

# ライティング・タスクにおける課題文の影響 —多変量アプローチによる言語分析—

小林雄一郎<sup>†1</sup> 阿部真理子<sup>†2</sup>

本研究の目的は、多変量アプローチを用いて、ライティング・タスクにおけるトピックと課題文の影響を調査することである。分析データは、International Corpus Network of Asian Learners of English (ICNALE) を用いる。そして、英語学習者のライティングがトピックと課題文の影響を強く受けていることを明らかにする。

## A Multivariate Approach to the Effects of Writing Task Prompts

YUICHIRO KOBAYASHI<sup>†1</sup> MARIKO ABE<sup>†2</sup>

The purpose of the present study is to investigate the effects of essay prompts and topics on the language use in the EFL writings using a multivariate approach. It draws on the International Corpus Network of Asian Learners of English (ICNALE). By so doing, we illustrate the essay writings are greatly influenced by the topics and prompts.

### 1. はじめに

近年、学習者のライティングやスピーキングを大量に集めた学習者コーパスの構築が盛んに行われている (Pravec 2002)。また、複数の学習者コーパスを比較した研究なども多く見られる (e.g. Granger 2013, Tono 2007)。しかしながら、コーパスは特定の目的のための明示的なデザイン基準に沿って構築されたものであり (Atkins & Clear 1992)、コーパスに基づく言語分析の結果は「コーパス次第」であるとも言われている (Sinclair 1991)。従って、異なる学習者コーパスのデータを比較する場合には、細心の注意が必要である。

Granger (1998) によれば、学習者コーパスのデザイン基準には、全てのサブコーパスで統一して採用される共通特徴 (年齢、学習状況、習熟度、媒体、ジャンル、専門性) と、サブコーパスごとに異なる変数特徴 (性別、母語、地域、他の外国語、学習歴、トピック、タスク) の2種類が存在する。そして、当然のことながら、これらの変数特徴の違いが、サブコーパスにおける言語使用に影響を与える可能性がある (Brook & Hirst 2013)。

変数特徴の中でも、ライティングのトピックと課題文の影響に関しては、英語教育の分野で長らく議論されてきた (e.g. Hamp-Lyons 1990)。例えば、Brossell & Hoetker Ash (1984), Carlson *et al.* (1985), Spaan (1993) などは、ライティングのトピックが異なっていたとしても、ライティングの評価に有意な影響を与えないと述べている。その一方で、Reid (1990) や Hinkel (2009) は、トピックがライティングにおける言語使用に影響を与えると報告している。つまり、学習者の「言語使用」を分析する学

習者コーパス研究にとって、ライティングのトピックやタイプ、そして課題文の影響を明らかにすることは急務であると言えよう。特に、課題文は「学習者に反応させるための刺激」(Kroll & Reid 1994) であり、学習者のパフォーマンスに大きな影響を与える。

### 2. 研究の目的

本研究の目的は、多変量アプローチを用いて、ライティング・タスクにおける課題文の影響を明らかにすることである。分析にあたっては、語彙、品詞、統語、談話などの様々な観点から、異なるトピックのライティングにおける言語使用を比較する。また、どのような言語項目にトピックの差が見られるのかという点についての分析も行う。

### 3. データと方法論

#### 3.1. 分析データ

本研究で分析対象とするデータは、ICNALE (International Corpus Network of Asian Learners of English) (Ishikawa 2011) に収録されている香港人学習者 (HKG)、韓国人学習者 (KOR)、台湾人学習者 (TWN)、日本人学習者 (JPN) が書いたライティング (2,000 人分) である。ライティングのトピックは、「大学生のアルバイトへの賛否」(PTJ) と「レストラン全面禁煙への賛否」(SMK) の2種類で、課題文は以下のとおりである。

Do you agree or disagree with the following statement?  
Use reasons and specific details to support your answer.

