

文書分類の手法と一般化線形モデルを用いた 英語ライティングにおける文法的誤りの影響

石井 雄隆 (早稲田大学 大学総合研究センター)

近藤 悠介 (早稲田大学 グローバルエデュケーションセンター)

本研究では、文書分類の手法と一般化線形モデルを用いて英語ライティングにおける文法的誤りの影響について調査した。日本人英語学習者のスピーキングやライティングのパフォーマンスを自動採点する試みが近年注目を集めているが、どの特徴量がどれほど評価に寄与しているかについては、まだ十分に明らかにされていない。本研究では、パフォーマンスを測る一つの指標である文法的正確さに焦点を当てて、その影響を調査した。その結果、エッセイ評価に影響を与えている文法的誤りは、動詞の語彙に関するエラーと語順に関するエラーの二種類の文法的誤りであった。英語学習者のライティング評価においては、動詞に関する誤りは、他の文法的誤りに比べて、エッセイの全体的評価に大きな影響を与えていることを示唆する結果となった。

Investigating effects of accuracy in English writing based on the technique of essay classification and generalized linear model

Yutaka Ishii (Center for Higher Education Studies, Waseda University)

Yusuke Kondo (Global Education Center, Waseda University)

This study investigates effects of grammatical accuracy in English writing based on the technique of essay classification and generalized linear model. In recent years, there has been an increasing interest in automated scoring of learners' production skills such as speaking and writing. However, one major issue in the research concerned what linguistic features contribute to essay grading. This paper assesses the significance of grammatical accuracy in essay scoring. The results showed that verb related errors (errors in word selection) and errors in word order affect learners' essay evaluation.

1. はじめに

本研究では、文書分類の手法と一般化線形モデルを用いて英語ライティングにおける文法的誤りの影響について調査した。日本人英語学習者のスピーキングやライティングのパフォーマンスを自動採点する試みが近年注目を集めているが、どの特徴量がどれほど評価に寄与しているかについては、まだ十分に明らかにされていない。本研究では、パフォーマンスを測る一つの指標である文法的正確さに焦点を当てて、その影響を調査する。

2. 先行研究

これまでのライティング研究の多くは、第二言語としての英語ライティングを対象として行われてきており、日本人英語学習者を対象とした研究はあまり行われてこなかった。また日本における英語ライティング研究には大きく分けて二つの問題が存在すると言われてきた。一つは、ライティングの評価は、教師にとって、大きな負担で

あるということと、評定者間の評価の不一致の問題である。

上記の問題を解決するために、近年注目を集めているのがライティングの自動評価である。ライティングの自動評価においては、さまざまな言語的指標が用いられる。代表例として Educational Testing Service (ETS) の e-rater は、下記 12 個の変数を用いている。

1. 総語数に対する文法エラーの割合
2. 総語数に対する語の使用法についてのエラーの割合
3. 総語数に対する手順のエラーの割合
4. 総語数に対するスタイルについてのエラーの割合
5. 必要とされる談話要素の数
6. 談話要素における平均語数
7. 作文を 6 点法で採点する際に語彙の類似度が一番近い点数
8. 最高点を取った作文との語彙の類似度
9. Type-Token Ratio
10. 語彙の困難度

11. 平均単語長

12. 総語数

(Burstein, Chodorow & Leacock, 2004)

上記の変数の中で全体の3分の1を占めるのが文法的誤りである。上記を鑑みると、ライティングの評価は文法的誤りに大きく寄与すると考えられる。一方で、Jones (2006) は、ライティングにおける46箇所の文法的誤りを訂正し、それをある自動採点システムに評価させたところ、訂正前と訂正後のスコアが全く同じであったということを経験している。この研究は、学習者の文法的誤りがライティングの評価において寄与しない可能性を示唆する研究である。そこで本研究では、文法的誤りという変数のみでどれほどライティングの評価を予測することができるかを検討する。

本研究が、文法的誤りに焦点を当てる理由としては、二点が挙げられる。一点目は、自然言語処理の教育応用で近年盛んに文法的誤りの自動検出に関する共通タスクなどが行われているという点である。例えば、これまでに下記表1のような共通タスクが行われている。

