Vol.2015-NL-222 No.11
2015/7/16

ベクトル空間モデルを用いた英文コロケーション誤り訂正

橋本 捷人[1,a)]　相澤 彰子[2,1,b)]

概要：文法誤り訂正 (GEC) タスクでは，構文的な誤りは比較的扱いやすいのに対して，意味的な誤りの訂正は難しいことが知られている．本研究では，意味的な誤りの代表例であるコロケーション誤りに注目して，ベクトル空間モデルの誤り訂正手法における有効性を検証する．本研究における誤り訂正では，まずコロケーション誤りが検出されたものとして，N グラムを用いたモデルで訂正語の候補を複数選び，次にその候補を誤り語との類似度によって並び替える．ここで，並べ替えの際に，コロケーション誤りと判断される「誤り語」は訂正後の正しい語と意味的に類似しているという観察に基づき，ベクトル空間モデルから得られる単語の類似度を用いる．実験では，注釈付きコーパスを用いて評価を行い，ベクトル空間モデルを用いることで，正答率が上昇することを示す．また，ベクトル空間の構成法による違いを分析する．

1. はじめに

文法誤り訂正 (grammatical error correction, GEC) は，文中の文法誤りを検出し訂正するタスクである．GEC は，自然言語処理の技術の発展により挑戦が可能になった総合的なタスクとして近年注目を集めており，例えば，2011 年と 2012 年に Helping Our Own (HOO) [1], [2]，2013 年と 2014 年に CoNLL [3], [4] で共通タスクが開催されている．CoNLL で用いられた注釈付きコーパスである NUCLE コーパス [5] では，誤りは 28 種類に分けられている．この分類に基づき，CoNLL-2014 での GEC タスクの分析 [6] では，意味的な誤りは構文的な誤りに比べて訂正が難しいことを指摘している．たとえば，冠詞や前置詞などの構文的な誤りは， HOO-2012 と CoNLL-2013 で主な訂正の対象とされるなど，多く研究されている．このような構文的誤りは，意味的な誤りに比べて，誤りのパターンが限られているため訂正が比較的扱いやすいのだと考えられる．

本稿では， NUCLE コーパスを使い，その中で全体の誤りの 11.8% を占めるコロケーション誤りの訂正に焦点をあてて検討を行う．コロケーション誤りとは，構文的に正しいが慣用的でない単語の組み合わせのことである．例えば，以下の文 “This situation may cause a serious challenge to the community.” は，単語 cause が不適切とされる．なぜなら，“cause a challenge” という組み合わせが使われないからであり，cause の代わりに， pose を使うなどの訂正が必要となる．意味誤りの中では，コロケーション誤りは頻出しているが，コロケーション誤りに注目した研究は少ない．

本稿では，誤り語と訂正語の類似性を利用して，コロケーション誤りを訂正することを試みる．ここで，単語間の類似度を測るために，単語のベクトル表現を与えるベクトル空間モデル (VSM) を用い，近年様々なタスクで有効性が示されている単語のベクトル表現の有効性を検証する．筆者らが知る限りでは，単語のベクトル表現を GEC に適用する手法はまだ用いられていない．

VSM では各単語はコーパス中の文脈で特徴づけられ，ベクトルとして表現される．単語間の類似度はベクトル表現の間の類似度で表現することができる．本稿ではベクトル間の類似度を測るために，広く用いられているコサイン尺度を用いる．提案手法は言語モデルに基づく手法で訂正の候補語を選び，それらをベクトル表現によって得られる類似度スコアによって並び替えて訂正語を得る．

提案手法を NUCLE コーパスを用いて評価し， VSM を用いた並び替えがコロケーション誤り訂正の正答率を改善することを示す．複数の VSM を用いて，空間の性質によって性能がどう変わるかを比較する．Turney は， VSM は構成法によって，分野的な類似性を捉えるものや，機能的な類似性を捉えるものがあるということを指摘し，ベクトル空間の性質を理解することの重要性を唱えた [7] ．先のコロケーション誤りの例 “This situation may cause a serious challenge to the community.” では，誤り語 cause とその訂正語 pose の間には，動詞原形であるという機能的な類似性と，「発生させる」という意味的な類似性が見ら