(A) *It is important for college students to have a part-time job.*

(B) *Smoking should be completely banned at all the*

<sup>†1</sup> 日本学術振興会  
<sup>†2</sup> 中央大学理工学部

*restaurants in Japan.*

表1は、分析データの概要である。このデータにおける2つのトピックの分量は、同じ数に揃えられている。

表1 分析データの概要 (人数)

	HKG	KOR	TWN	JPN
PTJ	100	300	200	400
SMK	100	300	200	400
Total	200	600	400	800

そして、制限時間は20~40分、辞書や参考書の使用は不可、語数は200~300語、Wordのスペルチェックの強制使用が求められている。なお、このライティング課題は、大学の英語授業の課外課題として実施され、本文のみをメール本体に貼り付けて送信するという形式で提出されたものである。

### 3.2. 分析対象とする言語項目

先行研究では、ライティングのトピックや課題文が言語使用に与える影響を調べる場合、語数、平均文長、代名詞の種類、内容語の割合などが分析対象とされてきた(e.g. Reid 1990, Spaan 1993)。しかし、少数の言語項目の分析からは非常に限定的な結論しか得られないため、より多くの言語項目を用いて、テキスト解析を行う必要がある(Biber 1986)。

そこで本研究では、Biber (1988) が英語母語話者の話し言葉と書き言葉の分析に用いた67種類の言語項目のうち、以下の58項目を分析対象とする(詳しくは、本論文のAppendixを参照)。このような語彙、品詞、統語、談話などに関わる様々な言語項目を分析することで、学習者のライティングを包括的かつ多角的に比較することが可能になる。

### 3.3. 分析の手順

まず、ICNALEにおける2種類の課題文と、2,000人分のライティングから前掲の58項目の頻度を求める(前処理)。

次に、2,000人分のライティングにおける58種類の言語項目の頻度を対象として、2つのトピックにおける平均値を比較する(予備的実験1)。比較に用いる統計量は、 $t$ 値と、その効果量 $d$ と $r$ (Cohen 1988)である。

また、決定木(Breiman *et al.* 1984)を用いて、2つのトピックを弁別する言語項目を特定し、それらが互いにどのように関連し合っているかを確認する(予備的実験2)。決定木とは、非線形判別分析の1つであり、特徴量の値を何らかの基準で分岐させ、分類モデルを構築する手法である。その分岐の過程は、木構造で図示することがで

き、IF-THENのような簡単なルールで表すこともできる。

そして、58種類の言語項目の頻度を特徴量の候補とするランダムフォレスト(Breiman 2001)を用いて、トピックの弁別に有効な言語項目、個々のライティングの類似度、トピックと言語項目の頻度の関係などを明らかにする(本実験)。ランダムフォレストでは、まず、与えられたデータセットから、 $n$ 組のブートストラップサンプルを作成する。次に、各々のブートストラップサンプルデータを用いて、未剪定の最大の決定木を生成する(但し、分岐のノードは、ランダムサンプリングされた特徴量のうち最善のものを使用する)。そして、全ての結果を多数決で統合し、新しい分類器を構築する。

## 4. 結果と考察

### 4.1. 言語項目の頻度の算出

実験の前処理として、課題文、そして個々のライティングに現れている言語項目の頻度(100語あたりの相対頻度)を算出した。表2は、2種類の課題文に現れていた言語項目のリストである。

表2 課題文で用いられている言語項目のリスト

PTJ	SMK
attributive adjectives	agentless passives
be as main verb	amplifiers
infinitives	necessity modals
predicative adjectives	present tense
present tense	total other nouns
pronoun <i>it</i>	total prepositional phrases
total other nouns	
total prepositional phrases	

(それぞれの言語項目の頻度は省略している)

