

統語ベース翻訳に対する統語的前処理の適用

波多腰 優斗^{†1,a)} Graham Neubig^{†1,b)} Sakriani Sakti^{†1} 戸田 智基^{†1} 中村 哲^{†1}

概要: フレーズベース翻訳において、翻訳精度を向上させるために統語情報を用いた様々な前処理の手法が提案されている。例えば英日翻訳において、英語を日本語の語順に近づける並べ替え処理や、日本語の助詞に相当する擬似単語を挿入する処理を行う Head Finalization[1] は、フレーズベース翻訳の精度を大幅に改善することが知られている。このような統語的前処理が精度の向上に貢献する理由として、フレーズベース翻訳の並べ替えモデルの貧弱さと事前並べ替えによる改善が挙げられてきた。しかし、アライメント精度の改善や、擬似助詞の挿入による翻訳性能改善も Head Finalization の重要な要素であり、これらの効果はフレーズベース翻訳に限ったものではないと考えられる。そこで、本稿では Head Finalization の各要素がどの程度翻訳精度に貢献しているかを検証し、並べ替えの問題が比較的少ない統語ベース翻訳に対しても有効であるか調査を行う。実験の結果、統語ベース翻訳に対して統語的前処理を適用することで翻訳精度が向上することを確認した。

1. はじめに

統計的機械翻訳では、複数の単語から成るフレーズ間の翻訳確率を計算し、目的言語として適切な語順となるように並べ替えモデルによる局所的なフレーズの移動を行う、フレーズベース機械翻訳 (phrase-based machine translation: PBMT)[2] が広く用いられている。PBMT は翻訳モデルの学習が容易であり、多くの言語対で高い精度を実現する一方、形態的・統語的な情報の扱いに乏しいという問題点もある。この問題を解決し、翻訳精度を向上させるために、これらの情報を用いた前処理を適用する手法が提案されている。例えば、動詞の接頭辞を分割して目的言語との単語対応を取りやすくする前処理 [3] や、名詞の格の不一致・動詞の活用誤り減らすための前処理 [4] などにより翻訳精度が向上しており、統語的前処理の有効性が示されている。

特に注目されている前処理の一つとして事前並べ替え [5] が挙げられる。PBMT の並べ替えモデルは、長距離の並べ替え確率を正確に推定することが困難であることから、英語と日本語のように語順が大きく異なる言語対では翻訳精度が低下する問題もある。そのため、原言語文を目的言語に近い語順に並び替える事前並べ替えを適用してから PBMT による翻訳を行うことで、翻訳精度が大幅に改善されることが知られている。Head Finalization[1] という手法は、英日翻訳において有効な統語的前処理手法であり、

二言語間の統語的な構造に基づくシンプルなルールによって、PBMT の翻訳精度が飛躍的に向上することが知られている。Head Finalization の処理の中でも、英語の主辞を句の末尾に移動させる事前並べ替え処理が特に注目されている。しかし、英語文を日本語に近い語順に並べ替える処理以外にも、より日本語に近い文とするために冠詞の削除、助詞の挿入といった語順の操作とは関係のない、単語に関する前処理も行っている。このような並べ替え以外の処理も重要な要素となっている場合があり、これらの統語的前処理による効果は PBMT に限ったものではないと考えられる。

PBMT 以外の翻訳手法としては、文の構文解析結果に基づいて翻訳を行う統語ベース翻訳 [6] がある。統語ベース翻訳に対する統語的前処理としては、構文木の変換によって文の構成要素と単語アライメントの対応を改善するための前処理 [7] が提案され、翻訳精度を向上させている。統語ベース翻訳に対しても統語的前処理の有効性は示唆されており、前述したルールに基づく並べ替え処理や、単語に関する前処理の適用についても検証する余地がある。しかし、統語ベース翻訳では並べ替え問題の影響も少なく、構文解析器が付与する豊富な情報を用いた翻訳が可能であることから、統語的前処理を適用した事例が殆どない。