[1] 東京大学大学院情報理工学系研究科
[2] 国立情報学研究所
[a)] hashi@nii.ac.jp
[b)] aizawa@nii.ac.jp

れる．このような違った性質の類似性を，どのような性質の VSM を用いることで捉えることができるのか，本稿では予備的な段階として分析を試みた．また，コーパスの注釈を基準訂正語とする自動評価に加え，人手による評価も行い，評価方法の正当性を確かめる．

2. 関連研究

文献 [4] では，GEC に対する手法を，ルールベース法，機械学習法，言語モデル法，機械翻訳法の 4 種類に分類し，コロケーション誤りに対しては，言語モデル法や機械翻訳法を用いたシステムが有効であったことと報告している．

ルールベースに基づく手法は，注釈付きコーパスなどから誤り訂正のルールを自動で生成し，それに従い訂正を行う手法である．この手法では，コーパスに出現する誤りに対しては有効で，正確な訂正が行える．しかし，コロケーション誤りを訂正するには非常に多くのルールが必要になるため，限られた注釈付きコーパスを元にルールを生成するルールベース手法は向いていない．

機械学習を用いる手法は，単語列をどのような誤りであるかというラベルに分類する分類器を構成する．このラベルから，どのような訂正を適用すれば良いのかを判断し，実際の訂正を行う．例えば，「名詞を複数形に変える」や「 the を加える」といった具体的な訂正がラベルとなる．この手法は，訂正の仕方にパターンの見られる動詞の時制や冠詞，名詞の単複といった構文的な誤りには適しているが，訂正が誤りによって異なるコロケーション誤りを扱うには向いていないと言える．

言語モデルに基づく手法では，コーパスから計算された N グラム頻度に基づいて訂正を行う． Kao らは， moving window という，ターゲット語の周辺の単語列の頻度を用いて訂正を決定する手法を提案した [8] ．Lee らは， moving window の頻度だけでなく，確率も考慮した手法を用い，さらにルールベース法と組み合わせた [9] ．コロケーション誤りは単語の組み合わせに関わる誤りなので， N グラムを用いた言語モデルはそのような誤りを訂正するのに適している．しかし，この手法では周辺の単語列の N グラム頻度しか考慮していない．つまり，誤り語とその訂正語の関連性を考慮に入れていないので，文の意味を変えてしまうような訂正語を選んでしまう可能性があるという問題がある．

機械翻訳法は，「誤りを含んだ英語」を「正しい英語」へ翻訳するような機械翻訳システムを構築する手法で，広範な誤りを扱うことができる．2014 年の CoNLL での誤り種類別の再現率を見ると，機械翻訳法を用いたシステムはコロケーション誤りの訂正の性能が良いことが分かる [4].

Dahlmeier らは， NUCLE コーパスにおけるコロケーション誤りの大半は，誤り語と正しい語の間の類似性によるものだという指摘をし，そのような類似性を利用した機械翻訳を用いた手法を提案した [10]．前節のコロケーション誤りの例では，正しい語 pose と誤り語 cause の間には類似性が見られる．どちらも動詞原形であり，「発生させる」という同じような意味を持っている．Dahlmeier らは，対訳コーパスを用いて，まず誤り語を書き手の母語に翻訳し，さらにその語を英語に翻訳し直して訂正の候補語を得るという手法を提案し， NUCLE コーパスを用いた実験により，この手法の有効性を示した．この手法は，コロケーション誤りは，母語で似た意味を持つ英単語の混同によって起こるという考えに基づいている．しかし，この手法を広範な誤りに対して適用するには，大規模な対訳コーパスが必要となるという点で問題がある．本稿の手法では Dahlmeier らの手法のように書き手の母語特有の誤りを捉えることは難しいと考えられるが，英語のコーパスしか用いていないためより汎用的である．

