

統計的機械翻訳を用いた中国語文法誤り訂正

趙寅琛[†] 小町守[†] 石川博[†]

概要: 近年、外国語学習者のための文法誤り自動訂正についての研究が盛んになってきている。統計的機械翻訳に基づく手法で英語や日本語などの誤り自動訂正を行う研究はすでに存在するが、統計的機械翻訳手法を適用できるだけの中国語学習者コーパスが存在しなかったため、中国語についての研究はまだ少ない。そこで、本研究では統計的機械翻訳手法に基づく中国語文法誤り自動訂正手法を提案する。本論文では、中国語文法誤り訂正のためのパラレルコーパスの構築、階層的フレーズベースの訂正モデルと言語モデルの使用、誤り訂正の最適化と評価について詳しく論じていく。

キーワード: 統計的機械翻訳, 文法誤り自動訂正

Chinese Grammatical Error Correction Using Statistical Machine Translation

YINCHEN ZHAO[†] MAMORU KOMACHI[†]
HIROSHI ISHIKAWA[†]

Abstract: Recently, statistical machine translation (SMT) has become increasingly popular for grammatical error correction for second language learners. There have been some such researches focused on grammatical error correction for English and Japanese but few for Chinese since there has not existed any large-scale parallel corpora for learners of Chinese. Thus, we propose an SMT-based method to deal with Chinese grammatical error correction. In this paper we focus on building a parallel corpus for Chinese grammatical error correction; adopting hierarchical phrase-based correction model and language model; and performing optimization and evaluation of the system.

Keywords: Statistical Machine Translation, Grammatical Error Correction

1. はじめに

誤りがある文を正しい文に「翻訳」というアイデアは、Brockett ら [1] によって初めて提案された。Brockett らはノイズチャネルモデルを用いた統計的機械翻訳で外国人英語学習者 (Learners of English as a Second Language) を支援するための英文法自動訂正の手法を提案した。統計的機械翻訳の手法に基づく文法誤り訂正には言語に対応する文法知識が不要というメリットがあり、ある程度の量の正-誤文パラレルコーパスを用意すれば、どんな言語でも処理できるため、それ以降、統計的機械翻訳の手法を用いた文法誤り自動訂正についての研究が盛んになってきている。たとえば、CoNLL-2014 で開かれた英語文法誤り訂正オープンタスク [2] に参加したチームは全部で 13 チームあり、その中の 4 チームはフレーズベース統計的機械翻訳手法を利用した。特に、トップの成績を残したチーム [3] は統計的機械翻訳、言語モデルでのチェック、ルールベース変換など複数のサブシステムからなる大きなシステムを提案した。また、統計的機械翻訳に基づく日本語誤り訂正についての研究 [4] は 2011 年に発表された。

一方、統計的機械翻訳手法での中国語誤り訂正に関する

研究はまだないが、これは大規模な中国語学習者コーパスが利用できないという問題のためであった。そこで本稿では、ウェブから大規模な中国語学習者コーパスを抽出し、対訳コーパスとして用いることで、統計的機械翻訳に基づく中国語文法誤り訂正手法を提案する。

本論文の主要な貢献を以下に示す。

- 中国語文法誤り訂正のための大規模中国語学習者コーパス (正-誤文パラレルコーパス) を構築した。
- 中国語文法誤り訂正タスクにおいて初めて統計的機械翻訳手法の実用性を検証した。

本論文の構成は以下の通りである。2 節ではフレーズベースの統計的機械翻訳手法に基づく訂正手法を紹介する。3 節で統計的機械翻訳に基づく誤り訂正に必要な中国語学習者パラレルコーパスの構築について述べる。さらに 4 節で誤り訂正の評価および最適化について詳述する。最後の 5 節で実験結果と考察を示す。

2. 統計的機械翻訳に基づく中国語文法誤り訂正

2.1 統計的機械翻訳による誤り訂正モデル

統計的機械翻訳の基本理論はノイズチャネルモデルに

[†]首都大学東京システムデザイン研究科

基づいている [5]. 与えられた入力文 f に対し, 翻訳モデル $P(f|e)$ と言語モデル $P(e)$ によって様々可能な翻訳文 e を列挙し, そのなかで翻訳確率が一番高い翻訳 \hat{e} を見つけ出す. 関連研究 [1] では, 誤り文を原言語 f , そして出力文を目的言語 e として訂正タスクを翻訳タスクと見なした. そうすることによって, 統計機械翻訳における技術を誤り訂正にも適用することができる.

