

Project Next 英文校正タスク: 前置詞誤りを対象とした誤り分析

水本智也 三田雅人
奈良先端科学技術大学院大学

{tomoya-m, mita.masato.mz2}@is.naist.jp

1 はじめに

英文校正と言っても、文章を書いた人のレベルなどによって行なう処理が異なる。例えば、母語話者が書いた文章であれば、校正する箇所は細かいところ（言い回し等）であるが、非母語話者の書いた文章であれば、校正する箇所はスペル誤りや文法誤りの修正が多くなる。本稿では、非母語話者の文章の誤りを訂正するタスクに焦点を絞る。

企業のグローバル化が進んでおり、英語非母語話者が英語を使う機会が多くなっている。英文校正の需要も高まっており、Ginger^{*1}や1Checker^{*2}といった無料で使用できる自動英文校正のサービスも増えている。英語学習者の文法誤りの訂正を行なう研究も盛んになっており、英語の文法誤り訂正の性能を競うコンペティションも4年連続で開催されている [3, 2, 12, 11]。

英語の文法誤り訂正の研究が数多く行なわれるようになり、様々な手法が提案されている。古くはルールベースによる文法チェックツール [8] があり、近年では統計的手法が用いられている。分類器を用いた文法誤り訂正 [13, 7, 15] や統計的機械翻訳の手法を応用した文法誤り訂正 [9, 4, 5] が研究されている。

多くの手法が提案されている一方で、どの手法がどのような誤りの種類に有効であり、どのような誤りを苦手としているかは分析されていない。そこで本稿では、現在一般的に用いられている分類器ベースの手法と統計的機械翻訳ベースの手法の誤り訂正の結果から、どのような違いが見られるかを分析する。また、前置詞の Semantic Relation に注目して分析を行なう。学習者の書いた前置詞が、Semantic Relation 内の前置詞で混同しているのか、システムが訂正した後の前置詞が正しい Semantic Relation 内の前置詞に訂正されているか調査する。

^{*1} <http://www.getginger.jp>

^{*2} <http://www.1checker.com>

表1 Konan-JIEM Corpus における誤りの分布

タイプ	割合 (%)	タイプ	割合 (%)
冠詞	19.23	動詞その他	4.09
名詞の単複	13.88	副詞	3.59
前置詞	13.56	接続詞	2.04
動詞の時制	8.77	語順	1.34
名詞の語彙選択	7.04	名詞その他	1.30
動詞の語彙選択	6.90	助動詞	0.88
代名詞	6.62	語彙選択その他	0.74
動詞の人称・数の不一致	5.25	関係詞	0.42
形容詞	4.30	疑問詞	0.04

2 分析対象の誤り

学習者の犯す誤りには様々なタイプがある。日本語を母語とする大学生の書いた英作文から作られた、誤り訂正情報付き学習者コーパス“Konan-JIEM Corpus” (以後、KJ コーパス) [10] の誤りの分布を表1に、シンガポールの大学生の書いた英作文から作られた“The NUS Corpus of Learner English” (以後、NUCLE コーパス) [1] の誤りの分布を表2に示す。アノテーションの方法等が異なるが、KJ コーパスで冠詞誤りの割合が19.23%、前置詞誤りの割合が13.56%、NUCLE コーパスで冠詞誤りの割合が12.88%、前置詞誤りの割合が15.69%^{*3}と、どちらの学習者コーパスにおいても前置詞と冠詞の誤りの数が多い。

冠詞と前置詞の誤りは学習者の作文で多く現れるため、現在の英語の文法誤り訂正ではこの2つの誤りが扱われることが多い。HOO2011 [3] と CoNLL2014 Shared Task [11] は全ての誤りタイプが対象であったが、HOO2012 [2] は冠詞と前置詞、CoNLL2013 Shared Task [12] は冠詞、前置詞、動詞の一致、名詞の単複、動詞の活用を対象であった。また、対象は全ての誤りタイプであったとしても、冠詞と前置詞に絞って訂正を行なうチームも多い。

