高度な英文訂正への統計的機械翻訳技術の適用

水嶋海都, 荒瀬由紀 †
大阪大学工学部電子情報工学科, † 大阪大学大学院情報科学研究科マルチメディア工学専攻
{mizushima.kaito, arase}@ist.osaka-u.ac.jp

1 はじめに

国際化やインターネットの普及に伴って世界共通言語のひとつである英語を使用する機会はますます増えており、ノンネイティブ話者による英語の使用を補助する技術が重要となっている。言語を習得すには「聞く・話す・書く」技術を身に着ける必要があるが、本稿では特に、英語を書く際の補助技術に取り組む。

ノンネイティブが英語を書く際の間違いをコンピュータによって自動訂正する技術は多数提案されている。英文の訂正においてその複雑さのレベルを 3 段階に分けることができる。

1. 単語の表記を訂正するスペルチェック
2. 文法的な間違いの訂正
3. ネイティブ話者から見てより自然な表現への訂正

1 のスペルチェックはワードプロセッサーなど多くのアプリケーションで実装されており、広く使用されている。また 2 の文法的な誤りの訂正もさかんに研究されている。特に大規模な学習用コーパスの整備と、近年大きく発展を遂げた機械学習、および統計的機械翻訳技術を用いることで、データから誤り訂正モデルを学習することが可能となり、訂正性能の向上が実現されている。

そこで本稿では、より高度な英文訂正技術である 3 の実現を目指し、文法的な誤り訂正で頻繁に用いられる統計的機械翻訳技術の適用と、代表的な翻訳モデルによる訂正性能の比較を行う。本稿では、訂正前の文を元言語、訂正後の文を目的言語とした対訳コーパスとみなす。このコーパスを用いて統計的機械手法であるフレーズベース [5]、階層的フレーズベース [1]、Tree-to-String[6] モデルを学習する。それぞれのモデルに対し、ノンネイティブ話者に書かれた訂正前の文を入力し、得られた翻訳文が正解データである訂正後の文にどれ程近いか比較評価する。これにより、英語学習者が書いた不自然なフレーズの訂正に対する統計的機械翻訳手法の有効性、および代表的な翻訳モデルの性能を比較・検証する。

2 関連研究

統計的機械翻訳を用いた文法的誤り訂正においては、例えば前置詞の誤りといった誤りのタイプに限定することが一般的である。Tajiri ら [11] は動詞の時制の誤りに注目し訂正を行い、Rozovskaya and Roth[12] は冠詞、名詞の単数形複数形に関する誤り、および動詞の誤りを訂正した。また文法的誤り訂正における統計的機械翻訳モデルの有効性の比較検証が水本ら [9] によってなされている。この実験では誤りのタイプを冠詞に関する間違いや動詞の時制に関する間違いなどに分類し、それぞれの誤りタイプに関して訂正の再現率、適合率、F 値を評価、分析している。実験の結果平均的にはフレーズベースモデルが最も良い性能を示している。一方、フレーズベース、階層的フレーズベースに比べて統語的モデルである Tree-to-String モデル、Tree-to-Tree モデルは高い訂正性能を示した誤りタイプもあったが全体的に性能が低い結果となったことが示されている。

本稿では水本らと同様の比較実験を、より高度な訂正であるノンネイティブによって書かれた不自然なフレーズから自然なフレーズへの訂正において行う。

3 コーパスの抽出

本稿の実験で使用するデータとしてユーザが学習している言語でエッセイを記述し、相互に添削する Web サービスである Lang-8[*1] のデータを用いた。例えば、日本語を母国語とする英語学習者は日本語学習者の書いた日本語の文章を添削し、自分の書いた英語の文章もまた英語を母国語とする他のユーザから添削してもらうといった Web サービスである。校正前後の文は対訳文とみなすことができ、大規模に訂正前後の文を収集することで、統計的機械翻訳システムを構築する。

[*1] http://lang-8.com/

表 1: 例文 (太字で示した部分が校正前後の差分を表す)

校正前	Sometimes he **is going to** around of world .
校正後	Sometimes he **goes** around **the** world .

本実験では水本ら [8] によって構築された Lang-8 Corpus of Learner English [*2] を用いる。本稿で注目する訂正はある限定的な文法間違いではなく不自然なフレーズである。そこでノンネイティブによる不自然なフレーズを次のように定義する。

条件 1. 3 単語以上連続で訂正されているもの
条件 2. 動詞の時制などの文法的な訂正は除く

条件 1 に関して　単語レベルの訂正が可能な文法的な誤り訂正と異なり、不自然なフレーズから自然なフレーズへの訂正はまとまったスパンの書き換えにより実現されると期待される。そこで本実験では 3 単語以上連続で訂正されている箇所にのみに注目してコーパスを構築する。