### 4.2. 言語項目の平均値の比較

まず、予備的実験として、2つのトピックにおける58種類の言語項目の生起頻度を比較した。比較に用いる統計量は、 $t$ 値である。また、サンプルサイズが大きいため、効果量 $d$ と $r$ も算出した。図1は、その結果である。この図において、それぞれの統計量が正の値を持つ言語項目はPTJに高頻度であり、負の値を持つ言語項目はSMKに高頻度であることを示している(因みに、58項目全てに関して、0.1%水準の有意差が見られた)。具体的には、PTJではinfinitives, attributive adjectives, third person pronouns, private verbs, first person pronounsなどが高頻度であり、SMKではagentless passives, necessity modals, amplifiers, analytic negation, place adverbialsなどが高頻度であった。

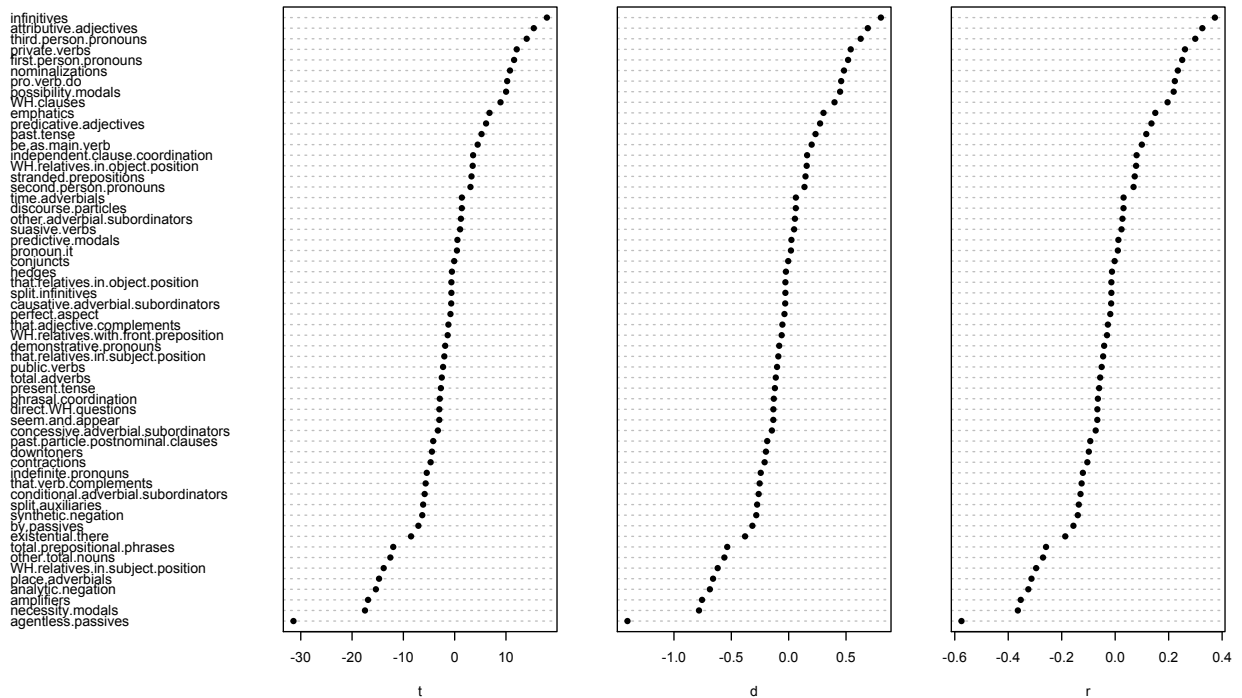


図1 2つのトピックにおける言語項目の頻度の比較

### 4.3. 決定木

次に、決定木を用いて、2つのトピックの弁別に有効な言語項目を特定し、それらの言語項目の関係を視覚的に表現する。図2は、CART (classification and regression tree) のアルゴリズムに基づく決定木である。