冠詞と前置詞の訂正を比べると、冠詞の方が訂正性能

^{*3} NUCLE コーパスでは、前置詞誤りだけでなくコロケーション、イディオムの誤りもカウントされている。

表2 The NUS Corpus of Learner English における誤りの分布

タイプ	割合 (%)	タイプ	割合 (%)
Wrong collocation/idiom/preposition	15.69	Verb tense	7.06
Local redundancies	13.71	Word form	4.81
Article or determiner	12.88	Subject-verb agreement	3.38
Noun number	8.49	Other errors	3.29
Mechanics	7.06	Verb form	3.29

表3 CoNLL2013 Shared Task, CoNLL2014 Shared Task の冠詞, 前置詞の訂正結果のそれぞれ最も高いものを示す。CoNLL2014 は Recall のみ報告された。

	CoNLL Shared2013 Shared Task			CoNLL2014 Shared Task	
	再現率	適合率	F 値	再現率	
冠詞	25.65	47.84	33.40	58.85	
前置詞	12.54	29.10	17.53	38.23	

が高い。表3に CoNLL2013 Shared Task の冠詞, 前置詞の訂正の結果で最も F 値が高いものと, CoNLL2014 Shared Task の冠詞, 前置詞の訂正の結果で最も再現率^{*4}が高いものを示す。CoNLL2013 Shared Task, CoNLL2014 Shared Task とともに前置詞の訂正結果よりも冠詞の訂正結果のほうがよい。

そこで本稿では, 学習者の誤りが多く, 訂正の難しい前置詞誤りを対象として分析を行なう。前置詞の種類は多いが, 学習者の文であまり見られない前置詞も多い。本稿では, 使用頻度の高い上位 12 個の前置詞 (約 91% をカバー) に対象を絞って分析を行なう。12 個の前置詞は, “of”, “in”, “for”, “to”, “by”, “with”, “at”, “on”, “from”, “as”, “about”, “since” である。

3 分析に用いるデータ

誤り分析には CoNLL2014 Shared Task で用いられた評価データ^{*5} (以後, CoNLL-2014 TestSet) を用いる。CoNLL-2014 Test Set は全部で 1,312 文あり, 対象の前置詞が含まれている文は 1,217 文であった。CoNLL-2014 Test Set では, 学習者の書いた文に対する訂正は, 複数のアノテータで行われており, 訂正箇所が異なる場合がある。本稿では, 分析の簡単化のためアノテータ 1 人の訂正のみを対象とする。

CoNLL2014 Shared Task は参加チームの出力結果^{*6}を公開している。本稿では, この Shared Task の参加チームの出力結果をエラー分析に使用する。公開されているシステム出力結果の内, ベースの手法の種類ご

とに最も精度の高いものを分析に用いる。次節でシステムの概要を述べる。

4 分析に用いるシステムの概要

英語の文法誤り訂正で, 現在主に用いられている手法は分類器ベースと統計的機械翻訳ベースである。本稿では, CoNLL2014 Shared Task で分類器ベース, 統計的機械翻訳ベースでそれぞれ最も高精度のチームの出力結果を用いる。分類器ベースの手法のものとして “Columbia University and the University of Illinois at Urbana-Champaign” チーム (以後, CUUI) [13] を用い, 統計的機械翻訳ベースの手法のものとして “University of Cambridge” チーム (以後, CAMB) [4] を用いる。

表4に各システムの手法と使用しているコーパスの概要を示す。CUUI はナイーブベイズ分類器と平均化パーセプトロンの出力をマージを行なうことで前置詞の訂正を行なった。ナイーブベイズ分類器は, Google Web 1T コーパスのような大規模な言語モデルから学習しており, 事前分布として学習者の母語による間違えやすさを考慮している。平均化パーセプトロンは, 学習者コーパスを用いて学習を行なっている。CUUI は学習者コーパスのデータ不足を解消するために, 学習者コーパスの分布に基づいて人工的にエラーを作った。

CAMB は統計的機械翻訳の結果を言語モデルによってランキングしている。コーパスは, CUUI と異なり多くのコーパスを使用しており, 学習者コーパスも大規模な “Cambridge Learner Corpus” を使用している。