$$\begin{aligned}\hat{e} &= \arg \max_e P(e|f) \\ &= \arg \max_e P(f|e)P(e)\end{aligned}$$

翻訳モデルは生成モデルをより一般化し, 事後確率 $P(e|f)$ を直接対数線形モデル (log-linear model) を用いて表現することができる. ここで, $h(\cdot)$ は w により重み付けされる M 次元の素性関数である. 使える素性は翻訳モデル $P(f|e)$ と言語モデル $P(e)$ に限らず, 翻訳手法によって単語ペナルティや語順並べ替えなどもある.

$$\begin{aligned}\hat{e} &= \arg \max_e P(e|f) \\ &= \arg \max_e w^T h(e, f)\end{aligned}$$

2.2 フレーズベース訂正モデル

本研究では, 訂正モデルとして普通のフレーズベースモデル [6] と階層的フレーズベースモデル [7] という2つの翻訳モデルを使用する. 階層的フレーズベースモデルでは同期文脈自由文法に従い生成ルールを抽出する. 中国語文法誤り訂正タスクでは, 抽出された生成ルールは以下のようである:

$X \rightarrow (X1 \text{ --- } \text{好消息 } X2, \quad X1 \text{ --- } \text{一个 } \text{好消息 } X2)$

左辺の X と右辺の X_i はすべての可能なフレーズを代表している. 機械翻訳において, 階層的フレーズベースモデルは語順の差が大きい言語対により強いことが知られている. 従って, 階層的フレーズベースモデルは普通のフレーズベースモデルと比べて語順誤り訂正の精度が高いと予想される.

2.3 言語モデル

言語モデルは目的言語の生コーパスから構築するが, 目的言語で翻訳対象とする分野の生コーパスが大量に入手できないことがあるため, パラレルコーパスのなかの目的言語コーパスを利用し言語モデルを作るのが, 統計的機械翻訳における一般的な手法である. 本研究でも3節で論じるコーパスにおける正解文で言語モデルを構築する.

それ以外に, 必ずしも分野が一致しなくても, 様々なジャンルのテキストが含まれた大規模な生コーパスから言語

モデルを作成することも広く行われている. ウェブコーパスは, 多様な内容のテキストが含まれており, 大規模に入手できるため, 言語モデルの構築に適している. 今回の実験では, 大規模ウェブコーパスである Google 中国語ウェブ 5-gram データ [8] を使用する. Google 中国語ウェブ 5-gram は約 1,000 億文のコーパスから計算した N グラム頻度のデータからなるとても強力なコーパスであり, 160 万の語彙と 12 億の 5 グラムが含まれている.

2.4 文法誤りの検出

本研究のタスクは, 誤り文が与えられたとき, 誤りが発生した場所を検出し, 語順 (Disorder) 冗語 (Redundant) 脱落 (Missing) 語彙 (Selection) のいずれかのタイプを判断するというタスクである.

統計的翻訳手法を用いて翻訳モデルを訓練したあと, 入力文 (誤り文) をデコーダに入れると, 翻訳文 (訂正文) が生成される. 誤り文と訂正文を同時に与えることによって, 動的計画法で誤り文から訂正文への文字単位の編集操作の系列がわかるようになる. 編集操作の系列情報を利用して, 図 1 で示されている手続きで誤りのタイプおよび文の中に誤りが発生した場所を求める. 入力文に対応する編集操作の開始位置と終了位置によって誤りの箇所を確定でき, 各編集操作を集計することによって誤りタイプを予測できる.