条件 2 に関して　文法的誤りの訂正のうち、動詞に関する訂正は条件 1 を満たすような比較的長いスパンに渡る訂正になりうる。例えば、“write”が “has been writing”に訂正されるなど訂正が連続して 3 単語以上になることも多い。学習者が書く不自然なフレーズに注目する今回の実験に関してこのような動詞句の文法的な訂正は適さないため、以下の手順により除去する。

まず Lang-8 Corpus of Learner English に含まれる文を構文解析器 Enju[*3] を用いて解析する。解析結果から、条件 1 を満たす 3 単語以上からなる訂正部分を特定し、訂正された全ての token ノードを含む最小の constituency ノードを特定する。そのカテゴリが動詞句（VP）であった場合、文法的誤りの訂正であるか判定する。

具体的には、動詞句に含まれる全ての単語の原型を調べ、該当する訂正先の単語が同一の原型をもつ場合、時制などの文法的な修正とみなしそれを含む対文をコーパスから除く。表 1 の例では校正前の “is going to”を全て含む最小の constituency ノードは動詞句を構成しており、各 token ノードの原型はそれぞれ { “be”,“go”,“to” } ある。校正後の “goes”の原型が { “go” } であり、“goes”の原型と一致するため、“is going to”→“goes”の訂正は文法的な訂正と判断する。

[*2] http://cl.naist.jp/nldata/lang-8/
[*3] http://www.nactem.ac.uk/enju/index.ja.html

4　翻訳モデルの比較実験

4.1　翻訳モデル

今回の実験では統計的機械翻訳における代表的な翻訳モデルを比較することで不自然なフレーズの訂正に有効なモデルを検討する。具体的にはフレーズベース、階層的フレーズベース、Tree-to-String を用いて比較検証する。

フレーズベースモデルでは翻訳の最少単位を任意の連続した単語列からなるフレーズとして翻訳する。一方、階層的フレーズベース、および Tree-to-String モデルは同期的文脈自由文法に基き、文の構文構造を利用して翻訳する。同期的文脈自由文法において非終端記号として X のみを用いるのが階層的フレーズベースモデルであり、句構造解析のカテゴリを用いるのが Tree-to-String モデルである。そのため、階層的フレーズベースモデルでは構文解析器は用いず、あるフレーズに含まれる部分フレーズを非終端記号に置き換えることで文法を抽出する。階層的フレーズベースおよび Tree-to-string モデルでは木構造を用いるため、不連続なフレーズの翻訳にも柔軟に対応できる。さらに構文構造に基づいた語順の変換が可能である。

4.2　実装

本実験では Moses[*4] を用いて統計的機械翻訳システムを構築した [4]。Moses は多言語間で使うことのできる統計的機械翻訳システムであり、任意の 2 言語間の対訳文を集めてコーパスを作り学習させることで多様な翻訳モデルを用いた統計的機械翻訳システムを構築できる。本実験では訂正前の文を元言語文、訂正後の文を目的言語文とし、訂正前後の文を対訳データとして翻訳システムを構築した。階層的フレーズベースではパラメータとして max phrase length を 5、Tree-to-String では Max span を 999 に設定した。

Lang-8 Corpus of Learner English には 1,037,562 文対含まれており、3 章で述べた条件に従って不自然なフレーズの訂正を含む訂正前後の文を抽出、翻訳モデルのトレーニングセットを 74,104 文、チューニング用のディベロップメントセットを 2,101 文、翻訳に用いるテストセットを 1,389 文とした。

[*4] http://www.statmt.org/moses/index.php?n=Main.HomePage

4.3 言語モデル

本実験で使用する言語モデルは Europarl v7、News Crawl:article from 2012, 2013[*5] (合計 36,846,615 文) を用い構築した。Europarl は欧州議会や国際連合などの国際的な議会の議事録であり、11 カ国語におよぶ対訳データが提供されている [3]。 そのうち英語のデータのみを言語モデルとして使用した。News Crawl は Web ソースからスクレイピングされたデータセットである。

4.4 評価指標・パラメータチューニング

実験の評価指標として自動評価尺度である BLEU 値 [10] を用いた。Lang-8 の校正後のデータを正解文として入力文に対する翻訳結果をこれと比較し、BLEU 値を計算した。

各モデルのパラメータはディベロップメントデータを用い、BLEU 値を最大化するよう Minimum error rate training [2] によりチューニングした。

4.5 実験結果

表 2 に各翻訳モデルの BLEU 値を示す。またノンネイティブによる元の文、つまり翻訳による訂正を行わない文と正解文の BLEU 値も示す。

各機械翻訳モデルの結果を見ると、フレーズベースモデルが最も高い BLEU 値を達成した。これは文法的誤りの訂正における翻訳モデルの性能を比較検討した水本らの実験結果と共通している。フレーズベースの性能が高い要因として、訂正された部分以外は元の文の状態が保持されており、構文的な構造変換が観察される割合が低かったためだと考えられる。一方で階層的フレーズベースモデルと Tree-to-string モデルを比較すると、後者の方が高い性能を示している。この結果から、カテゴリのようにより詳細な構文情報が不自然なフレーズから自然なフレーズへの訂正に有効であることが示される。今後、構文情報を訂正に適した形で用いる手法を設計することで、訂正性能を向上できる可能性がある。