5 分析方法

本稿では以下の大きく分けて 2 種類の分析を行なう。

1. 各システムで訂正できる誤りの訂正タイプによる

^{*4} 全ての誤りを対象とした場合, 訂正する必要のない箇所を訂正した場合の False Positive のカウントが難しく適合率の計算ができない。

^{*5} <http://www.comp.nus.edu.sg/~nlp/conll14st/conll14st-test-data.tar.gz>

^{*6} http://www.comp.nus.edu.sg/~nlp/conll14st/official_submissions.tar.gz

表4 CUUI (分類器ベース) と CAMB (統計的機械翻訳ベース) の手法の概要

	手法	使用しているコーパス
CUUI	分類器	CoNLL-2013 Test Set, Google Web 1T
CAMB	言語モデル, 統計的機械翻訳	Cambridge Learner Corpus, CoNLL-2013 Test Set, First Certificate in English corpus, English Vocabulary Profile corpus, Microsoft Web LM

分析

2. 前置詞の Semantic Relation に注目した分析

以下, 前置詞の誤り訂正タイプによる分析と Semantic Relation による分析について説明を行なう。

5.1 前置詞の誤り訂正タイプによる分析

前置詞の誤り訂正タイプごとに訂正性能を調べることで, 各手法の訂正の特徴を明らかにする。前置詞誤りの訂正タイプは, 訂正の仕方から選択, 抜け, 余剰の3つに分類できる。選択は下の例のように, 学習者が間違っている“by”を“to”に訂正するような誤りである。

学習者: ... face by face
 正解: ... face to face

3つの訂正タイプの中では最も頻度が高い。分類器を使った訂正が容易タイプである。

抜けは下記のように, 本来は必要である“in”を学習者が書いていないような誤りである。

学習者: ... social media to get _ contact with our friends.
 正解: ... social media to get in contact with our friends.

システムによる訂正の場合, 前置詞の抜けている箇所を検出する必要があるため, 訂正が難しいタイプであり, 前置詞誤り訂正の研究では対象としていないことが多い。

余剰は下記の例のように, 学習者が余分に書いた“for”を削除するような誤りである。

学習者: It is normal for people to demand for treatment
 正解: It is normal for people to demand _ treatment

余剰の場合, 元からある単語が必要かどうか判断を行なうだけでよいと, 抜けに比べると対象としている研究が多い。

5.2 前置詞の Semantic Relation による分析

前置詞の Semantic Relation は, 前置詞とその係り関係にある単語によって決まる意味的關係性であり, 文献 [14] で提案された。Srikumar and Roth [14] は, 前置詞の Sence を使って意味的に関連している前置詞をグループにまとめ, Semantic Relation を定義した。

前置詞の Sence は The Proposition Project (TPP) [6] によって, Oxford Dictionary of English の定義に基づいて 334 個の前置詞に対して 847 個定義された。SemEval

2007 Shared Task における前置詞の語義曖昧性解消のタスクでは, TPP で定義された前置詞の Sense を基に, 34 個の前置詞を対象に前置詞の Sence がアノテーションされた。

Srikumar and Roth [14] は, SemEval 2007 Shared Task でアノテーションされた Sense を使って, 意味的に関連している前置詞をグループにまとめ, 32 個の Semantic Relation ラベルを定義した。以下の例を使って, Sense Relation ラベルについて簡単に説明する。

1. Poor care led to her death from pneumonia.
2. copied from the film
3. recognized from the start

(1) において前置詞“from”は, 係り関係にある単語 death, pneumonia によって Semantic Relation は“Cause”となり, (2) の“from”の Semantic Relation は“Source”となり, (3) の“from”の Semantic Relation は“Temporal”となる。このように同じ前置詞でも異なる Semantic Relation となる。また, 以下の例,