表-1 自然言語処理の教育応用の共通タスク

会議名	タスク
Helping Our Own (HOO) 2011	論文の文法的誤り訂正
Helping Our Own (HOO) 2012	前置詞と限定詞の文法的誤り訂正
Native Language Identification Shared Task	英作文から英語学習者の母語推定
CoNLL 2013	限定詞, 前置詞, 数, 動詞の形, 一致, スペル, 句読点の文法的誤り訂正
誤り検出・訂正ワークショップ 2012	前置詞誤り・動詞(主語・動詞の一致)誤りの2つのトラックに加え、誤りの種類を限定しないパイロットトラック

上記のような研究における知見は、外国語教育研究の発展に大きく寄与するものであると考えられる。しかしながら、外国語教育研究者にとっても最も関心があるのは、どういった文法的誤りがライティングの評価に寄与するかという点である。その研究成果は、教室におけるライティング指導などにも生かすことが可能である。

二点目は、外国語教育への活用が進められているヨーロッパ言語共通参照枠における文法的正

確さの記述の精緻化である。下記表2は、文法的誤りに対する能力記述文であるが、その記述を見てみると、段階ごとの誤りの程度について漠然とした記述がなされているのみである。学習者が習熟度ごとにおかしやすい文法的誤りを特定することは、こうした能力記述文の精緻化にも応用することが可能である。

表-2 ヨーロッパ言語共通参照枠の文法的正確さの能力記述文

レベル	文法的正確さ
C2	(例えば、これから言うことを考えている時や、他人の反応をモニターしているような時といった)他のことに注意を払っている時でも、複雑な言葉について常に高い文法駆使力を維持している。
C1	常に高い文法的正確さを維持する。誤りは少なく、見つけることは難しい。
B2	高い文法駆使力がある。時には「言い間違い」や、文構造での偶然起こした誤りや些細な不備が見られる場合があるが、その数は少なく、後で見直せば訂正できるものが多い。 比較的高い文法駆使力が見られる。誤解につながるような間違いは犯さない。
B1	馴染みのある状況では、割合正確にコミュニケーションを行うことができる。多くの場合高いレベルでの駆使能力があるが、母語の影響が明らかである。誤りも見られるが、本人が述べようとしていることは明らかに分かる。 比較的予測可能な状況で、頻繁に使われる「繰り返し」やパターンのレパートリーを、割合正確に使うことができる。
A2	いくつかの単純な文法構造を正しく使うことができるが、依然として決まって犯す基本的な間違いがある — 例えば、時制を混同したり、性・数・格などの一致を忘れていたりする傾向がある。しかし、本人が何を言おうとしているのかはたいていの場合明らかである。
A1	学習済みのレパートリーについて、いくつかの限られた単純な文法構造や構文を使うことはできる。

日本人英語学習者のライティング評価と文法的誤りの関係性を検討したものとして Kitamura (2011) が存在する。Kitamura (2011) では、決定木分析を用いて、文法的誤りのエッセイ評価への影響を調査した。その結果、トピックによっては文法的誤りがエッセイ評価に寄与することが示された。しかしながら、対象とされた文法的

誤りは主語と動詞の一致 (He have been living there since June.), 動詞の形 (I can't skiing well, but...), 不完全な文 (Because people's interesting thing is not the same.) の三種類であったため, 本研究では, 20 個の文法的誤りを対象とし, 検討を試みる.

Tang and Liu (2005) では, 文書分類タスクで良く用いられる素性選択として, Information Gain, Chi-squared test, Odds ratio, Bi-Nominal Separation を挙げている. 本研究では, Bi-Nominal Separation を用いる.

3. データ

本研究で用いたデータは, Konan-JIEM learner corpus と呼ばれる言語資源である (Nagata, Whittaker, & Sheinman, 2011). このコーパスは, 甲南大学と教育測定研究所 (JIEM) が共同で収集し, アノテーションを行ったコーパスである. 日本人英語学習者の 170 エッセイから成り, また, 文法誤り情報と品詞/句情報が人手で付与されている. 研究目的のためであれば言語資源協会を通じて, 購入が可能である. 概要を表 3 に, エラータグの種類を表 4 に示す.

表-3 データの概要

エッセイの数	170
エッセイを書いた学生の数	10
総文数	2409
総単語数	19285
異なり語数	2054

成田 (2013) では, Konan-JIEM learner corpus の個々のエッセイに日本の大学で英語教育に従事している教員が評価を与えている. 成田 (2013) ではすべてのエッセイを 3 人の教員が評価して優れたものとそうでないものに分類している.

本研究は, 成田 (2013) の評価を採用し, 全体で 170 あるエッセイからランダムに優れたもの 50 とそうでないもの 50 を選び 100 のエッセイを分析の対象とした.