そこで本稿では、ルールに基づく統語的前処理を統語ベース翻訳に適用することを検討する。具体的には、Head Finalization の各処理がどの程度 PBMT の翻訳精度の向上に貢献しているかを検証するとともに、統語ベース機械翻訳に対しても同様の処理が有効であるかを確かめる。さ

^{†1} 現在、奈良先端科学技術大学院大学
Presently with Nara Institute of Science and Technology
a) hatakoshi.yuto.hq8@is.naist.jp
b) neubig@is.naist.jp

らに、デコード時に用いる素性として追加するなど、ソフトな制約としてルールを取り入れた場合の翻訳精度についても調査する。実験で用いたタスクにおける英日翻訳で、Head Finalization における並び替え以外の処理も PBMT の翻訳精度の向上に寄与していることを確認し、統語ベース翻訳に対しても Head Finalization の統語的前処理が有効であることが分かった。

2. Head Finalization

Head Finalization は、英日翻訳における文法の不一致を緩和した手法であり、後述する日本語の主辞後置性に基づくシンプルなルールによる並び替えと、機能語に関する処理を行う。Head Finalization によって前処理された英語と日本語の対訳データを用いて PBMT の翻訳モデルを学習することで、翻訳精度が飛躍的に向上する。

2.1 日本語の主辞後置性に基づく並び替え

英語文に対する Head Finalization の適用例を図 1 に示す。Head Finalization の並び替え処理は、日本語の主辞が、主辞を修飾する語の後方に置かれるという統語的な特徴を利用したものであり、英語の構文木の各非終端ノードにおいて、その子ノードの中の主辞を末尾に移動させるというものである。図 1 の構文木では、主辞となっている要素への枝を太線で示しており、この要素が子ノードの先頭であるときに入れ替えを行う。例えば、動詞句 VP(黒色のノード)の主辞である VBD を末尾に移動させることで、“John hit a ball”(Original English) が “John a ball hit” のように日本語語順の英語 (Head Final English) に変換される。

2.2 単語に関する処理

Head Finalization では、語の並び替えと関係のない以下の 3 つの処理を行う。

- (1) 擬似助詞の挿入
- (2) 冠詞 “a”, “an”, “the” の削除
- (3) 単数化

(1) 擬似助詞の挿入は、英語文にない助詞を補うことで、翻訳時の助詞の欠如や誤訳を防ぐことを目的としている。具体的には、構文情報を利用して、日本語の格助詞「が」「は」(主格)や「を」(目的格)に相当する以下の 3 つの擬似助詞を挿入する。

- va0: 文の主辞動詞の主格助詞
- va1: その他の動詞の主格助詞
- va2: 動詞の目的格助詞

図 1 の例では、動詞 “hit” の主語となる “John” の後ろに主格を付与する va0, 目的語となる “ball” の後ろに目的格を付与する va2 を挿入する。

