Factored Translation Models を用いた事後並べ替えによる日英翻訳

小林和也　Kevin Duh　松本裕治
奈良先端科学技術大学院大学 情報科学研究科
{kazuya-ko, kevinduh, matsu}@is.naist.jp

1　はじめに

統計的機械翻訳の翻訳精度は，翻訳を行う言語ペアによって大きく変化する．例えば，英語から日本語への翻訳精度は，フランス語への翻訳精度よりも低くなる．この翻訳精度の違いの主な要因として，言語間の文構造の違いが挙げられる．英語やフランス語は，”John hit a ball.”のように主語－動詞－目的語という語順のSVO 言語であるのに対し，日本語は”ジョンはボールを打った。”のように主語－目的語－動詞という語順の SOV 言語である．

英語と日本語のように文構造が異なる言語間の翻訳を行う場合，長距離の語順の並べ替えを考慮しなければならない．もし，出力する単語数が n 個で並べ替えの距離を制限しない場合，単語列の候補は $n!$ 個となり，探索空間が非常に大きくなるため，全ての単語列の候補を考慮することは計算量の問題から不可能である．また，現在の統計的機械翻訳システムの並べ替えモデルは長距離の語順の並べ替えを解決するには充分ではない．以上の 2 点より，現在の統計的機械翻訳システムは長距離の語順の並べ替えが必要となる言語間の翻訳を不得手としている．

文構造が異なる言語間での翻訳における長距離の語順の並べ替えの問題を解決するために，事前並び替えと事後並び替えと呼ばれる手法が提案されてきた．これらの手法は単語の翻訳と語順の並べ替えを別々に行うことで翻訳精度を向上させている．事前並べ替えでは，前処理として原言語を目的言語の語順に並べ替えたあとに翻訳を行う．翻訳を行う前に原言語の語順を目的言語に近づけることで，翻訳中の語順の並べ替えの距離を少なくしている．一方の事後並べ替えは，翻訳を行ったあとに語順の並べ替えを行う手法である．

本研究では日英翻訳における語順の並べ替えの問題を解決するために，事後並べ替えに着目し，単語の表層以外の情報を考慮する手法を提案する．提案手法では factored translation models を用い，単語の表層と品詞，大規模データに対してクラスタリングを行うことによって求めた単語のクラスタの情報を考慮した翻訳を行う．また，考慮する情報による翻訳への影響も実験によって調査する．

2　事後並べ替え

英語の日本語語順へと並べ替えはいくつかの単純な規則で高精度に実現できる．これは，日本語が典型的な主辞後続言語（head-final language）であるため，統語主辞を対応するフレーズや節の最後尾に移動させればよいからである．Isozaki らは日本語のこのような特徴に着目した並べ替え規則を提案して英日翻訳の事前並べ替えを行った．一方，日英翻訳では日本語を英語語順に並べ替える規則は簡単に書けないため，事前並べ替えでは翻訳精度はなかなか向上しない．そこで，Sudoh らは Isozaki らの提案した並べ替え規則を用いて，事後並べ替え [9] を提案した．Sudoh らの提案した事後並べ替えでは最初に，Isozaki らの提案した並べ替え規則 [3] を用いて日本語語順の英語 (Head-Final English:HFE) を獲得する．その後に，日本語と HFE のペアと HFE と英語のペアを使って 2 つの機械翻訳システムを学習する．そして，日本語から HFE への翻訳と HFE から英語への翻訳の 2 段階の翻訳を行うことで日英翻訳を行う．事後並べ替えにおける翻訳の流れを図 1 に示す．

2.1　Head Finalization

本節では，英語を HFE に並べ替える Head Finalization の規則 [3] について説明する．本研究では，英語における統語主辞を出力するための解析器として

Copyright(C) 2014 The Association for Natural Language Processing.
All Rights Reserved.

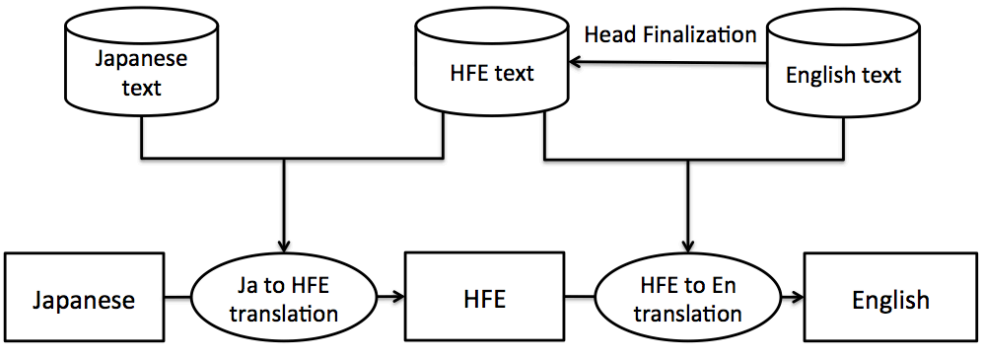

図 1: 事後並べ替えの流れ

Enju[1][5] を用いた．Enju はそれぞれのノードについて最大 2 つのノードを子として出力する．片方が統語主辞で，もう片方が従属部である．Head Finalization は英語に対して以下の規則を適用して並べ替えを行う．