表-4 エラータグの種類

タグ	内容
n_num	名詞・単複エラー 加算, 不可算エラー
n_lxc	名詞・語彙選択エラー
n_o	名詞-その他のエラー
pn	代名詞に関するエラー
vAgr	動詞・人称・数の不一致
v_tns	動詞・時制エラー
v_lxc	動詞・語彙選択エラー

v_o	動詞-その他のエラー
mo	助動詞に関するエラー
aj	形容詞に関するエラー
av	副詞 (句) に関するエラー
prp	前置詞に関するエラー
at	冠詞に関するエラー
con	接続詞に関するエラー
rel	関係詞に関するエラー
itr	疑問詞に関するエラー
o_lxc	二語以上から成る成句での 語彙選択ミス
ord	語順エラー
uk	特定不能なエラー 構成上の致命的なミス
f	フラグメント (断片, 未完の文等)

4. 分析

まず, エッセイ中の誤りの Bi-Normal Separation (BNS: Forman, 2003) を算出した (表 5). BNS は, ある事象があるカテゴリに出現したかどうかだけでなく, その事象が何回出現したかという頻度を表した指標である. 文法的誤りはその種類だけでなく, その出現頻度がエッセイの評価に影響を与えると考えられるため, 本研究で指標として採用した. 本研究において BNS は, 優れたエッセイとそうでないエッセイにおいてそれぞれの誤りが出現している量の差を示すことになる. 出現頻度が極端に低い (10 以下) の指標を除くと, 優れたエッセイとそうでないものにおいて大きな値を示すものは v_lxc (動詞の語彙に関するエラー) と ord (語順に関するエラー) の 2 種類の文法的誤りであった.

表-5 エラーの頻度と Bi-Normal Separation 値

タグ	出現頻度		BNS
	Good	Poor	
n_num	52	60	0.18
n_lxc	15	8	0.78
n_o	41	63	0.54
pn	33	38	0.18
vAgr	67	68	0.02
v_tns	40	69	0.68
v_lxc	15	49	1.45
v_o	9	7	0.31
mo	22	40	0.74
aj	28	27	0.05
av	79	125	0.57
prp	122	169	0.41
at	17	18	0.07
con	5	5	0.00

rel	2	0	1.00
itr	6	2	1.35
o_lxc	9	8	0.15
ord	20	60	1.35
uk	2	4	0.86
f	0	4	1.00

次に、文法的誤りの頻度を予測変数、エッセイの評価を基準変数としてロジスティック回帰を行い、すべての文法的誤りを予測変数として始め、ひとつずつ変数を減らし、AICが改善しなくなるまで繰り返した。最終的にAICは125.6となり、残った変数は、v_lxc（動詞の語彙に関するエラー）とord（語順に関するエラー）の2種類の文法的誤りであった。以下にv_lxc（動詞の語彙に関するエラー）とord（語順に関するエラー）の頻度を評価別に示す。図1および2が示すように、評価がGoodのエッセイのほとんどにこれらの文法的誤りは含まれていない一方で、評価Poorのエッセイの約半数にこれらの誤りが含まれている。また、評価Goodのエッセイにこれら2つのエラーが3つ以上含まれていることはなかった。例外はあるが、この傾向は他の文法的誤りにおいても見られた。

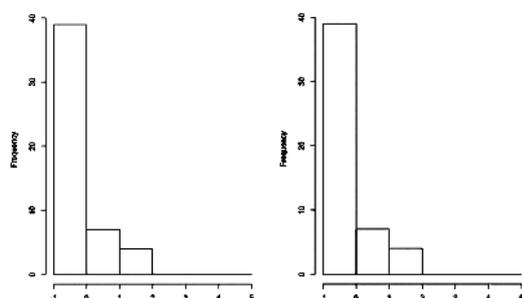


図-1 評価 Good の v_lxc (左) と ord (右) の頻度

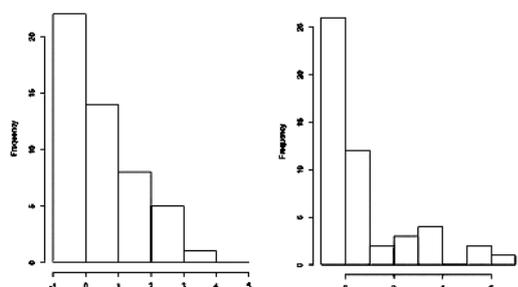


図-2 評価 Poor の v_lxc (左) と ord (右) の頻度

このようなデータの傾向から、文法的誤りの回数に評価の境界線が引けると考え、決定木分析を用

いて評価の予測を試みた。決定木分析による予測結果を表6に示す（適合率: 0.62, 再現率: 0.81, 特異度: 0.69, 正確度: 0.74, F値: 0.70）。図3に決定木分析の結果を示す。ロジスティック回帰分析の結果とは異なるが、この結果もエッセイ評価において動詞の語彙に関するエラーが重要な文法的誤りであることを示している。