また、日本語には冠詞が存在せず、名詞の語尾変化による

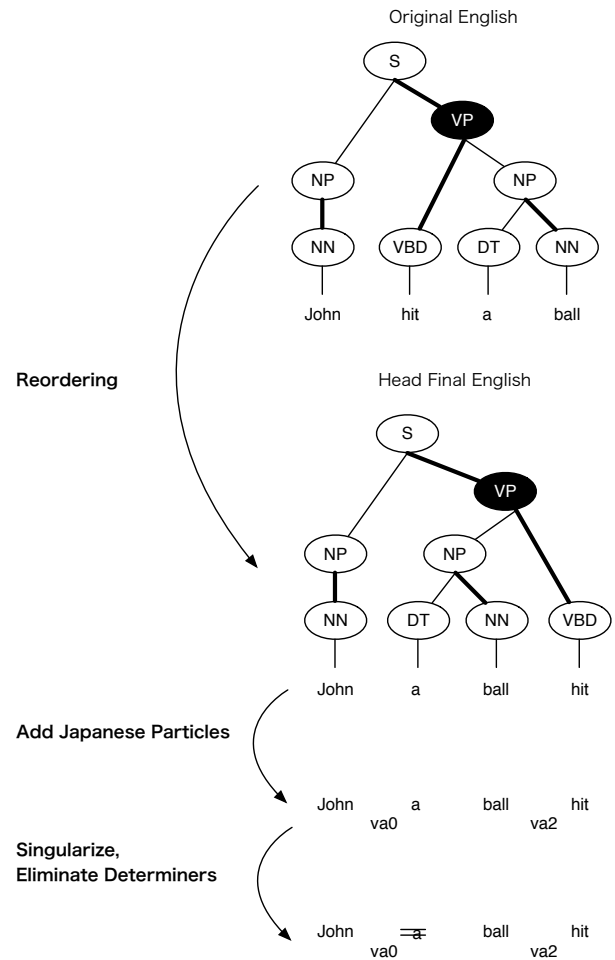


図 1 Head Finalization の適用例

単数と複数の区別がないため、(2)(3)の処理を行い、より日本語らしい文を生成することで、単語対応を取りやすくする。これらの処理をすべて適用することで、“John va0(wa) ball va2(wo) hit” という日本語に近い文が得られる。

3. 統語ベース翻訳

統語ベース翻訳は、文の構文情報を用いて翻訳を行う手法である [6]。統語ベース翻訳では、翻訳パターンを構文木の部分木の構造を用いて与えるため、フレーズベース翻訳よりも並び替えを正確に行うことができる。以下、統語ベース翻訳の中でも原言語側のみの構文情報を用いた、同期木置換文法 (STSG) [8] に基づく T2S(Tree-to-String) 翻訳について述べる。

T2S 翻訳は原言語文の構文解析結果を利用することで、二言語間の関係を統語的な構造により捉えることができ、正確な翻訳が可能となる。翻訳パターンはフレーズベース翻訳のように単語列ではなく、変数 x を含むルールも利用して原言語文の部分木の構造として表現される。以下の例は、置き換え可能な 2 つの NP を翻訳パターンに直接含んでおり、 x_0, x_1 に当てはまる候補の確率と翻訳パターン自体の確率を考慮して訳文を生成する。

S (x0:NP) (VP (VBD hit) (x1:NP)))

→ x0 は x1 を打つ

原言語文の部分木を用いることで訳出候補が少なくなるメリットもあり、探索空間が小さくなることで翻訳に必要な時間も短縮される。一方で、構文解析の結果を基に翻訳パターンを生成するため、翻訳精度が構文解析器の精度に依存してしまうというデメリットもある。

4. 統語ベース翻訳に対する Head Finalization の適用

2節で述べたように、Head Finalization では英語文の並べ替え以外にも、擬似助詞の挿入、冠詞の削除、単数化といった翻訳精度向上のための処理も行っており、フレーズベース翻訳以外の翻訳手法においてもその効果が期待できる。本研究では統語ベース翻訳の中でも、原言語側の構文情報を T2S 翻訳に対して Head Finalization を適用する。

実験では、T2S の前処理として適用した場合、翻訳パターンの素性として追加した場合の2つの手法の効果を確かめた。

4.1 T2S に対する統語的前処理

PBMT に対して適用する場合と同様に、表1に示す2つの処理を T2S の前処理として適用する。以下、これらの処理の適用方法と期待される効果について詳細に述べる。

表 1 T2S に適用する統語的前処理

処理名	処理内容
Reordering	日本語の主辞後置性に基づく並べ替え
Word Processing	擬似助詞の挿入、冠詞の削除、単数化

4.1.1 T2S のための事前並べ替え

Reordering は図1に示したように、日本語の主辞後置性に基づいて、英語文を日本語の語順に変換する処理である。PBMT における前処理は、原言語文を目的言語の語順に近づけることで、並べ替えモデルにおける長距離の語順移動問題を緩和する目的で用いていた。一方で、T2S における翻訳パターンは原言語文の部分木の構造を用いて表現されるため、両言語の語順が大きく異なる場合でも並べ替えの問題は比較的少なく、PBMT の場合ほど事前並べ替えの効果は期待できない。