1. 統語主辞はその従属部の後ろに置く．ただし，並列句については並べ替えを行わない．
2. 日本語の助詞の”は”や”を”に相当する擬似単語を挿入する．
   - va0：文章の主辞動詞の主語．
   - va1：その他の動詞の主語．
   - va2：動詞の目的語．

規則 2 は英語と日本語の単語のアラインメントを取りやすくするための規則である．

2.2 2 段階の機械翻訳

事後並べ替えにおける最初の翻訳である日本語から HFE への翻訳では，原言語に日本語のコーパス，目的言語に HFE のコーパスを用いて統計的機械翻訳システムを構築する．HFE のコーパスは英語のコーパスに対して Head Finalization を適用したものを使用する．日本語から HFE への翻訳の大きな目的はフレーズの翻訳であり，長距離の語順の並べ替えは行わない．よって，デコードの際は短い距離の単語の移動のみを許すか，単語の移動を全く許さないようにディストーションリミットを設定する．

HFE から英語への翻訳では，HFE と英語のコーパスを用いて統計的機械翻訳システムを構築する．HFE から英語への翻訳では単語の長距離の並べ替えが大きな目的である．よって，長距離の並べ替えが行えるようにディストーションリミットを設定しデコードを行う．

[1]http://www.nactem.ac.uk/tsujii/enju/index.html

3 Factored Translation Models を用いた事後並べ替え

本研究では，単語以外の情報を考慮するために factored translation models[4] を用いる．

f を出力文，e を入力文としたとき，本来の統計的機械翻訳は翻訳確率 $p(f|e)$ を翻訳モデル $p(f_{word}|e_{word})$ と言語モデル $p(f_{word})$ を使って最適な f を出力する．Factor translation models ではそれに加えて翻訳モデル $p(f_{word}, f_{factor}|e_{word})$ と言語モデル $p(f_{factor})$ を考慮する．Factor を考慮することで，翻訳における情報量や制限を増やしている．

本研究では，Factor は品詞とクラスタ，それらの組み合わせの 3 種類を考慮する．品詞は Enju の出力を用い，クラスタは訓練データに対し，Brown Clustering[1] を適用したものを用いる．Brown Clustering のクラスタ数は 50 と 1,000 の 2 種類で分類を行った．50 クラスタでの分類は Enju の出力する品詞が約 50 個であるため，品詞による分類と Brown Clustering による分類の違いの影響を評価するために用いる．1,000 クラスタでの分類は品詞と単語の間の粒度の分類による翻訳への影響を評価するために用いる．

3.1 日本語から HFE への翻訳

日本語から HFE への翻訳では，出力側である HFE の品詞とクラスタを考慮し，言語モデルと翻訳モデルを学習する．この際，HFE のクラスタは英文の訓練データではなく，HFE の訓練データに対して Brown Clustering を適用したものを用いる．また，実験におけるディストーションリミットは 6 とする．

Copyright(C) 2014 The Association for Natural Language Processing.
All Rights Reserved.

3.2 HFE から英語への翻訳

HFE から英語への翻訳では英語の factor を考慮して，言語モデルと翻訳モデルを学習する．また，ディストーションリミットは 12 として実験を行う．

4 事後並べ替えにおける Factor の影響の評価実験

4.1 実験設定

実験には Wikipedia 日英関連文書対訳コーパスを対訳コーパスとして用いる．Wikipedia 日英関連文書コーパスは京都に関連する Wikipedia の記事を使用した対訳コーパスである．今回の実験では，訓練データに 318,443 文，パラメータのチューニングデータに 1,166 文，テストデータに 1,160 文をそれぞれ用いた．単語アラインメントの獲得には GIZA++[2][7] を用い，言語モデルの学習には SRILM[3] を用いた．言語モデルは単語の表層は 5-gram，factor は 7-gram までを学習した．パラメータのチューニングには MERT[6]，デコーダには Moses[4] を用いた．翻訳結果の評価には BLEU[8] と RIBES[2] を用いた．BLEU は 1-gram から 4-gram までの適合率と短い文へのペナルティから計算される．RIBES は語順の評価を行うために用いられ，ケンドールの順位相関係数と 1-gram の適合率によってスコアが計算される．いずれの評価指標もスコアが高いほど良い翻訳が得られていると言える．

4.2 実験結果・解析

実験結果を表 1 に示す．事後並べ替えの有無にかかわらず factor を考慮することで BLEU スコアが上昇している．一方で，RIBES は事後並べ替えを用いない場合は低下しているが，事後並べ替えを用いた場合は上昇している．このことから factored translation models は事後並べ替えに有効であるといえる．

事後並べ替えにおけるそれぞれの factor の影響については，品詞と 1000 クラスタを考慮した場合に BLEU と RIBES のスコアが最も高くなっている．factor 間の違いを見ると，品詞よりも 50 クラスタを考慮することで翻訳精度が向上していることが分かる．これは Brown Clustering がデータに沿った分類を行なっているため，よりデータの傾向に沿ったモデルが学習され，翻訳精度の向上につながったと考えられる．