1. reported on September 26

の前置詞“on”では, Semantic Relation は“Temporal”となり異なる前置詞でも, 同じ Semantic Relation を取る。Semantic Relation によって取りうる前置詞が決まっており, 本稿対象前置詞では, Semantic Relation “Cause”は, “at”, “for”, “from”, “of”, “with”の5つの前置詞のみ取りうる。

Semantic Relation による分析では, 以下のことを明らかにする。

1. 学習者の前置詞と正解の前置詞の Semantic Relation の違い
2. システムの訂正後の前置詞と正解の前置詞の Semantic Relation の違い

本稿では上記のことを調べるため, CoNLL-2014 TestSet の選択誤りを含む文 97 文中の 66 文の学習者の文の前置詞, 訂正後の前置詞それぞれに Semantic Relation の付与を行なった。訂正前もしくは訂正後が前置詞ではない場合, 訂正前もしくは訂正後が To 不定詞となっている場合, 訂正が複数単語にまたがる場合, 本稿で対象と

している前置詞ではない場合は Semantic Relation 付与の対象外とした。結果として 45 個の前置詞に対して、Semantic Relation の付与を行なった。

正解の文に対して Semantic Relation を付与する場合と違い、学習者の書いた誤った前置詞に Semantic Relation を付与する場合は、前置詞が正解の Semantic Relation に属していない場合などがあり難しい場合がある。例えば、“death from pneumonia.” の “from” は Semantic Relation “Cause” であるが、学習者が “death by pneumonia.” と書いた場合、“by” は “Cause” を取らないため、“Cause” を付与することができない。学習者の文に対して、Semantic Relation を付与する場合は、(1) 学習者の前置詞が取りうる Semantic Relation 群にあり、(2) 前置詞の governor, object の品詞が Semantic Relation の取りうる品詞と同じであり、もっとも適切であると判断したものを選択した。

6 各手法ごとの前置詞訂正の性能評価

4 節で挙げたチームの出力結果を用いて分析を行なう。各チームがどの程度前置詞誤りを訂正できるか調査する。精度を測る指標として、適合率、再現率、 $F_{0.5}$ を用いる。 $F_{0.5}$ は CoNLL2014 Shared Task で用いられており、学習者の文法誤り訂正においては再現率よりも適合率が重要であるとされているためである。適合率、再現率、 $F_{0.5}$ は以下の通りである。

$$\text{適合率} = \frac{TP \text{ の総数}}{TP \text{ の総数} + FP \text{ の総数} + FP_NA \text{ の総数}}$$

$$\text{再現率} = \frac{TP \text{ の総数}}{TP \text{ の総数} + FN \text{ の総数}}$$

$$F_{0.5} = \frac{(1 + 0.5^2) \times \text{適合率} \times \text{再現率}}{(0.5^2) \times \text{適合率} + \text{再現率}}$$

適合率、再現率の計算には、TP, FN, FP, FP_NA の数を用いる。TP は、システムが訂正し、実際に正しかった箇所を示す。FN は、システムが訂正しなかったが、実際は訂正する必要があった箇所を示す。FP と FP_NA は、システムが訂正したが、異なる前置詞（もしくは前置詞を削除）に訂正した箇所を示す。FP はアノテーションが付いている箇所、すなわち他の前置詞に訂正する必要がある箇所である（訂正しなければ FN となる）。一方、FP_NA はアノテーションが付いていない箇所、すなわち学習者の前置詞が正しく訂正する必要がない箇所である（訂正しなければ TN となる）。FP と FP_NA は、間違っている箇所を間違っただけで訂正しているか、正しい箇所を間違っただけで訂正しているかで性質が異なるため区別する^{*7}。

^{*7} FP は誤り検出の観点では正しくできている。

表7 学習者の前置詞、システムの出力した前置詞と正解の前置詞の Semantic Relation ラベルの違いの比較。学習者、システムの Semantic Relation ラベルが正解と同じ場合 Correct、そうでない場合 Incorrect とする。

	Correct	Incorrect
学習者	14	31
CUUI	17	28
CAMB	21	24

表5に全ての前置詞対象とした場合の結果を示す。適合率、再現率ともに統計的機械翻訳を用いた CAMB の方が高く、 $F_{0.5}$ においても CAMB の方が高い。特に再現率では2倍以上の差がある。