3. VSM を用いたコロケーション誤りの訂正

3.1 タスクの設定

本稿では， Dahlmeier らが用いたタスク設定 [10] にしたがって，誤り箇所は 1 文につき 高々 1 つであるものとして，誤り箇所は前段階の処理ですでに検出済であると想定する．すなわち，誤りを含む対象文 s が $w_1, \cdots, w_K$ の K 語の並びで表されるとき，本タスクにおけるシステムへの入力および出力は以下となる．

入力 文 $s = (w_i)_{i=1}^{K}$，誤り語の位置 k $(1 \leq k \leq K)$

出力 訂正語 w'_k

ここで，コロケーション誤りには句動詞などフレーズの誤りも含まれる．例えば NUCLE コーパス [5] では，コロケーション誤りのうち約 49 % が一語置き換えの訂正で，残り 51% では，誤りまたは訂正が複数語である．しかしながら，現在一般的に用いられる word2vec [11] などの VSM 構築法は単語のみを対象とする場合が多いことから，本稿の実験では，一語置き換えの訂正のみを扱う．複合語を含む VSM 構成法の実装として， [12] なども提案されており，より一般的には文献 [13], [14] などによる意味合成法の適用も考えられる．これらを用いたフレーズを含むコロケーション誤り訂正への対応は今後の課題である．

図 1 に提案システムの概要を示す．提案システムは，候補語選択と並び替えの 2 つのステップからなる．候補語選択では，別途準備した大規模コーパスから得られる N グラム統計に基づき訂正の候補となる語を複数選択する．並び替えのステップでは，得られた候補語を VSM から得られた類似度に基づいて並び替える．以下，各ステップについて述べる．

3.2 候補語選択

候補語選択のステップでは，Kao らや Lee らの手法

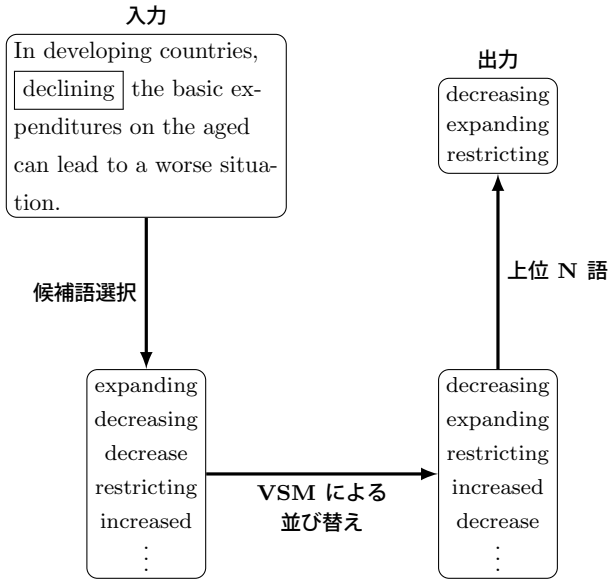

図 1 システムの概要図

表 1 s が “This situation may cause a serious challenge to the community.” のときの抽出されたフレーム集合 $\mathcal{F}_4(s,4)$

長さ	フレーム
2	may □ / □ a
3	situation may □ / may □ a / □ a serious
4	This situation may □ / situation may □ a / may □ a serious / □ a serious challenge

[8], [9] を参考に入力の文と誤り語から「フレーム」を抽出する．フレームとは，任意の単語と置換可能なワイルドカード（本稿では空白文字 □ で表す) をただ 1 つを含む N グラムである．Kao らや Lee らの手法は，訂正する前と訂正した後のフレームの頻度を比較することで，訂正を行うかどうかを決定するものであるが，本稿ではフレームを用いて訂正の候補となる語を選択するため，従来手法とはことなる方法を用いている．

フレーム F と単語 w に対して， F 中の □ を w で置き換えた N グラムを $F(w)$ で表す．そのような N グラムをフレーム N グラムと呼ぶことにする．いま，誤りを含む入力文と誤り語の位置の組 (s,k) に対して， 誤り語 w_k を □ で置き換えて得られるフレームを考える．(s,k) から抽出されるフレームのうち，長さ l 以下のものの集合を $\mathcal{F}_l(s,k)$ と表す．たとえば，表 1 は $l=4$ のときに前出の例文に対して抽出されるフレーム集合である．