例えば, 誤り文の「所以我不会让失望她」と訂正文の「所以我不会让她失望」が与えられたとすると, 文字単位の編集操作の系列 {equal, equal, equal, equal, equal, equal, equal, delete, delete, equal, insert, insert} が得られる. 入力文字列に対応する非 equal の操作の位置は「7, 9」, そして削除および挿入数がともに 0 より大きいので「Disorder」という誤りタイプを返す.

```
I = input sentence, O = translated sentence;
L = get_operations(I,O);
i = num_insert(L); d = num_delete(L);
r = num_replace(L)
f = position_firstoperation(L);
l = position_lastoperation(L);
if d > 0 and i > 0
    type ← "Disorder"
else if r > 0
    type ← "Selection"
else if d > 0
    type ← "Redundant"
else if i > 0
    type ← "Missing"
else
    type ← "correct"
end if
return type, f, l
```

図 1 誤り検出アルゴリズム

3. 中国語学習者コーパス

3.1 コーパスの抽出

本研究で構築した中国語学習者コーパスは、Lang-8¹学習者コーパスと HSK 作文コーパスという 2 つのサブコーパスからなる。

Lang-8 とは言語学習者向けの相互添削型 SNS である。Lang-8 で言語学習者が学習中の言語で日記を書くと、その言語を母語とする人が文単位で書き直してくれる。Lang-8 のデータベースから大量の正-誤対応文を抽出でき、パラレルコーパスを作ることができる [4]。

HSK とは外国人中国語学習者向けの試験である。HSK 作文コーパス²では毎回 HSK 試験の受験者が書いた作文を収集し、中国語教師による文訂正が以下のようなアノテーションとして文の中に追加されている。

アノテーション例：

这就{CQ 要}由有关部门和政策管理制度来控制。

中括弧によってアノテーションされた場所は誤りが発生した場所である。英文字が誤りタグであり、ここでは「要」という単語が欠けているという意味である。

このようなアノテーションを利用すれば対応文を簡単に抽出できると思われる。語順誤り (Disorder) についてのアノテーションには正しい語順が示されていないため、誤りタグから語順誤りの対応文を作るのが不可能である。それ以外に、「被」「把」など中国語での特殊文法についてのアノテーションを処理することも非常に難しい。そのため、我々は語順誤りに関する学習者コーパスの抽出を断念した。最終的に HSK 作文コーパスから抽出した対応文に含まれる誤りは冗語 (Redundant) 脱落 (Missing) 語彙 (Selection) という 3 つの誤用に限られる。

3.2 クリーニング

誤り文とそれに対応している正しい文との単語単位の編集距離を測れば、コーパス全体の正-誤文編集距離の分布が得られる。表 1 が示すように、本研究で最終的に評価を行う中国語学習者コーパスと比較すると、前節でウェブから抽出した中国語学習者コーパスは編集距離的にかなりの違いがあり、このコーパスをそのまま使うと適切な対応関係を学習できない恐れがある。

そこで、テストセットの編集距離の分布に合うようにコーパスをダウンサンプリングしクリーニングを行った。最終的に得られたコーパスの編集距離分布は表 1 の「クリーニング済」で示されている。抽出しクリーニングした結果

として Lang-8 学習者コーパスから 58,249 文、HSK 作文コーパスから 59,027 文、合わせて 117,276 文の正-誤文パラレルコーパスが得られた。

表 1 コーパスでの単語単位の編集距離の分布

	0	1	2	3	4	5	6	7	>7
テストセット	0.00	33.83	41.56	10.09	8.19	3.93	1.40	0.86	0.13
Lang-8	1.11	17.05	24.90	15.30	12.57	9.05	6.38	4.38	11.26
クリーニング済	0.00	36.20	44.22	8.57	6.46	2.03	1.44	0.60	0.30

4. システムの評価と最適化

用意しておいたテストセットを入力としてシステムに入れ、得られた出力と正解を参照して出力の

- 正解率 (計算式: $(TP+TN)/(TP+TN+FP+FN)$)
- 適合率 (計算式: $TP/(TP+FN)$)
- 再現率 (計算式: $TP/(TP+FP)$)
- F1 値 (計算式: $2 \times \text{適合率} \times \text{再現率} / (\text{適合率} + \text{再現率})$)
- FP 率 (計算式: $FP/(FP+TN)$)