しかし、事前並べ替えは翻訳時の並べ替え問題を緩和する以外にも、翻訳モデル学習時の精度向上に貢献することが考えられる。例えば、IBM Model や HMM などの単語対応を獲得する手法 [9] は語順に強く影響されているため、二言語間の語順が近づくことによるアライメント精度が改善が期待できる。この要素も翻訳精度の向上に寄与するため、前処理として適用する。

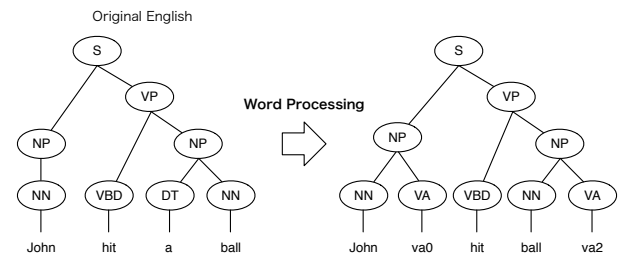


図 2 Word Processing の適用例

4.1.2 T2S のための単語に関する処理

Word Processing は、Head Finalization における擬似助詞の挿入、冠詞の削除、単数化の処理を行う。Word Processing の適用例を図2に示す。

擬似助詞の挿入、冠詞の削除では、簡単な構文木の変換処理を行っている。擬似助詞の挿入では、2.2節に示した3種類の擬似助詞“va0”、“va1”、“va2”を終端ノードとして、非終端ノードVAとともに追加する。図2の例では、動詞“hit”の主語となるNPノードの末尾に主格を付与する“va0”、目的語となるNPノードの末尾に目的格を付与する“va2”を挿入する。冠詞の削除では、終端ノード“a”、“an”、“the”とともにDTノードを削除する。例の場合、“a”をDTノードとともに削除する。

Word Processing は PBMT に適用する場合と同様に、英語文に無い助詞を補う処理や日本語に無い冠詞を削除する処理を行うことで、これらの単語に関する翻訳性能を上げることが目的としている。

4.2 並べ替えルールに従う翻訳パターンに対する素性の追加

4.1.1節で述べたように、T2S では語順が大きく異なる言語対を用いた場合でも、並べ替えにおける問題の影響が少ない。そのため、二言語間の語順を近づけるための前処理として Reordering を適用しても、並べ替えの精度が大幅に向上することは期待できない。そこで本研究では、並べ替えルールに従う翻訳パターンに対して素性を追加する。並べ替えに用いた統語的な特徴は、T2S の翻訳モデルにおいて、精度の向上に役立つ翻訳パターンを判別する指標として利用できる可能性がある。

PBMT や T2S 翻訳では、原言語文 f が与えられた時、全ての翻訳可能性 e' の中から最良と思われる訳出 e を求めるために、翻訳精度の向上に役立つ様々な素性を用いて、対数線形モデルによって定式化する。(1)式において、 $h(\cdot)$ は w により重み付けされる M 次元の素性関数であり、言語モデル確率、フレーズ翻訳確率、語彙化翻訳確率といった翻訳パターンの良し悪しを判別するのに役立つ素性が用いられる。

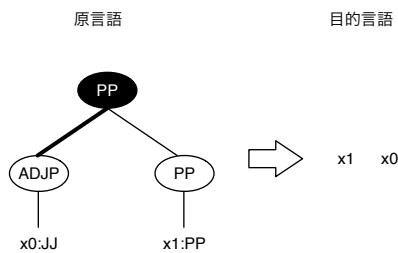


図 3 素性を追加する翻訳パターン

表 2 NTCIR7 のデータ内訳

Dataset	Lang	Words	Sentences	Average length
train	En	99.0M	3.08M	32.13
	Ja	117M	3.08M	37.99
dev	En	28.6k	0.82k	34.83
	Ja	33.5k	0.82k	40.77
test	En	44.3k	1.38k	32.11
	Ja	52.4k	1.38k	37.99

$$\hat{e} = \arg \max_e \frac{\exp(\mathbf{w}^T \mathbf{h}(\mathbf{f}, \mathbf{e}))}{\sum_{e'} \exp(\mathbf{w}^T \mathbf{h}(\mathbf{f}, \mathbf{e}'))} \quad (1)$$