ここでは，与えらえたコーパス中での N グラム出現頻度に基づき，フレーム N グラム $F(w)$ に対してスコア $S_{\mathrm{f}}(F(w))$ を以下のように割り当てる．$\mathrm{freq}(F(w))$ を $F(w)$ のコーパス内での頻度，W をコーパス中に出現するすべての語の集合として，$\mathrm{freq}(F(w))>0$ の場合， S_{f} を次式で定義する．

$$S_{\mathrm{f}}(F(w)) = -\log \frac{\mathrm{freq}(F(w)) - d}{\sum_{w' \in W} \mathrm{freq}(F(w'))} \qquad (1)$$

$\mathrm{freq}(F(w))=0$ のときは，$Z=\{w \in W \mid \mathrm{freq}(F(w))>0\}$ として，次式で定義する．

$$S_{\mathrm{f}}(F(w)) = -\log \frac{|Z|d}{\sum_{w' \in W} \mathrm{freq}(F(w'))} \frac{1}{|W|-|Z|} \qquad (2)$$

ここで，d は absolute discounting [15] における discounting 定数で，これによって，頻度がゼロの語にもスコアを割り当てるようにしている．

さらに，入力 (s,k) の誤り語 w_k に対する訂正語候補 w のスコア $S_{\mathrm{s}}(w;s,k)$ を，上記の $\mathcal{F}_l(s,k)$ および S_{f} を用いて，次のように定義する．

$$S_{\mathrm{s}}(w;s,k) = \sum_{F \in \mathcal{F}_l(s,k)} S_{\mathrm{f}}(F(w)) \qquad (3)$$

最後に，S_{s} が高い順から上位 m 語を候補語として選び，次の並び替えステップに受け渡す．

3.3 類似度による並び替え

並び替えのステップでは，単語のベクトル表現に基づき以下で計算される並び替えスコア S_{r} によって候補語集合を並び替える．

$$S_{\mathrm{r}}(w;s,k) = S_{\mathrm{s}}(w;s,k) \times (1 - \cos(\boldsymbol{v}_w, \boldsymbol{v}_{w_k})) \qquad (4)$$

ただし，$\boldsymbol{v}_w$ を単語 w のベクトル表現とする． このベクトル空間の構成法について以下に述べる．

ベクトル空間で，各次元が周辺語などの特定の文脈情報に対応しているとき，その空間を explicit 空間と呼び，それ以外のときは embedded 空間と呼ぶ．Mikolov らはニューラルネットワークを用いて embedded 空間を構築するモデルを提案した [11]．これに対して， Levy らは explicit 空間の有効性を示した [16]．Levy らは explicit 空間の構成法として，bag-of-words (BOW) と依存関係 (DEP) の 2 種類の文脈を用いて explicit 空間を構築し，前者が分野的類似度を測るのに適しているのに対して，後者は機能的類似度を捉えることができると指摘した [17].

本稿では，このような空間の性質の違いが及ぼす影響について検証するため，embedded 空間モデルの代表的な実装の一つである word2vec を用いて embedded 空間を構築する．また，Levy らの 2 種類の文脈を用いて，それぞれに対応する explicit 空間を構築する．BOW 文脈はコーパス中のターゲット語の前後の語を文脈とする．例えば，w_1, w_2, w_3, w_4, w_5 という単語列がコーパスにあるとき，w_3 に対する文脈は $w_1^{-2}, w_2^{-1}, w_4^{+1}, w_5^{+2}$ である．ここで，文脈にはターゲット語からの相対位置が情報として付加されている．一方，DEP 文脈は単語間の依存関係を文脈とする．図 2 は文に対する単語の依存関係の例である．単語間の依存関係がラベルとともに示されている．各単語の文脈は，その単語と依存関係がある単語とそのラベルの組として表される．この例からは，単語 challenge の文脈は a/*det*, serious/*amod*, pose/$dobj^{-1}$ の 3 つとなる．ここで