を計算する。正解率、F1 値、FP 率はシステム性能を表した主な尺度として使われる。

CoNLL-2014 のオープンタスクにおいては訂正精度を向上させるためにいくつかシステムの最適化手法が提案された。例えば、[9] では $F_{0.5}$ を評価指標としてチューニングを行った。[10] は MERT と kb-MIRA という 2 つの最適化アルゴリズムを結合した。[9] と [10] の研究によって、適切なチューニングを用いればシステムの性能を向上させることができることが示されている。そこで、本研究では、評価尺度としての正解率、F1 値、FP 率を全部用いて最適化するために、以下の線形結合を用いる。

$$\text{Score} = \alpha \times \text{正解率} + \beta \times \text{F1} + \gamma \times (1 - \text{FP 率})$$

正規化に $\alpha + \beta + \gamma = 1.0$

正解率、F1 値、FP 率のトレードオフはハイパーパラメータによって調整する。予備実験によって、 $(\alpha, \beta, \gamma) = (0.5, 0.0, 0.5)$ に設定した。チューニングに関する問題は、次の節で詳しく述べる。

5. 実験

5.1 実験設定

WAT2015*のベースラインに従い、フレーズベースと階層的フレーズベースの中国語文法誤り訂正システムを構築した。単語分割のツールは Stanford Word Segmenter (version 2014-01-04) 単語アライメントのツールは GIZA++ v1.07 である。Berkeley Parser で中国語解析を行った。コーパスによる言語モデルと Google 中国語 5-gram 言語モデルは両方

1 <http://lang-8.com/>

2 <http://202.112.195.192:8060/hsk/index.asp/>

* <http://orchid.kuee.kyoto-u.ac.jp/WAT/>

とも IRSTLM v5.80.06 で作成した。デコーダは Moses v2.11 である。ZMERT [11] というツールによって最適化した。

テストセットは中国語文法誤り訂正オープンタスク NLP-TEA2 [12] で公式に使われ、以下のように誤り文と訂正文があり、人手で誤りタイプと位置がアノテーションされている。1,400 文の中では正しい文 (Correct) と誤り文の割合が半々で与えられ、そして誤り文では語順 (Disorder) 冗語 (Redundant) 脱落 (Missing) 語彙 (Selection) の誤りの割合が均等になっている。また文の構成が同様な開発セットが 2,100 文用意されている。

テストセットのサンプル	
Error sentence :	我送你那里
Correct sentence :	我送你到那里
(日本語訳: 私はあなたをあなたに送ってあげます)	
TYPE:	Missing

以上の設定によって、以下の 3 種類の実験を行った。

1. コーパスのサイズは結果に影響を与えるのかを明らかにするために、コーパスを分割し、サイズが違う部分コーパスを利用してシステムを訓練した。
2. 2つの訂正モデルを対比するために、全コーパスでフレーズベースシステムと階層的フレーズベースモデルを訓練した。また、大規模言語モデルを評価するために、中国語学習者コーパスによる言語モデルを Google 中国語 5-gram 言語モデルに入れ替え、実験を行った。
3. 評価尺度に対する最適化の有効性を示すために、チューニングしたかどうかの対照実験を行った。

テストセットでの評価では Position レベル (誤りのタイプと箇所) の誤り検出によって行ったが、開発セットでのチューニングと評価は Identification レベル (誤りタイプのみ) で行った。これは、後述のように本手法の Position レベルの訂正精度はまだ高くなく、予備実験では Position レベルでチューニングしても最後の結果への影響がわずかであったためである。

5.2 中国語学習者コーパスの実験結果と考察

表 2 が表しているように、コーパス全体の 1/2 しか使わない場合、訂正システムによる予測はほとんど間違っている。コーパスのサイズを線形的に増やすと、F1 値もほぼ線形的に上がっていく。一つの正一誤対応文には、誤りのある箇所は多くないため、全ての単語を翻訳しなければならない普通の機械翻訳タスクと比較すると、学習者文が正しい場合には翻訳しなくてよい誤り訂正タスクにおいては、翻訳モデルが学習した「訂正に有用な」アライメントは少ないと思われる。つまり、統計的機械翻訳を用いた誤り訂

正タスクは普通の翻訳タスクより大きなコーパスが必要である、と推測される。この知見は、Mizumoto ら [13] の研究によって英語学習者コーパスの分量が英語文法誤り訂正に与える影響を調査した結果と一致する。

