーズ対の集合を表す。また $\mathbf{1} = (1 \dots 1)^T$ である。

3 提案手法

3.1 有向グラフのパスとしてのフレーズ対応

[3] の定式化では各フレーズ対に対して変数を置くため、フレーズ対同士の位置関係を表すリオーダーリングモデルを一次式として目的関数に組込むのは難しい。そこで本節では、上記の問題を解決するフレーズ対応問題の新たな定式化を提案する。

本稿で提案する定式化では、対訳文に適用可能なフレーズ対を目的言語側について有向グラフ上に表すことを考える。図 1 のフレーズ対を有向グラフ上に表すと図 2 のようになる。図 2 において $e_1 \sim e_3$ は目的言語文中の単語である。フレーズは四角い枠で表されており、各フレーズの位置と大きさは被覆する単語位置に対応する。またフレーズの番号は図 1 のフレーズ対の番号に対応している。フレーズ同士を結ぶ有向枝に割り振られた数字は枝番号 a を表す。フレーズ対応は、原言語側について式 (5) の原言語側制約式 $F\mathbf{x} = \mathbf{1}$ を満たし、かつ目的言語側グラフにおいて開始ノード s から終端ノード g へのパスとなっている必要がある。例えば図 2 中、フレーズ 2, 3, 4 を通るパスは対訳文に対するフレーズ対応候補の一つである。フレーズ対応に含まれるフレーズ対は、目的言語側グラフのパスに含まれる枝からも導くことができる。このモデル化のもとでは、リオーダーリング確率を目的言語側グラフの枝に対する重みとして組込むことができる。

[3] の定式化で用いられている記号に加えて、目的言語側有向グラフ中の枝 a に対し、 a がパスに含まれる場合は 1、そうでない場合は 0 をとる仮変数 y_a を新たに導入する。このとき、

フレーズ対応問題は以下のように定式化される。

$$\begin{cases} \text{maximize} & \sum_{k \in K} x_k \log p_k + \sum_{a \in A} y_a \log d_a \\ \text{subject to} & F\mathbf{x} = \mathbf{1}, \\ & M\mathbf{y} = \mathbf{b}, \\ & N\mathbf{y} = \mathbf{x}, \\ & x_k \in \{0, 1\} \quad (\forall k \in K), \\ & y_a \in \{0, 1\} \quad (\forall a \in A). \end{cases} \quad (6)$$

ここで $M\mathbf{y} = \mathbf{b}$ は目的言語側グラフでノード s からノード g へのパスとなるための条件であり流量保存則と呼ばれる。 N は目的言語側パスに含まれる枝と各フレーズ対との関係を与える行列である。また A はすべての枝の集合を、 d_a は目的言語側の各枝に割り当てられるリオーダーリング確率を表す。

図 2 について、流量保存則を書き下すと式 (7) のようになる。左辺第一項は行列 M であり、各行は s 、フレーズ 1、...、フレーズ 4、 g のノードに対応する。例えば式 (7) の 5 行目は、フレーズ対 4 のノードについての流量保存則に対応し、 $y_4 + y_5 = y_6$ を表す。

$$\begin{pmatrix} -1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad (7)$$

また、図 2 のグラフにおいてフレーズ対 4 のノードには枝 4 ないしは枝 5 からパスが流入することになる。したがって枝に対する変数 y_4 、 y_5 のいずれかが 1 をとるとき、フレーズ対 4 に対する変数 x_4 の値は 1 となる。以上より $y_4 + y_5 = x_4$ という関係式が導かれる。その他のフレーズ対についても同様にして有向枝との関係式を求めると、制約式 $N\mathbf{y} = \mathbf{x}$ は次のようになる。ここで、左辺第一項の行列が N である。