本研究では Head Finalization の並べ替えルールに従う翻訳パターンに対して新たなバイナリ素性 HF-feature を追加する。Head Finalization の並べ替えルールは、日本語の主辞後置性に基づいて英語文を日本語の語順に近づけるため、翻訳精度向上のための指標になり得る。この並べ替えルールに従う翻訳パターンに対して素性を与える事により、学習された T2S の翻訳パターンの中で、二言語の統語的な特徴を考慮したものとそうでないものを区別することが可能になる。

具体的には、図 3 のように、原言語側の部分木に Head Finalization を適用した時のルール変数 x の並びと、目的言語側のルール変数 x の並びが一致する翻訳パターンに対して HF-feature を追加する。原言語-目的言語間でルール変数 x 以外の非終端ノードの単語対応を取ることができないため、ルール変数 x の数が 1 以下の翻訳パターンについては素性の追加を行わない。

5. 実験

実験では、Reordering と Word Processing の前処理の組み合わせにより、PBMT の翻訳精度がどの程度向上するかを検証する。T2S に対しては 2 つの前処理と HF-feature の追加の組み合わせによる翻訳精度を調べる。

5.1 実験条件

実験データには NTCIR-7 特許機械翻訳テストコレクションの英日翻訳データを用いた。実験データに関して、学習データ (train)、開発データ (dev)、テストデータ (test) の詳細を表 2 に示す。

英語側の文に対する構文解析は Enju[10]^{*1} を、日本語側の単語分割には KyTea[11]^{*2} を用いた。目的言語である日本語の言語モデルは SRILM[12] を用いて 5-gram で学習し、各素性の重みは BLEU が最大となるように MERT[13] を用いて最適化した。単語アライメントを取るツールとして GIZA++[9]^{*3} を用いた。実験における翻訳精度は、BLEU[14] と RIBES[15] の 2 つの自動評価尺度を用いて測った。PBMT は Moses[16]、T2S は Travatar[17]^{*4} に実装されているものをデフォルトの設定で用いた。

5.2 実験結果

実験結果を表 3 に示す。表中の太字は、危険率 5% の下でブートストラップ・リサンプリング法 [18] を用いて、最も精度の高い条件と比較した結果、統計的有意性がない数値を示している。また、“+” は各処理を行った場合、“-” は行っていない場合を表している。

PBMT の翻訳精度は、Reordering を適用した 2 つの条件下で BLEU と RIBES が最も高くなった。Reordering 適用下では、Word Processing の効果が確認できなかったが、Word Processing のみを適用した場合に BLEU と RIBES が高くなった。

T2S の翻訳精度は前処理を適用しない状態で、PBMT の最も良い条件下よりも高い精度となった。BLEU はスコアがほぼ横並びであり、5 つの条件下で最も高い値となった。RIBES は Word Processing を適用した 2 つの条件下で最も高いスコアとなった。

5.2.1 前処理の影響

事前並べ替えの先行研究で述べられているように、Reordering の処理により原言語文を目的言語の語順に近づけた場合に、BLEU と RIBES が大きく上昇することが確認された。また、Word Processing のみを追加した場合、追加していない場合と比較して BLEU スコアが向上した。Word Processing における 3 つの処理 (助詞の挿入、冠詞の削除、単数化) について、個別に翻訳精度への影響を調べたが、冠詞の削除、単数化の処理の有無による翻訳精度への影響は確認できなかった。

Reordering の適用によって PBMT の翻訳精度は大きく改善されたが、それとは対照的に、T2S に対しては効果が見られなかった。この理由としては、T2S が並べ替えに関して比較的頑健な手法であることと、二言語間の語順が近づくことによるアライメント精度の改善に関して、期待するほどの効果が得られなかったことが挙げられる。