本研究では実際に学習者が書いた文から誤り訂正モデルを構築したが、テストセットは正しい文と誤り文が均等になるように構成され、また誤りタイプも同数含まれているので、単に編集距離の分布が同じようにダウンサンプリングするだけでは、必ずしもテストセットの性質とパラレルコーパスの性質が同じではなかった可能性がある。従って、テストセットと誤りが同様の分布になるように、学習者コーパスをサンプリングする、といった手法が考えられる。

表 2 コーパスサイズが訂正精度に与える影響

文数	58,000	73,000	88,000	103,000	117,000
F1 値	0.013	0.025	0.064	0.094	0.108

5.3 訂正モデルと言語モデルに関する実験結果と考察

表 3 上部に示されているのは 2 つの訂正モデルによる結果である。この結果からは、フレーズベースモデル (PB) と階層的フレーズベース (HPB) との優劣をつけることはできない。HPB は PB より誤りを正しく検出できた事例の数が多いが、FP 率が高く、HPB は正しい入力文を変えてしまう傾向がある。さらに、訂正手法として重視される適合率 (Precision) が少し劣っている。

表 3 違うモデルの組み合わせによる訂正結果

		FP 率	正解率	適合率	再現率	F1 値
中国語学習者 コーパス	PB	0.134	0.449	0.193	0.032	0.055
	HPB	0.350	0.362	0.175	0.074	0.104
Google 中国語 5-gram	HPB	0.163	0.493	0.193	0.050	0.080

しかし、予想通り階層的フレーズベースモデルは普通のフレーズベースモデルより多くの語順誤りを正しく予測できた。フレーズベースモデルでは正しく予測した誤り文のなかには語順のタイプが 1 つしかない。それに対して、階層的フレーズベースモデルでは 7 つの語順誤りを正しく検出できた。一つの例を挙げてみると：

入力文	其中有一人丢护照了。		
	(日本語訳: そのなかの一人はパスポートを紛失した)		
PB 出力文	其中有一人丢护照了。		(未訂正)
HPB 出力文	其中有一人丢了护照。		(正解)

のように、「护照」と「了」いった離れたところにある単語同士を入れ替えることができています。

さらに、普通のフレーズベースモデルでは語彙誤りを1つも検出できなかったが、階層的フレーズベースモデルでは5つの誤りを検出した。

しかしながら、他の手法と比較してフレーズベース訂正手法の精度は高くない。中国語文法誤り訂正オープンタスク NLP-TEA2 [12] の最終結果を F1 値でランキングすると、本研究の階層的フレーズベースモデルによる結果はちょうど真ん中に位置する。5.2 節の結果と合わせて考えると、フレーズベースの訂正手法は大規模な学習者コーパスを必要とするため、10 万文規模の学習者コーパスではカバー率が不足し、分類器に基づく手法と比較して訂正性能が劣っている、と推測される。

本手法は誤り訂正をした後に誤り分類を行うが、誤り分類を先に行うことで訂正可能な誤りタイプに訂正先を限定して FP 率を下げる手法や、誤り訂正と誤り分類を同時に学習する手法も考えられる。

階層的フレーズベースの実験設定に基づいて、中国語学習者コーパスによる言語モデルを Google 中国語 5-gram 言語モデルに入れ替えると、表 3 の下のような結果となった。言語モデルを入れ替えた結果、F1 値が3ポイント下がったが、FP 率と正解率が10ポイント以上改善した。中国語学習者コーパスと分野は違っても、大規模なコーパスから言語モデルを構築した方が、全体的な訂正精度は向上したと考えられる。

統計的機械翻訳でもリカレントニューラルネットワーク言語モデル [14] のように過去の履歴を考慮することで予測性能を向上させる手法が提案されており、誤り訂正にも有効である可能性がある。

5.4 評価と最適化の実験結果と考察

4 節で述べたように、以上のすべての実験結果はハイパーパラメータセット $(\alpha, \beta, \gamma) = (0.5, 0.0, 0.5)$ の設定の上で得られた。このハイパーパラメータセットの有効性を証明するために、フレーズベースと階層的フレーズベースモデルでそれぞれ対照実験を行った。その結果は表 4 で示されている。 $\beta = 0$ と設定しても正解率と FP 率の改善とともに F1 値も上がった。このようなハイパーパラメータセットは望ましい設定と思われる。