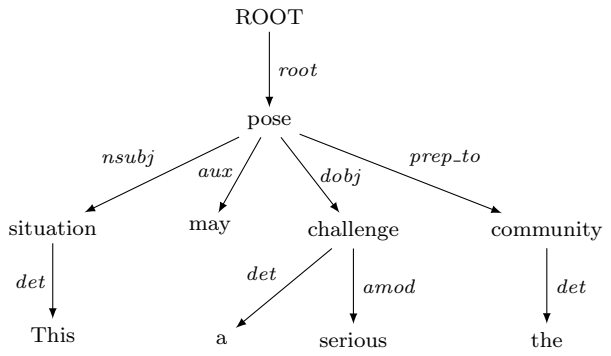

図 2 文 “This situation may pose a serious challenge to the community.” に対する依存関係

challenge に向かう依存関係の場合はラベル $dobj$ が $dobj^{-1}$ に変えたものを文脈としていることに注意されたい．

以下，word2vec，および BOW, DEP 文脈を用いて構成した 3 つの空間モデルをそれぞれ，W2V，BOW, DEP 空間と呼ぶ．

4. 実験

4.1 評価方法

評価に用いたデータセットは，NUCLE コーパス [5] から以下の方法で生成した．NUCLE コーパスではそれぞれの誤りに，誤り種類と訂正語が注釈として付加されている．そこで NUCLE コーパスから，一語置き換えで訂正されているコロケーション誤りを含む文を抽出し，さらに抽出した文から，コロケーション誤り 1 つだけが残るようにその他の誤りを注釈に沿って訂正した．これによって，最終的に 2,337 文からなる評価用集合を得た．

実験では，各文入力に対して候補語選択および並べ替え処理を適用し，得られる訂正候補語リストを評価した．自動評価では，NUCLE コーパス上で注釈の形で与えられている正解を基準訂正語とし，基準訂正語がシステム出力の上位 N 語の中に含まれるかどうかを調べ，2,337 文中での正答率を比較した ($N = 1, 5, 10, 50$)．あわせて平均逆順位 (MRR) も計算した． MRR は，最上位の正しい訂正候補語の順位の逆数の平均である．正しい訂正候補語がリストにない場合，順位の逆数を 0 とした．

ここで，コロケーション誤り訂正の正解は 1 通りとは限らない．そこで，自動評価の正当性を確認するため，実験ではさらに手動評価も行った．手動評価では，入力文集合から 100 文をランダムに選び，システム出力の上位 10 語について，人手で正しい訂正かどうかを判断して正答率を比較した．

4.2 モデル構築

候補語選択モデルと VSM を構築するためのコーパスは，英語版 Wikipedia から抽出し， splitta[*1] [18] で文分割し Stanford tokenizer でトークン化したものを用いた．最終的に約 18 億 6 千万トークン， 7460 万文のコーパスを得た．コーパスから， SRILM [19] を用いて長さ 5 以下の N グラムの頻度を数え，フレームスコアを計算した．その際，discounting 定数は 0.75 とした．

ベクトル空間も同じ Wikipedia コーパスから構築した．ただし， 4 語以下，または 41 語以上の長さの文はコーパスから除き，14 億 1 千万トークン， 6590 万文のコーパスを用いた．すべてのトークンを小文字化し，コンマとピリオド，英数字と隣り合うハイフン・アポストロフィ以外の記号は除いた．DEP 空間に対しては，各トークンは Stanford POS tagger [20] を用いて POS タグを付加し，Stanford parser [21] を用いて構文解析した．

W2V 空間に関しては，word2vec[*2] を用いて skip-gram モデルで 500 次元の空間を構築した．ネガティブサンプリング数は 15 とし，サブサンプリングのパラメータは 10^{-5} を，ウィンドウ幅は 5 を用いた．頻度が 100 よりも小さい語は除き，最終的な語彙の大きさは 174,367 となった．