一方、Word Processing は T2S に対しても有効に機能していることが分かった。Word Processing を適用した

*1 <http://www.nactem.ac.uk/enju/index.ja.html>

*2 <http://www.phontron.com/kytea/index-ja.html>

*3 <https://code.google.com/p/giza-pp/>

*4 <http://www.phontron.com/travatar/>

表 3 3つの処理の組み合わせによる各翻訳手法の精度

HF-feature	Reordering	Word Processing	PBMT		T2S	
			BLEU	RIBES	BLEU	RIBES
-	-	-	32.04	68.95	38.67	78.51
-	-	+	33.08	70.07	38.97	79.31
-	+	-	37.53	77.54	38.30	78.35
-	+	+	37.60	77.67	39.19	78.90
+	-	-	—	—	38.83	78.23
+	-	+	—	—	39.18	79.38
+	+	-	—	—	38.43	78.54
+	+	+	—	—	37.83	78.37

表 5 各条件における最適化された HF-feature の重み

HF-feature	Reordering	Word Processing	HF-feature の重み
+	-	-	-0.032579
+	-	+	0.0385629
+	+	-	0.117788
+	+	+	0.0902383

2つの条件下で最も RIBES が高くなっており、(+HF-feature, +Reordering, +Word Processing) を除く全ての条件で Word Processing 適用後に RIBES が高くなった。BLEU スコアでは Word Processing の適用による有意な翻訳精度の改善が見られなかったが、表 4 の翻訳結果の例に示すように助詞の翻訳性能の向上が確認できた。

5.2.2 T2S に対する HF-feature の効果

HF-feature の追加による翻訳精度の向上は確認できず、HF-feature を追加したことによって翻訳精度が低下してしまう例もあった。HF-feature が機能しなかった理由としては、MERT による最適化が上手く行えなかった点が挙げられる。表 5 に各条件における最適化された HF-feature の重みを示す。

(+HF-feature, -Reordering, -Word Processing) の条件では、負の重みが推定されており、素性を追加した翻訳パターンが重要とみなされなかった。一方、(+HF-feature, +Reordering, -Word Processing) の条件では、比較的高い重みが推定されており、HF-feature を追加した翻訳パターンがデコード時に選択されすぎることにより、局所解に陥っていることも考えられる。統語的な素性を用いて MIRA による最適化 [19] を行うことで翻訳精度が向上するなど、MERT 以外の最適化手法の効果も報告されており、これらについても検討していく必要がある。

最適化の問題に加えて、素性の追加方法の問題も原因の一つとして考えられる。今回の実験では、翻訳パターンの目的言語側のルール変数 x の並びのみを考慮して素性の追加を行った。この適用方法では、ルール変数 x 以外の表層の単語の並びの正確さを判断できないため、今後はこの点を改良していく必要がある。

6. おわりに

本稿では、統語的前処理の手法である Head Finalization の PBMT に対する効果をより詳細に分析するとともに、統語ベース翻訳の T2S に適用した。並べ替えを行う処理に加えて、擬似助詞の挿入などの単語に関する前処理を行う Word Processing も Head Finalization の処理の中で重要な要素となっており、Word Processing は T2S の前処理としても有効であることを確認した。

Head Finalization の並べ替えルールに従う翻訳パターンに対して素性を追加する手法については、最適化の方法や素性の追加方法など問題点が多く、今後改良すべき課題である。