表 4 チューニングが訂正精度に与える影響 (F1 値)

	PB	HPB
未チューニング	0.0513	0.0868
チューニング後	0.0701	0.1080

一方、理論的には $(\alpha, \beta, \gamma) = (0.0, 1.0, 0.0)$ に設定すれば、F1 値を最大化するように最適化することになるので、MERT でチューニングしたあと一番高い F1 値を得るはずだが、実際に開発セットで実験してみたところ、表 5 のよう

な結果になった。F1 値は 0.2 ポイント近く向上したが、FP 率が異常に高い。これは、システムが元の文を訂正しすぎ、ほとんどの文を誤り文と認定してしまった結果である。実際の訂正文も元の文から大きくかけ離れたものになってしまっているため、訂正結果として使えない。

表 5 $(\alpha, \beta, \gamma) = (0.0, 1.0, 0.0)$ に対する
 フレーズベース訂正モデルでの実験結果

	FP 率	正解率	適合率	再現率	F1 値
PB	0.997	0.146	0.144	0.993	0.251

それでは、 $(\alpha, \beta, \gamma) = (0.5, 0.0, 0.5)$ というハイパーパラメータセットは万能であるのか？ それについてさらに追加実験で string-to-tree 翻訳モデルを訂正モデルとして試した。実験設定ではまた WAT2015 のベースラインに従い、Moses デコーダを用いた。最適化の効果を検討するため、開発セットでチューニングしたシステムとチューニングしていないシステムを比較した。

実験結果は表 6 で示されている。正解率と FP 率が改善されたが、F1 値の結果が極めて低い。これは、 $\beta = 0$ としたために F1 値を無視するチューニングとなってしまったせいである。システムはほとんどの誤りタイプを「Correction」に予測してしまうので、F1 値で最適化した実験結果とは逆で、ほとんどの入力では元の文を全く変えないため、訂正結果としては受け入れられない。string-to-tree 翻訳モデルでの訂正タスクには、 $(\alpha, \beta, \gamma) = (0.5, 0.0, 0.5)$ とは別の適切なハイパーパラメータセットが必要だと思われる。

一つの方針として、機械翻訳の複数の評価尺度に対して最適化する際に、Duh ら [15] の研究でパレート最適化を用いる手法が提案されており、誤り訂正にもパレート最適化が有効である可能性が考えられる。

表 6 $(\alpha, \beta, \gamma) = (0.5, 0.0, 0.5)$ に対する
 string-to-tree 訂正モデルでの実験結果

	FP 率	正解率	適合率	再現率	F1 値
未チューニング	0.3973	0.4087	0.1042	0.0787	0.0896
チューニング後	0.1029	0.4747	0.0480	0.0057	0.0102

追加実験では人手でいくつかのハイパーパラメータセットを設定し行われた。MERT などの最適化アルゴリズムでハイパーパラメータセットを自動学習することが可能であるが、本実験では ZMERT によりパラメータセットを最適化しており、ハイパーパラメータセットまで自動で最適化するのは計算量が非常に膨大になるため、我々の実験では試さなかった。

6. おわりに

本稿では、統計的機械翻訳を用いた中国語文法誤り訂正手法について説明した。中国語文法誤り訂正に必要な学習者コーパスを構築する手法を紹介し、語順誤りに強い階層的フレーズベースの使用を提案した。さらに訂正精度を向上するために、評価尺度で線形スコアを作りチューニングをした。最後に、システムのそれぞれの構成要素の効果を検証した実験について述べた。

5.2節に述べたように、学習者コーパスの規模が11万文では、本稿で対象としたような誤りを訂正するために統計的機械翻訳手法を適用するにはまだ不足していると思われる。コーパスをさらに拡充するために、Lang-8やHSKのような文法誤り訂正に特化したリソースからコーパスを抽出するだけでなく、関連研究 [3] で用いられた手法のように人工的に擬似負例を作成しコーパスを構築することも可能である。また、5.4節で論じた最適化の問題に対して、ベイジック最適化 [16] など、洗練された最適化手法を適用することが今後の課題である。