FP_NA を考慮しない場合の適合率と $F_{0.5}$ についても計算を行なったが、アノテーションの付いている箇所のみであれば高い値を達成できていることがわかる。CUUI の FP_NA が 75 個、CAMB の FP_NA が 179 個と、両手法とも訂正する必要のない箇所を訂正している数が多い。訂正する必要のない箇所のエラーを減らすことが今後の1つの課題である。

表6に前置詞誤りの訂正タイプ別の結果を示す。選択、抜けに関しては CAMB の方が適合率、再現率、 $F_{0.5}$ 全てで CUUI を上回っている。抜けに関しては、分類器ベースの CUUI は対象としていないため訂正できていないが、統計的機械翻訳ベースの CAMB では訂正ができている。余剰に関しては、CUUI の方が適合率が高く、 $F_{0.5}$ においても CAMB より高い値となっている。

7 Semantic Relation に基づく分析

5.2 節で Semantic Relation をアノテートしたデータを用いて、学習者の書いた前置詞と正解の前置詞で Semantic Relation に違いがあるか、システムの出力した前置詞は正解の前置詞の Semantic Relation と同じラベルになっているかを調べる。表7に、学習者、システムの前置詞の Semantic Relation が正解の前置詞の場合と同じである数、そうでない数を示す。学習者が誤っている前置詞の半数以上は、Semantic Relation を越えた誤りであった。Semantic Relation を越えた誤りに関しては学習者の母語に影響が大きいと考える。

システムの出力した前置詞と正解の Semantic Relation を比較すると、学習者の場合よりも Correct が増えており、正解の前置詞と同じ Semantic Relation ラベルの取りうる前置詞に訂正している。しかしながら、Correct になっているものはシステムが正しい前置詞に訂正したもの (True Positive) であり、False Positive で正解と同じ Semantic Relation ラベルを取りうる前置詞に訂正したものはなかった。

本稿で用いた学習者コーパスは、学習者のレベルも

表5 全ての前置詞に対する誤り訂正の結果

	TP	FN	FP	FP_NA	適合率	再現率	$F_{0.5}$	適合率 w/o FP_NA	$F_{0.5}$ w/o FP_NA
CUUI	18	177	11	75	17.31	9.23	14.73	62.07	28.94
CAMB	47	136	26	179	18.65	25.68	19.73	64.38	49.47

表6 前置詞誤りの訂正タイプによる結果. FP_NAはカウントしていないため、適合率及び $F_{0.5}$ の値は正確ではない。

		TP	FN	FP	FP_NA	適合率	再現率	$F_{0.5}$
CUUI	選択	10	87	9	-	52.63	10.31	28.90
	余剰	8	58	2	-	80.00	29.41	60.98
	抜け	0	32	0	-	-	-	-
CAMB	選択	19	75	14	-	57.57	20.21	42.04
	余剰	20	38	10	-	66.67	34.48	56.18
	抜け	8	23	2	-	80.00	25.81	56.34

高く、またドメインも健康などに対して意見を述べる論述形式であった。そのため一般的に第2言語学習者が作文中でよく間違える Location の at, in, on の混同, Temporal の at, in, on の混同はほとんど出現しなかった。

8 おわりに

本稿のエラー分析では、英文校正の中でも特に学習者の書いた作文の誤りを訂正するタスクに焦点を絞った。学習者の犯す誤りのタイプは数多くあるが、数が多く訂正の難しい前置詞を対象とした。

英語誤り訂正の手法として様々な手法が提案されているが、どの手法がどのようなタイプの誤りを訂正可能であるかは議論されていない。本稿では、現在主に使用されている分類器ベースの手法と統計的機械翻訳の手法で比較、分析した。適合率、再現率、 $F_{0.5}$ の値全てで統計的機械翻訳をベースとした方が高い値であった。適合率については、両手法において訂正する必要のない箇所を多く訂正していたため低い値であった。