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \quad (8)$$

3.2 フレーズアライナを用いたリランキングの手順

本節では、3.1 節で提案したフレーズアライナを翻訳システムに適用して翻訳候補のフレーズ対応を最適化しリランキングを行う流れを説明する。

まず、フレーズベースデコーダを用いて入力文に対する翻訳候補上位 N 個すなわち N best 翻訳を求める。次に、デコーダに入力した原言語文と各翻訳候補のペアを対訳文として、フレーズアライナを用いて最適フレーズ対応を求める。なお、フレーズ対応最適化時には翻訳時と異なり、目的言語文である翻訳候補文が固定されているため、言語モデル確率や目的言語文

表 1 コーパス詳細

データセット	文数	単語数	語彙サイズ
学習セット (英)	1,798,571	59,974,173	142,435
学習セット (日)	1,798,571	64,184,179	121,652
dev. セット (英)	915	30,028	3,986
dev. セット (日)	915	32,427	3,653
テストセット (英)	1,381	45,334	4,116
テストセット (日)	1,381	48,737	3,882

表 2 評価実験条件

項目	条件
デコーダ	Moses (08/02/20 release)
ビーム幅	10, 20, 50, 100, 200, 500, 1,000
beam-threshold	1e-5
ttable-limit	20
リオーダーリングモデル	msd-bidirectional-fe
歪み距離制限	なし
数値計画問題の Solver	CPLEX 11.0

長によるペナルティは考慮しなくてよいことに注意されたい。最後に、最適化されたフレーズ対応のスコアに言語モデルスコアおよび目的言語文長ペナルティを加味し、各翻訳候補の再順位付けを行い、最も確率の高い候補を翻訳システムの出力とする。ここでスコアとは対数確率である。すなわち提案手法は、従来の統計的機械翻訳システムとは異なり、式 (1) の \max_c の部分を近似することなく厳密に行う。

4 評価実験

4.1 実験条件

実験には、NTCIR-7 特許翻訳タスク [4] で配布された英日対訳コーパスを用いた。フレーズテーブルの学習には同学習セットを、デコーダの各素性に対する重み学習には development セットを使用した。またテストセットは NTCIR-7 フォーマルランで配布されたテストセットとした。コーパスの詳細を表 1 に示す。翻訳方向は日英とし、翻訳精度の評価には BLEU を、ベースラインには Moses デコーダ^{*1}を用いた。デコーダパラメータ等の実験条件を表 2 に示す。また原言語文一文あたりの N best 数はビーム幅と同値とした。

また、Moses は N best 翻訳を求める際、探索途中で recombine した翻訳仮説を再展開し N best 候補に含める。しかし、Moses の recombine 条件はリオーダーリングスコアの一致を保障しないため、Moses が付与した N best 翻訳のリオーダーリングスコアは近似されている。そのため、Moses をそのまま利用して N best 翻訳を得ると提案手法によるスコアの改善を正確に評価することができない。そこで、recombine 条件をデフォルトのままとした場合に加えて、リオーダーリングスコアが一致することを保障するための recombine 条件を追加した場合についても、スコア改善幅を評価するために実験を行った。なお、recombine 条件を追加した場合、同時に N best 翻訳に目的言語文が一致しないための Moses のオプション: distinct を適用した。

4.2 実験結果

Moses および提案手法による翻訳結果の BLEU を表 3 および図 3 に示す。表および図中、rerank は提案手法を、default は Moses 本来の recombine 条件であり、かつ distinct オプションなしの場合を、recombine+ は recombine 条件を Moses に新たに追加し distinct オプションを適用した場合をそれぞれ意味する。提案手法と Moses を比較すると提案手法は常に Moses より BLEU 値が高くなっている。特に recombine+ の場合、有意水準 5% として有意差検定を行ったところ、ビーム幅 200 以下の全測定点において提案手法が Moses を有意に上回っているという結果が得られた。また default と recombine+ とで大きな BLEU の差は見られなかった。