Explicit 空間は， Levy らの手法 [16] に基づき，負の値を 0 にした自己相互情報量 (PPMI) を用いて構成した．また， W2V 空間と同じ語彙を用いた．BOW 文脈に対しては，ウィンドウ幅は 2 とし，DEP 文脈に対しては，頻度が 100 よりも小さい文脈は除いた．BOW 文脈集合の大きさは 697,463 ， DEP 文脈集合の大きさは 869,507 となった．

表 2 自動評価による正答率と MRR (%)．VSM が示されていないものは並び替えなしのモデル（ベースライン）．星印はベースラインとの有意差を表す （ McNemar 検定， $p < 0.05$）．

	正答率				
VSM	Top 1	Top 5	Top 10	Top 50	**MRR**
—	16.82	33.72	41.63	61.36	24.91
W2V	17.33	**38.85**⋆	**47.24**⋆	61.36	27.11
BOW	15.62	33.80	42.23	61.36	24.32
DEP	**17.80**⋆	35.94⋆	44.97⋆	61.36	26.43

4.3 結果

表 2 は自動評価による提案手法の正答率と MRR である．いずれの評価指標についても，VSM 類似度による並び替えで正答率が改善していることが分かる．ただし，BOW 空間を用いたときはベースラインからほとんど改善が見られない．DEP 空間は上位 1 語に対しては最も正答率が高いが，それ以外では W2V 空間が最も良い結果を残した．

表 3 は，並び替えによって基準訂正語の順位がベースラインから比べてどう変わったかの割合を表している．W2V 空間はもっとも順位が上昇したデータの割合が大きいが，DEP 空間は順位が下がったデータが一番少ない．このことから，DEP 空間は基準訂正語をベースラインよりも上位に並び替える割合が大きく，下位に並び替える割合が小

[*1] https://code.google.com/p/splitta/

[*2] https://code.google.com/p/word2vec/

表 3 並び替えによるベースラインからの順位の変化．数字はデータの中で順位が変動したものの割合 (%) を表す．

	W2V	BOW	DEP
順位上昇	24.95	19.98	23.83
順位降下	17.50	17.50	14.72
順位不変	18.91	23.88	22.81
上位 50 語外	38.64	38.64	38.64

表 4 ランダム 100 標本に対する自動・手動評価による正答率と MRR (%)

		正答率			
評価	VSM	Top 1	Top 5	Top 10	MRR
自動	—	21.00	39.00	43.00	28.17
	W2V	19.00	45.00	53.00*	30.23
手動	—	32.00	56.00	66.00	42.52
	W2V	31.00	65.00*	73.00*	45.81

さいが上昇幅はあまり大きくなく，一方で W2V 空間は下位に並び替えてしまう割合が大きいが，上位に並び替えたときの上昇幅は大きいと考えられる．

W2V 空間と DEP 空間によって順位が上昇したデータを見ると，W2V/DEP 空間両方で順位が上昇したものは 430 例， W2V 空間のみで上昇したものは 153 例， DEP 空間のみで上昇したものは 127 例となっている．このことから，これらの 2 つの空間はある程度性質を共有しながらも，捉えられる関係に差異があると考えられる．

4.4 人手による評価

表 4 はランダムに選んだ 100 標本に対する自動・手動評価による正答率と MRR である．手動で評価したとき，自動評価に比べて正答率が上がっていることが分かる．これは自動評価がシステムの性能を過小評価していることを示しているが，どちらの評価手法も一貫した傾向，すなわち，上位 1 語を除いては，並び替えによって正答率が上昇しているという傾向が見られる．このことから，自動評価の結果は，モデルを比較する際には有効であると考えられる．

4.5 システム出力の分析

表 5 と 6 はシステム出力の例である．太字の単語は，人手で正しいと判断された語を表している．

表 5 では，基準訂正語 *formulate* は候補語の中にないが，並び替えによって出力の中に適切な訂正語 *create*, *develop* そして *make* が現れている．3 つの空間の中では， W2V 空間を用いたときに最も高い順位に正しい訂正語が現れている．