表-6 決定木分析の分類精度

		真の結果	
		Good	Poor
予測結果	Good	31	19
	Poor	7	43

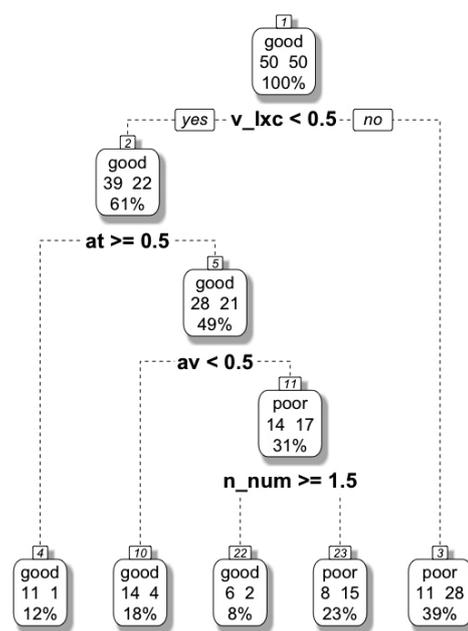


図-3 決定木分析の結果

実際に学習者の産出した英文の例を下記に示す。

Last, I want to earn much money and I want to <v_lxc crr="travel">trip</v_lxc> <ord crr="everywhere in Japan">in Japan everywhere</ord>.

上記のように、動詞 travel を使わなければならない箇所に名詞 trip が書かれている。こういった動詞を書かなければならない箇所に名詞を記してしまうタイプの誤りの産出は v_lxc（動詞の語彙に関するエラー）に多く、こうした誤りは作文の評価に寄与しやすいということを教師は教室で教える必要があると考えられる。

5. 考察

本研究では、BNS、ロジスティック回帰分析、決定木分析を用いて、英語学習者のエッセイに対する評価に与える文法的誤りの影響を調査した。それぞれの分析において異なる結果が得られた。共通する結果としては *v_lxc* (動詞の語彙に関するエラー) が優れたエッセイとそうでないものを分ける上で重要な文法的誤りであることが示された。

決定木分析では、訓練データのみでの結果であるが、4種類の文法的誤りのみを用いてかなり高い精度で評価を予測できることが示された。成田(2013)による評価は文法的誤りに焦点を当てた評価ではなく、エッセイ全体の評価である。この評価を文法的誤りのみを用いて予測ができるということは、エッセイ評価における文法的誤りの重要度が高いと解釈できる。

本研究結果は、英語学習者のライティング指導や自動評価において重要な示唆を含む。ライティング指導においては、動詞に関する知識がライティングの評価に大きく寄与することから、教師は、そういった知識については、重点的に指導をすることが求められる。また自動採点に関する研究開発においては、句構造などの統語解析や *n-gram* を用いて語順に関するエラーを検出し、他の誤りに比べて重みづけする必要があることが示唆された。

6) 成田真澄. Konan-JIEM 学習者コーパスにおける名詞後置修飾要素の分析. 東京国際大学論叢—言語コミュニケーション学部編.Vol.9,pp. 1-12. (2013)

7) Tang, L. & Liu, H. Bias analysis in text classification for highly skewed data. ICDM '05: Proceedings of the Fifth IEEE International Conference on Data Mining, pp. 781-784. (2005)

参考文献

- 1) Burstein, J., Chodorow, M., & Leacock, C. Automated essay evaluation: The Criterion online writing service. *AI Magazine*, 25(3), pp.27-36. (2004)
- 2) Forman, G. An extensive empirical study of feature selection metrics for text classification. *The Journal of machine learning research*, 3, pp.1289-1305. (2003)
- 3) Jones, E. ACCUPLACER's essay-scoring technology: When reliability does not equal validity. In P. F. Ericsson & R. Haswell (Eds.), *Machine scoring of student essays*. (pp. 93-113). Logan, UT: Utah State University Press. (2006)
- 4) Kitamura, M. Influence of Japanese EFL Learner Errors on Essay Evaluation. *Annual Review of English Language Education in Japan*, Vol.22, pp.169-184. (2011)
- 5) Nagata, R. Whittaker, E., & Sheinman, V. Creating a manually error-tagged and shallow-parsed performance learner corpus. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp.1210-1219. (2011)