しかし、いずれのモデルも翻訳前の文と比べると大きく BLEU 値を落としており、不自然なフレーズの訂正が非常に困難な課題であり、単純な統計的機械翻訳手法の適用では解決できないことが分かる。原因としてまず、自然なフレーズへの訂正は文の一部にのみなされており、それ以外の箇所は元の文のまま保持されているため、訂正部分にフォーカスした学習が困難であることが挙げられる。さらに、ノンネイティブによって書かれた入力文は不完全な構文構造を持っていると考えられ、構文情報を有効に利用できない点、また不自然なフレーズから自然なフレーズへの訂正という自由度の高い訂正を実現するような翻訳モデルを構築するにはコーパスが不足していた点が考えられる。N-best の構文木を用いるなど、構文解析誤りにロバストな手法を用いることで、訂正精度を高められると期待できる。また、訂正された部分のみ抽出したコーパスを用いた翻訳モデルを構築することで性能の改善が期待できる。

表 2: 訂正前の文および各翻訳モデルのテストセットにおける BLEU 値

翻訳前	フレーズベース	階層的フレーズベース	Tree-to-String
45.05	40.30	34.76	38.81

5 まとめ

本稿ではノンネイティブ話者によって書かれた不自然なフレーズをネイティブ話者から見た自然なフレーズに訂正することを目的とし、代表的な統計的機械翻訳モデルを用いた訂正の有効性を比較・評価した。フレーズベース、階層的フレーズベース、Tree-to-String モデルを用いて比較したところフレーズベースが最も高い BLEU 値を示したが、訂正前の文と比べるとどのモデルを用いても大きく BLEU 値が低下することが分かった。このことから自然なフレーズへの訂正は困難な課題であり、単純な統計的機械翻訳手法の適用のみでは解決困難であることが分かる。

今後、訂正部分のみ抽出したコーパスからのモデル学習、N-best の構文解析木を用いたロバストな構文情報を用いた訂正手法を検討する予定である。

参考文献

[1] David Chiang, “Hierarchical phrase-based translation,” Computational Linguistics, 33(2), pp.204-211, 2007.

[2] Franz Josef Och, “Minimum Error Rate Training in Statistical Machine Translation,” Proceedings of the Annual Meeting of the Associa-

[*5] http://www.statmt.org/wmt14/translation-task.html#download

tion for Computational Linguistics, pp.106-167, 2003.

[3] Philipp Koehn, “Europarl: A Parallel Corpus for Statistical Machine Translation,” Summit, no page numbers, 2005.

[4] Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, Evan Herbst, “Moses: Open Source Toolkit for Statistical Machine Translation,” Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), pp.177-180, 2007.

[5] Philipp Koehn, Franz Josef Och, Daniel Marcu, “Statistical Phrase-Based Translation,” Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, pp.48-54, 2003.

[6] Yang Liu, Qun Liu, Shouxun Lin, “Tree-to-string alignment template for statistical machine translation,” In Proc. ACL, pp.610-613, 2006.

[7] Yusuke Miyao, Jun’ichi Tsujii, “Feature Forest Models for Probabilistic HPSG Parsing,” Computational Linguistics, 34(1), pp.35-80, 2008.

[8] Tomoya Mizumoto, Mamoru Komachi, Masaaki Nagata, Yuji Matsumoto, “Mining Revision Log of Language Learning SNS for Automated Japanese Error Correction of Second Language Learners,” In Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP), pp.147-155, 2011.

[9] 水本智也, 松本裕治, “統計的機械翻訳に基づく英語文法誤り訂正におけるフレーズベースと統語ベースの比較と分析,” 言語処理学会第 20 回年次大会 発表論文集, pp.259-261, 2014.

[10] Papineni.K, Roukos.S, Ward.T, Zhu.W.J, “BLEU: a method for automatic evaluation of machine translation,” Proceedings of the Annual meeting of the Association for Computational Linguistics, pp.311-318, 2002.

[11] Toshikazu Tajiri, Mamoru Komachi, Yuji Matsumoto, “Tense and Aspect Error Correction for ESL Learners Using Global Context,” Proceedings of the Annual meeting of the Association for Computational Linguistics, pp.198-202, 2012.

[12] Alla Rozovskaya, Dan Roth, “Algorithm Selection and Model Adaptation for ESL Correction Tasks,” Proceedings of the Annual Meeting of the Association for Computational Linguistics, pp.924-933, 2011.