表 6 では，基準訂正語 *bode* は並び替えによって元の順位よりも低く順位付けされている．これは，基準訂正語 *bode* と誤り語 *bore* の類似性の低さによると考えられる．しかし， W2V 空間を用いた並び替えでは，別の正しい訂正語である *augur* も上位に順位付けされている．

表 5 入力 “Governments and policy makers will take the main role to invent the best policies to deal with the long-term economic and social challenge of global aging, so that we are fully prepared face this issue.” に対するシステム出力の上位 5 語．基準訂正語は *formulate* ．

	並び替え	並び替えあり		
順位	なし	W2V	BOW	DEP
1	be	be	be	be
2	have	**create**	have	have
3	“	**develop**	**make**	**make**
4	ensure	use	play	play
5	play	have	ensure	become

表 6 入力 “In this economy-driven society, taking too much of the public’s resources to support the elderly does not bore well for the future of the country.” に対するシステム出力の上位 5 語．基準訂正語は *bode* ．

	並び替え	並び替えあり		
順位	なし	W2V	BOW	DEP
1	**bode**	go	go	go
2	go	**bode**	**bode**	**bode**
3	work	work	work	work
4	do	do	do	do
5	end	**augur**	end	end

5. 結論

本稿では，コロケーション誤りの訂正を決定するために， VSM を用いた並び替え手法を提案した．単語間の類似度を測るため， embedded と explicit の両方のベクトル空間を構築し，類似度による並び替えで正答率が改善されることを示した．この結果から，コロケーション誤り語とその訂正語の間には類似性があり，VSM を用いることでその類似度を測ることができると考えられる．さらに実験では， word2vec を用いた embedded 空間は explicit 空間よりも良い結果を残したが，語の依存関係を文脈として用いた explicit 空間では，順位を下げるような並び替えが embedded 空間よりも少ないなど，ベクトル空間の性質によって性能に差が現れることが分かった．並び替えが結果を劇的に改善する場合もある一方で，誤り語とその訂正語の間に類似性が見られない場合には，性能を低下させてしまうことがあることを見た．本論文の結果から，総括的な GEC システムにおいて， VSM の利用することで性能が改善する可能性があることが分かる．

謝辞 本研究は JSPS 科学研究費補助金 15H02754 の助成を受けたものです．

参考文献

[1] Dale, R. and Kilgarriff, A.: Helping Our Own: The HOO 2011 Pilot Shared Task, *Proceedings of the 13th European Workshop on Natural Language Generation*, ENLG ’11, Strouds-

burg, PA, USA, Association for Computational Linguistics, pp. 242–249 (online), available from ⟨http://dl.acm.org/citation.cfm?id=2187681.2187725⟩ (2011).

[2] Dale, R., Anisimoff, I. and Narroway, G.: HOO 2012: A Report on the Preposition and Determiner Error Correction Shared Task, *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 54–62 (online), available from ⟨http://dl.acm.org/citation.cfm?id=2390384.2390390⟩ (2012).

[3] Ng, H. T., Wu, S. M., Wu, Y., Hadiwinoto, C. and Tetreault, J.: The CoNLL-2013 Shared Task on Grammatical Error Correction, *Proceedings of the Seventeenth Conference on Computational Natural Language Learning: Shared Task*, Sofia, Bulgaria, Association for Computational Linguistics, pp. 1–12 (online), available from ⟨http://www.aclweb.org/anthology/W13-3601⟩ (2013).

[4] Ng, H. T., Wu, S. M., Briscoe, T., Hadiwinoto, C., Susanto, R. H. and Bryant, C.: The CoNLL-2014 Shared Task on Grammatical Error Correction, *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*, Baltimore, Maryland, USA, Association for Computational Linguistics, pp. 1–14 (2014).

[5] Dahlmeier, D., Ng, H. T. and Wu, S. M.: Building a Large Annotated Corpus of Learner English: The NUS Corpus of Learner English, *Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications*, Atlanta, Gergia, USA, pp. 22–31 (2013).

[6] Felice, M., Yuan, Z., E. Andersen, Ø., Yannakoudakis, H. and Kochmar, E.: Grammatical error correction using hybrid systems and type filtering, *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*, Baltimore, Maryland, USA, Association for Computational Linguistics, pp. 15–24 (online), available from ⟨http://aclweb.org/anthology/W14-1702⟩ (2014).