recombine+ におけるテストセットに対する処理時間と BLEU の関係を図 4 に、ビーム幅と平均スコア改善幅との関係を図 5 にそれぞれ示す。ここで処理時間は、Moses の場合にはテストセット全文の翻訳に要した時間を、rerank の場合には Moses の翻訳時間とフレーズ対応最適化に要した時間との和を表す。平均スコア改善幅は、 N best 各候補のフレーズ対応最適化前後のスコアの差分を平均したものである。なお、各候補のスコアの大きさはおよそ -10 から -100 程度のオーダーであった。図 4 において Moses と提案手法とで処理時間を比較すると、提案手法は Moses と同程度の翻訳精度のとき、およそ Moses の 2 倍程度の時間を要することがわかる。また探索時間が短いとき、すなわちビーム幅が小さいときには rerank による BLEU の改善が大きいが、ビーム幅を大きくしていくと次第に提案手法の効果が小さくなっていく。最終的に、ビーム幅 500 以上では Moses と提案手法とで翻訳精度はほぼ変わらなくなる。これはビーム幅を大きくとることで、Moses の探索の精度が向上していくためである。このことは図 5 からも確認することができ、ビーム幅を大きくしていくとスコア改善幅が小さくなっていく様子が見られる。

最後に、提案手法によりフレーズ対応および翻訳精度が改善された例を図 6 に示す。図中、Moses は Moses による翻訳結果を表し、rerank は提案手法によりスコアが 1 位となった翻訳候補を表す。また図 6 には翻訳結果の他に、ランキングによりスコア 1 位になった候補の最適化前後のフレーズ対応も示してあり、四角い枠がフレーズを、フレーズ同士を結ぶ線がフレーズ対応を表している。図 6 では、正解翻訳例中の “control electrode (gate)” というフレーズが、Moses の翻訳結果においては “control electrode” と “(gate)” とに分解され、“(gate)” の係り先が変わってしまっている。一方、rerank は正解例と同様に “control electrode” と “(gate)” とが連続しており、Moses に比べてよい翻訳となっていると言える。また図 6 において、ランキングの結果スコア 1 位になった翻訳候補の最適化前後のフレーズ対応を比較すると、最適化前は “ている。” と “.” とが対応していたのに対し、最適化後は “接続されている” と “is connected”、および “。” と “.” といったより自然なフレーズ対応に改善されていることがわかる。

5 おわりに

本稿では、リオーダーリングモデルを考慮したフレーズ対応問題の新しい定式化を提案した。また、それを用いてフレーズベース翻訳システムの翻訳候補に対しフレーズ対応最適化を行い、 $P(\mathbf{f}, \mathbf{c} | \mathbf{e})$ を厳密に最大化することによってランキングを行う手法について検討した。提案手法は NTCIR-7 特許翻訳

*1 <http://www.statmt.org/moses/>

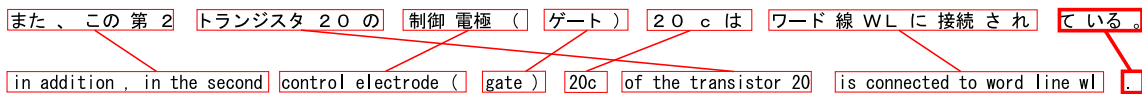
入力文： また、この第2トランジスタ20の制御電極（ゲート）20cはワード線WLに接続されている。

正解例： the control electrode (gate) 20c of the second transistor 20 is connected to the word line wl .

Moses： in addition , in the second control electrode of the transistor 20 is connected to the word line wl (gate) 20c .

rerank： in addition , in the second control electrode (gate) 20c of the transistor 20 is connected to word line wl .

最適化前のフレーズ対応



最適化後のフレーズ対応

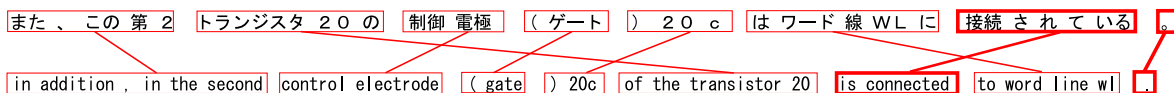


図6 提案手法によりフレーズ対応および翻訳精度が改善された例

表3 翻訳精度 (単位: %BLEU)