参考文献

- [1] Isozaki, H., Sudoh, K., Tsukada, H. and Duh, K.: Head Finalization: A Simple Reordering Rule for SOV Languages, *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, Uppsala, Sweden, Association for Computational Linguistics, pp. 244–251 (2010).
- [2] Koehn, P., Och, F. J. and Marcu, D.: Statistical phrase-based translation, *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, Association for Computational Linguistics, pp. 48–54 (2003).
- [3] Nießen, S. and Ney, H.: Improving SMT quality with morpho-syntactic analysis, *Proceedings of the 18th conference on Computational linguistics-Volume 2*, Association for Computational Linguistics, pp. 1081–1085 (2000).
- [4] Avramidis, E. and Koehn, P.: Enriching Morphologically Poor Languages for Statistical Machine Translation., *Association for Computational Linguistics*, pp. 763–770 (2008).
- [5] Xia, F. and McCord, M.: Improving a statistical MT system with automatically learned rewrite patterns, *Proceedings of the 20th international conference on Computational Linguistics*, Association for Computational Linguistics, p. 508 (2004).
- [6] Yamada, K. and Knight, K.: A syntax-based statistical translation model, *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*,

表 4 Word Processing の適用による T2S の翻訳結果の改善例

原文	another connector 96 , which is matable with this cable connector 90 , comprises a plurality of male contacts 98 aligned in a row in an electrically insulative housing 97 as shown in the figure .
参照訳	このケーブル コネクタ 90 と 嵌合 接続される 相手 コネクタ 96 は、図示のように、絶縁ハウジング 97 内に雄 コンタクト 98 を 整列 保持して 構成される。
- Word Processing	このケーブル コネクタ 90 は 相手 コネクタ 96 は、図に示すように、電気絶縁性のハウジング 97 に一列に並ぶ複数の雄型コンタクト 98 とから構成されている。
+ Word Processing	このケーブル コネクタ 90 と 相手 コネクタ 96 は、図に示すように、電気絶縁性のハウジング 97 に一列に並ぶ複数の雄型コンタクト 98 を有して構成される。

- Association for Computational Linguistics, pp. 523–530 (2001).
- [7] Burkett, D. and Klein, D.: Transforming trees to improve syntactic convergence, *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Association for Computational Linguistics, pp. 863–872 (2012).
- [8] Graehl, J., Knight, K. and May, J.: Training tree transducers, *Computational Linguistics*, Vol. 34, No. 3, pp. 391–427 (2008).
- [9] Och, F. J. and Ney, H.: A systematic comparison of various statistical alignment models, *Computational linguistics*, Vol. 29, No. 1, pp. 19–51 (2003).
- [10] Yusuke, M. and Jun'ichi, T.: Maximum entropy estimation for feature forests, *Proceedings of the second international conference on Human Language Technology Research*, Morgan Kaufmann Publishers Inc., pp. 292–297 (2002).
- [11] Neubig, G., Nakata, Y. and Mori, S.: Pointwise prediction for robust, adaptable Japanese morphological analysis, *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, Association for Computational Linguistics, pp. 529–533 (2011).
- [12] Stolcke, A., Zheng, J., Wang, W. and Abrash, V.: SRILM at sixteen: Update and outlook, *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop*, p. 5 (2011).
- [13] Och, F. J.: Minimum error rate training in statistical machine translation, *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, Association for Computational Linguistics, pp. 160–167 (2003).
- [14] Papineni, K., Roukos, S., Ward, T. and Zhu, W.-J.: BLEU: a method for automatic evaluation of machine translation, *Proceedings of the 40th annual meeting on association for computational linguistics*, Association for Computational Linguistics, pp. 311–318 (2002).
- [15] Isozaki, H., Hirao, T., Duh, K., Sudoh, K. and Tsukada, H.: Automatic evaluation of translation quality for distant language pairs, *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, pp. 944–952 (2010).
- [16] Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R. et al.: Moses: Open source toolkit for statistical machine translation, *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, Association for Computational Linguistics, pp. 177–180 (2007).
- [17] Neubig, G.: Travatar: A forest-to-string machine translation engine based on tree transducers, *ACL 2013*, p. 91 (2013).
- [18] Koehn, P.: Statistical Significance Tests for Machine Translation Evaluation, *Proceedings of EMNLP 2004* (Lin, D. and Wu, D., eds.), Barcelona, Spain, Association for Computational Linguistics, pp. 388–395 (2004).
- [19] Chiang, D., Marton, Y. and Resnik, P.: Online large-margin training of syntactic and structural translation features, *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, pp. 224–233 (2008).