[7] Turney, P. D.: Domain and Function: A Dual-space Model of Semantic Relations and Compositions, *Journal of Artificial Intelligence Research*, Vol. 44, No. 1, pp. 533–585 (online), available from ⟨http://dl.acm.org/citation.cfm?id=2387933.2387945⟩ (2012).

[8] Kao, T.-h., Chang, Y.-w., Chiu, H.-w., Yen, T.-H., Boisson, J., Wu, J.-c. and Chang, J. S.: CoNLL-2013 Shared Task: Grammatical Error Correction NTHU System Description, *Proceedings of the Seventeenth Conference on Computational Natural Language Learning: Shared Task*, Sofia, Bulgaria, Association for Computational Linguistics, pp. 20–25 (online), available from ⟨http://www.aclweb.org/anthology/W13-3603⟩ (2013).

[9] Lee, K. and Lee, G. G.: POSTECH Grammatical Error Correction System in the CoNLL-2014 Shared Task, *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*, Baltimore, Maryland, USA, Association for Computational Linguistics, pp. 65–73 (online), available from ⟨http://www.aclweb.org/anthology/W/W14/W14-1709⟩ (2014).

[10] Dahlmeier, D. and Ng, H. T.: Correcting Semantic Collocation Errors with L1-induced Paraphrases, *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 107–117 (online), available from ⟨http://dl.acm.org/citation.cfm?id=2145432.2145445⟩ (2011).

[11] Mikolov, T., Chen, K., Corrado, G. and Dean, J.: Efficient Estimation of Word Representations in Vector Space, *CoRR*, Vol. abs/1301.3781 (online), available from ⟨http://arxiv.org/abs/1301.3781⟩ (2013).

[12] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. and Dean, J.: Distributed representations of words and phrases and their compositionality, *Advances in Neural Information Processing Systems*, pp. 3111–3119 (2013).

[13] Mitchell, J. and Lapata, M.: Vector-based models of semantic composition, *Proceedings of dACL-08: HLT*, pp. 236–244 (2008).

[14] Mitchell, J. and Lapata, M.: Composition in distributional models of semantics, *Cognitive science*, Vol. 34, No. 8, pp. 1388–1429 (2010).

[15] Ney, H., Essen, U. and Kneser, R.: On Structuring Probabilistic Dependencies in Stochastic Language Modelling, *Computer Speech and Language*, Vol. 8, pp. 1–38 (1994).

[16] Levy, O. and Goldberg, Y.: Linguistic Regularities in Sparse and Explicit Word Representations, *Proceedings of the Eighteenth Conference on Computational Natural Language Learning*, Ann Arbor, Michigan, pp. 171–180 (online), available from ⟨http://aclweb.org/anthology/W14-1618⟩ (2014).

[17] Levy, O. and Goldberg, Y.: Dependency-Based Word Embeddings, *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Vol. 2, Baltimore, Maryland, Association for Computational Linguistics, pp. 302–308 (online), available from ⟨http://aclweb.org/anthology/P14-2050⟩ (2014).

[18] Gillick, D.: Sentence Boundary Detection and the Problem with the U.S., *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers*, Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 241–244 (online), available from ⟨http://dl.acm.org/citation.cfm?id=1620853.1620920⟩ (2009).

[19] Stolcke, A.: SRILM – an extensible language modeling toolkit, *Proceedings of International Conference on Spoken Language Processing*, pp. 257–286 (2002).

[20] Toutanova, K., Klein, D., Manning, C. D. and Singer, Y.: Feature-rich Part-of-speech Tagging with a Cyclic Dependency Network, *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, NAACL '03, Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 173–180 (online), DOI: 10.3115/1073445.1073478 (2003).

[21] Klein, D. and Manning, C. D.: Accurate Unlexicalized Parsing, *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, ACL '03, Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 423–430 (online), DOI: 10.3115/1075096.1075150 (2003).