謝辞 Lang-8 のデータベースを提供していただいた喜洋洋氏に、謹んで感謝の意を表す。

参考文献

- 1) Chris Brockett, William Dolan, Michael Gamon. (2006). Correcting ESL Errors Using Phrasal SMT Techniques. *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, pages 249–256. Sydney, Australia.
- 2) Hwee Tou Ng, Siew Mei Wu, Ted Briscoe, Christian Hadiwinoto, Raymond H. Susanto, Christopher Bryant. (2014). The CoNLL-2014 Shared Task on Grammatical Error Correction. *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*, pages 1–14, Baltimore, Maryland.
- 3) Mariano Felice, Zheng Yuan, Øistein E. Andersen, Helen Yannakoudakis, Ekaterina Kochmar. (2014). Grammatical Error Correction using Hybrid Systems and Type Filtering. *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*, pages 15–24, Baltimore, Maryland.
- 4) 水本智也, 小町守, 永田昌明, 松本裕治. (2013). 日本語学習者の作文自動誤り訂正のための語学学習 SNS の添削ログからの知識獲得. *人工知能学会論文誌*, Vol.28, No.4, pp.420–432.
- 5) Peter F. Brown, John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin. (1990). A Statistical Approach to Machine Translation. *Computational Linguistics*, Volume 16, Number 2, pages 79–85, MIT Press, Cambridge, MA, USA
- 6) Philipp Koehn, Franz Josef Och, Daniel Marcu. (2003). 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*, pages 48–54, , Edmonton, Alberta, Canada
- 7) David Chiang. (2005). A Hierarchical Phrase-Based Model for

- Statistical Machine Translation. *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL '05)*, pages 263–270, Ann Arbor, Michigan.
- 8) Fang Liu, Meng Yang, Dekang Lin. (2010). Chinese Web 5-gram Version 1. Linguistic Data Consortium, Philadelphia.
- 9) Anoop Kunchukuttan, Sriram Chaudhury, Pushpak Bhattacharyya. (2014). Tuning a Grammar Correction System for Increased Precision. *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*, pages 60–64, Baltimore, Maryland.
- 10) Marcin Junczys-Dowmunt, Roman Grundkiewicz. (2014). The AMU System in the CoNLL-2014 Shared Task: Grammatical Error Correction by Data-Intensive and Feature-Rich Statistical Machine Translation. *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*, pages 25–33, Baltimore, Maryland.
- 11) Omar F. Zaidan. (2009). Z-MERT: A Fully Configurable Open Source Tool for Minimum Error Rate Training of Machine Translation Systems. *The Prague Bulletin of Mathematical Linguistics*, Number 91, pages 79–88.
- 12) Lung-Hao Lee, Liang-Chih Yu, Li-Ping Chang. (2015). Overview of the NLP-TEA 2015 Shared Task for Chinese Grammatical Error Diagnosis. *Proceedings of The 2nd Workshop on Natural Language Processing Techniques for Educational Applications*, pages 15–19, Beijing, China.
- 13) Tomoya Mizumoto, Yuta Hayashibe, Mamoru Komachi, Masaaki Nagata, Yuji Matsumoto. (2012). The Effect of Learner Corpus Size in Grammatical Error Correction of ESL Writings. *Proceedings of The 22th International Conference on Computational Linguistics: Posters*, pages 863–872, Mumbai, India.
- 14) Tomas Mikolov, Martin Karafiat, Lukas Burget, Jan “Honza” Cernocky, Sanjeev Khudanpur. (2010). Recurrent neural network based language model. *Proceedings of International Speech Communication Association*, pages 1045–1048, Makuhari, Chiba, Japan.
- 15) Kevin Duh, Katsuhito Sudoh, Xianchao Wu, Hajime Tsukada, Masaaki Nagata. (2012). Learning to Translate with Multiple Objectives. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 1–10, Jeju, Republic of Korea.
- 16) Jasper Snoek, Hugo Larochelle, Ryan P. Adams. (2012). Practical Bayesian Optimization of Machine Learning Algorithms. *Proceedings of the Advances in Neural Information Processing Systems 25*. Harrahs and Harveys, Lake Tahoe, USA.