ビーム幅	default		recombine+	
	Moses	rerank	Moses	rerank
10	28.2	28.4	28.2	28.3
20	28.5	28.6	28.5	28.7
50	28.8	29.0	28.9	29.1
100	29.1	29.2	29.0	29.2
200	29.2	29.4	29.2	29.5
500	29.4	29.6	29.5	29.5
1,000	29.4	29.5	29.6	29.6

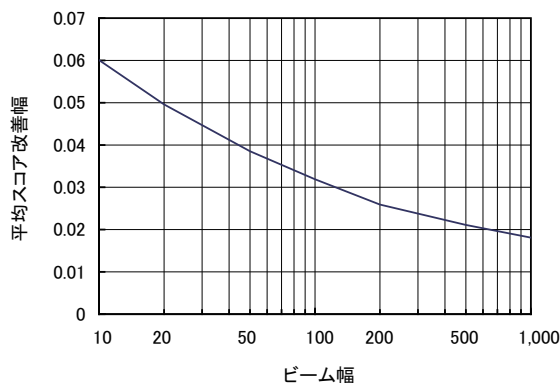


図5 ビーム幅に対する平均スコア改善幅

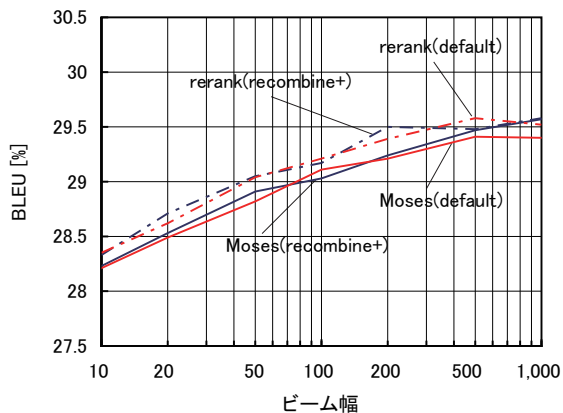


図3 ビーム幅に対する翻訳精度

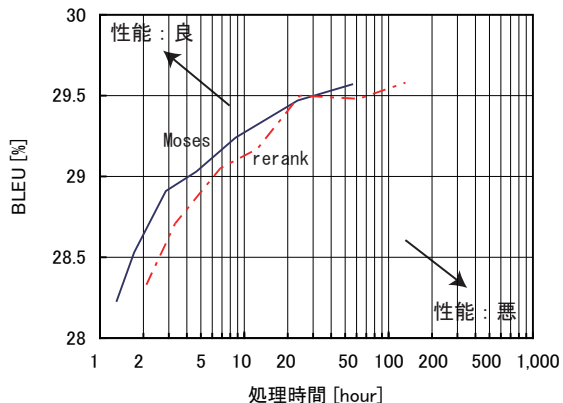


図4 翻訳時間とBLEUとの関係

タスクのデータセットを用いた評価実験において、常に Moses が翻訳候補に与えたスコアを改善し、BLEU を向上させることが確認できた。しかし、提案手法によるスコア改善はそれほど大きくなく、従来から用いられているヒューリスティック探索であっても、フレーズ対応についてはかなり高い精度の解が得られていると言える。今後は、翻訳時に最適化手法を取り入れる方法について検討していきたい。

参考文献

- [1] P.Koehn, F.J.Och and D.Marcu. 2003. "Statistical phrase-based translation." In *NAAACL '03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, pp.48-54.
- [2] S.Hasan, R.Zens and H.Ney. 2007. "Are Very Large N-Best Lists Useful for SMT?." In *Proceedings of Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers*, pp.57-60.
- [3] J.DeNero and D.Klein. 2008. "The Complexity of Phrase Alignment Problems." In *Proceedings of ACL-08: HLT, Short Papers*, pp.25-28.
- [4] A.Fujii, M.Utiyama, M.Yamamoto, and T.Utsuro. 2008. "Overview of the Patent Translation Task at the NTCIR-7 Workshop." In *Proceedings of the 7th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-lingual Information Access*, pp.389-400.