資料-1 学習者のエッセイの例 (University Life)

My university life is very interesting. Because I `<v_lxc crr="do">act</v_lxc>` many things `<prp crr="">since</prp>` now. First I `<uk crr="am a member of">join</uk>` `<at crr="a"></at>` cercle. I feel `<ord crr="very good about this"><prp crr="about"></prp>` this very good`</ord>`. `<uk crr="I kill time by">My killing time is</uk>` writing `<n_num crr="novels">novel</n_num>` or drawing `<n_num crr="pictures">picture</n_num>`. `<uk>This has many people like me</uk>`. So I concentrate `<prp crr="on"></prp>` this. Second is summer vacation. I `<v_tns crr="did"><v_lxc crr="do">act</v_lxc></v_tns>` many `<n_num crr="things">thing</n_num>` in `<pn crr="my"></pn>` summer vacation. My best memory is `<at crr="the"></at>` seminar on the sea. I went to Ho-chi-min and Singapore. I got many friends `<prp crr="from">around</prp>` Hyogo university. And I `<av crr="sometimes">sometime</av>` meet `<pn crr="them">friends</pn>`. Last I have many friends `<prp crr="from">in</prp>` high school, junior high school and `<aj crr="other">etc</aj>` `<n_num crr="groups">group</n_num>`. We always talk about each `<n_o crr="other's">other</n_o>` `<n_num crr="lives">life</n_num>` `<prp crr="by">in</prp>` e-mail or internet. And We play `<prp crr="">in</prp>` inside or outside home. We play funny `<n_num crr="games">game</n_num>`. For example, `<n_lxc crr="one of us">a friend</n_lxc>` `<v_agr crr="calls">call</v_agr>` `<prp crr="">in</prp>` Macdonald `<con crr="and"></con>` `<v_lxc crr="says"></v_lxc>` "Please give me `<at crr="a"></at>` hundred `<n_num crr="hamburgers">hunbergar</n_num>`. " And others look `<prp crr="at"></prp>` him and laugh. I have many friends, so my university life is very interesting.

資料-2 エラータグの例

タグ	例文
n_num	This is the only one <code><n_num crr="thing"></code> things <code></n_num></code> you have to do.
n_lxc	She listened to his <code><n_lxc crr="speech">speak</n_lxc></code> .
n_o	I went to <code><n_o crr="Nihonbashi in Osaka">Osaka Nihonbashi</n_o></code> .
pn	I took Martin and a frien of <code><pn crr="his">him</pn></code> to the park.
v_agr	The number of students who work part-time after school <code><v_agr crr="has been increasing">have been increasing</v_agr></code> .
v_tns	I'll make researvations for the ferry as soon as I <code><v_tns crr="find">will find</v_tns></code> out the schedule.
v_lxc	He wanted to <code><v_lxc crr="conceal">cancel</v_lxc></code> his guilt.
v_o	If it <code><v_o crr="is forgotten"><v_agr crr="forgets">forget</v_agr></v_o></code> , plants are going to die.
mo	"The phone is ringing." "I <code><mo crr="will"></code> 'm going to <code></mo></code> answer it."
aj	It was a <code><aj crr="genuine">genius</aj></code> diamond.
av	He worked <code><av crr="hard">hardly</av></code> today.
prp	He took full advantage <code><prp crr="of">with</prp></code> his position.
at	She is active in <code><at crr="the">a</at></code> development of low cost water pumps.
con	Clint hit a home run, <code><con crr="but">and</con></code> I didn't.
rel	I phoned all his friends, none of <code><rel crr="whom">who</rel></code> could tell me where he was.
itr	<code><itr crr="Which">What</itr></code> would you like to eat, Japanese or Chinese food?
o_lxc	He <code><o_lxc crr="made an attempt">had an attempt</o_lxc></code> at the conquest of the peak.
ord	When did you buy that <code><ord crr="large old brown wooden">old brown large wooden</ord></code> table?
uk	<code><uk crr="X">...</uk></code> In case of the UK tag, correction (crr="X") is not annotated in the tag.
f	<code><f>The last day,</f></code>