システム	BLEU	RIBES
PBMT	15.65	68.35
品詞	16.06	**68.62**
50 クラスタ	**16.32**	68.36
1000 クラスタ	15.61	68.39
品詞 + 50 クラスタ	16.17	68.06
品詞 + 1000 クラスタ	16.09	68.44

表 2: 日本語から HFE への翻訳における評価値

システム	BLEU	RIBES
PBMT	59.69	82.31
品詞	58.85	81.85
50 クラスタ	60.09	82.27
1000 クラスタ	60.74	**83.36**
品詞 + 50 クラスタ	60.08	82.58
品詞 + 1000 クラスタ	**60.99**	83.16

表 3: HFE から英語の翻訳における評価値

4.3 日本語から HFE への翻訳における factor の影響

単語の翻訳に有効な factor を調査するために日本語から HFE への翻訳を行い，翻訳精度を比較した．正解データには英語のテストデータに対して Head Finalization を適用したものを用いた．実験結果を表 2 に示す．これを見ると，50 クラスタを考慮した際に，BLEU が最も高くなっている．このことから，単語の翻訳は 50 クラスタを考慮することが有効であると言える．

4.4 HFE から英語への翻訳における factor の影響

単語の並べ替えに有効な factor を調査するために，factor 毎の HFE から英語への翻訳の精度を比較した．本実験では英語のテストデータに対して Head Finalization を適用したデータを入力文として使用し，HFE から英語への翻訳を行った．この結果を表 3 に示す．表 2 と異なり，クラスタを考慮すると BLEU と RIBES

[2]https://code.google.com/p/giza-pp/
[3]http://www.speech.sri.com/projects/srilm/download.html
[4]http://www.statmt.org/moses/

Copyright(C) 2014 The Association for Natural Language Processing.
All Rights Reserved.

	ディストーションリミット		BLEU	RIBES
	日本語 → HFE	HFE → 英語		
PBMT	12		16.25	**70.13**
PBMT + 品詞			16.68	64.54
PBMT + 50 クラスタ			17.36	65.25
PBMT + 1000 クラスタ			17.56	65.88
PBMT + 品詞 + 50 クラスタ			17.41	65.23
PBMT + 品詞 + 1000 クラスタ			**17.47**	65.50
[*Sudoh et al.*, 2012]	6	12	16.22	65.73
事後並べ替え + 品詞			16.22	65.77
事後並べ替え + 50 クラスタ			16.69	65.39
事後並べ替え + 1000 クラスタ			16.16	65.89
事後並べ替え + 品詞 + 50 クラスタ			16.55	65.45
事後並べ替え + 品詞 + 1000 クラスタ			**16.79**	**65.99**

表 1: 各手法に対する評価指標の数値

の両方のスコアが上昇し，品詞と 1,000 クラスタの両方を考慮したときに BLEU が最も高くなっている．

5 おわりに

本論文では，日英翻訳に対する事後並べ替えに factored translation models を考慮した手法を提案した．実験から，品詞やクラスタの情報を考慮することによって並べ替えの精度が向上することが確認できた．また，factored translation models は BLEU による n-garm の適合率と RIBES による語順の適合率に対して異なる影響を持つことがわかった．今後の課題として，単語の翻訳に有効な factor の調査や，原言語側の factor を考慮したモデルの提案が考えられる．

参考文献

[1] Peter F Brown, Peter V Desouza, Robert L Mercer, Vincent J Della Pietra, and Jenifer C Lai. Class-based n-gram models of natural language. *Computational linguistics*, Vol. 18, No. 4, pp. 467–479, 1992.

[2] Hideki Isozaki, Tsutomu Hirao, Kevin Duh, Katsuhito Sudoh, and Hajime Tsukada. Automatic evaluation of translation quality for distant language pairs. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp. 944–952, 2010.

[3] Hideki Isozaki, Katsuhito Sudoh, Hajime Tsukada, and Kevin Duh. Head finalization: A simple reordering rule for sov languages. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, pp. 244–251, 2010.

[4] Philipp Koehn and Hieu Hoang. Factored translation models. In *EMNLP-CoNLL*, pp. 868–876, 2007.

[5] Yusuke Miyao and Jun'ichi Tsujii. Feature forest models for probabilistic hpsg parsing. *Computational Linguistics*, Vol. 34, No. 1, pp. 35–80, 2008.

[6] Franz Josef Och. Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, pp. 160–167, 2003.

[7] Franz Josef Och and Hermann Ney. A systematic comparison of various statistical alignment models. *Computational linguistics*, Vol. 29, No. 1, pp. 19–51, 2003.

[8] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pp. 311–318, 2002.

[9] Katsuhito Sudoh, Xianchao Wu, Kevin Duh, Hajime Tsukada, and Masaaki Nagata. Post-ordering in statistical machine translation. In *Proc. MT Summit*, 2011.

Copyright(C) 2014 The Association for Natural Language Processing.
All Rights Reserved.