学習者コーパスに対して Semantic Relation ラベルのアノテーションを行ない、学習者の書く前置詞が Semantic Relation 内で関係のある誤りをしているか調べた。誤りの半数以上は Semantic Relation を越えた誤りであり、母語など他の影響を大きく受けていると考える。

本稿では、システムとして CoNLL Shared Task の出力を用いたが、今後は自分たちで構築したシステムを用い、コーパス条件などを揃えて実験を行なう。また、学習者コーパス全体に Semantic Relation のアノテーションを行ない、さらに深い分析を行なう予定である。

参考文献

- [1] D. Dahlmeier, H.T. Ng, and S.M. Wu, "Building a Large Annotated Corpus of Learner English: The NUS Corpus of Learner English," Proceedings of BEA, pp.22-31, 2013.
- [2] R. Dale, I. Anisimoff, and G. Narroay, "HOO 2012: A Report

- on the Preposition and Determiner Error Correction Shared Task," Proceedings of BEA, pp.54-62, 2012.
- [3] R. Dale and A. Kilgarriff, "Helping Our Own: The HOO 2011 Pilot Shared Task," Proceedings of ENLG, pp.242-249, 2011.
- [4] M. Felice, Z. Yuan, Ø.E. Andersen, H. Yannakoudakis, and E. Kochmar, "Grammatical error correction using hybrid systems and type filtering," Proceedings of CoNLL Shared Task, pp.15-24, 2014.
- [5] M. Junczys-Dowmunt and R. Grundkiewicz, "The amu system in the conll-2014 shared task: Grammatical error correction by data-intensive and feature-rich statistical machine translation," Proceedings of CoNLL Shared Task, pp.25-33, 2014.
- [6] K. Litkowski and O. Hargraves, "The Preposition Project," Proceedings of ACL-SIGSEM Workshop, pp.171-179, 2005.
- [7] X. Liu, B. Han, and M. Zhou, "Correcting Verb Selection Errors for ESL with the Perceptron," Proceedings of the 12th international conference on Computational linguistics and intelligent text processing - Volume Part II, CICLing'11, Berlin, Heidelberg, pp.411-423, Springer-Verlag, 2011.
- [8] N.H. MacDonald, T.F. Lawrence, S.G. Patricia, and A.K. Strcey, "The Writer's Workbench: Computer Aids for Text Analysis," IEEE Transactions on Communications, vol.30, no.1, pp.105-110, 1982.
- [9] T. Mizumoto, Y. Hayashibe, M. Komachi, M. Nagata, and Y. Matsumoto, "The Effect of Learner Corpus Size in Grammatical Error Correction of ESL Writings," Proceedings of COLING, pp.863-872, 2012.
- [10] R. Nagata, E. Whittaker, and V. Sheinman, "Creating a Manually Error-tagged and Shallow-parsed Learner corpus," Proceedings of ACL-HLT, pp.1210-1219, 2011.
- [11] H.T. Ng, S.M. Wu, T. Briscoe, C. Hadiwinoto, R.H. Susanto, and C. Bryant, "The CoNLL-2014 Shared Task on Grammatical Error Correction," Proceedings of CoNLL Shared Task, pp.1-14, 2014.
- [12] H.T. Ng, S.M. Wu, Y. Wu, C. Hadiwinoto, and J. Tetreault, "The CoNLL-2013 Shared Task on Grammatical Error Correction," Proceedings of CoNLL Shared Task, pp.1-12, 2013.
- [13] A. Rozovskaya, K.W. Chang, M. Sammons, D. Roth, and N. Habash, "The illinois-columbia system in the conll-2014 shared task," Proceedings of CoNLL Shared Task, pp.34-42, 2014.
- [14] Srikumar, Vivek and Roth, Dan, "Modeling Semantic Relations Expressed by Prepositions," Transactions of the Association for Computational Linguistics, vol.1, pp.231-242, 2013.
- [15] P. Wang, Z. Jia, and H. Zhao, "Grammatical error detection and correction using a single maximum entropy model," Proceedings of CoNLL Shared Task, pp.74-82, 2014.