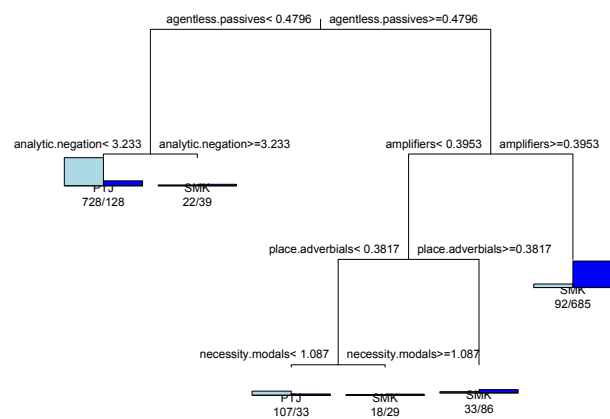


図2 トピックの弁別に有効な言語項目の関係

図2を見ると、最初に agentless passives の頻度によって、2つのトピックで書かれたライティングが分類されている。そして、agentless passives の値が一定未満の場合は左側へ、一定以上の場合は右側へ分岐する。次に、左側に分岐した場合は、analytic negation の値に基づいて、

PTJ か SMK かの判断が行われる。一方、最初のノードで右側に分岐した場合は、amplifiers の値を参照する。以下、必要に応じて、place adverbials や necessity modals の値を参照し、PTJ か SMK かの判断を下す。

また、図1と図2を見比べると、この決定木モデルで選択された言語項目は、5つとも SMK に高頻度なものであることが分かる。そのうち、agentless passives, amplifiers, necessity modals の3つは、SMK の課題文に含まれている言語項目である (表2を参照)。

### 4.4. ランダムフォレスト

これまでの予備的実験の結果から、分析データにおける2つのトピックの間には、言語使用に関する一定の差が存在することが確認された。そこで、決定木 (CART) よりも頑健で高精度な樹木モデルであるランダムフォレストを用いて、トピック間の言語使用の差を詳しく見ていく。

分析データにランダムフォレストを実行するにあたって、ランダムサンプリングする特徴量の数は全特徴量数の正の平方根を取り、生成する決定木の数は1,000本とする。図3は、生成する木の数と誤判別率の関係を表したものである。横軸が木の数で、縦軸が誤判別率をそれぞれ表している。この図を見ると、木の数が400本に達するあたりで誤判別率が比較的安定しているため、本実験における1,000本という木の数が妥当であることが分かる。

項目を対象とする部分従属プロットを作成する (図 5)。また、箱ひげ図を用いて、2 つのトピックにおける重要度上位 10 項目の頻度分布を視覚化する (図 6)。

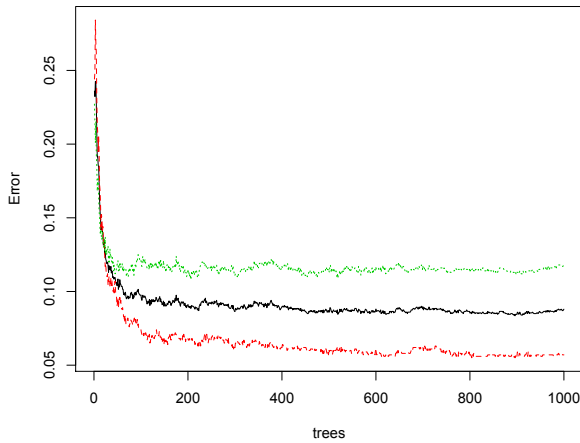


図 3 木の数と誤判別率の関係

本実験で構築されたランダムフォレストのモデルを見よう。図 4 は、2 つのトピックで書かれたライティングの弁別における特徴量の重要度をプロットしたものである。なお、重要度は、ジニ係数の平均減少数 (MeanDecreaseGini) に基づいて算出されている。このモデルにおける重要度上位 10 項目は、agentless passives, necessity modals, amplifiers, third person pronouns, infinitives, attributive adjectives, analytic negation, place adverbials, first person pronouns, other total nouns である。

図 4 の重要度を見るだけでは、それぞれの言語項目とトピックの関係が分からない。そこで、重要度上位 10

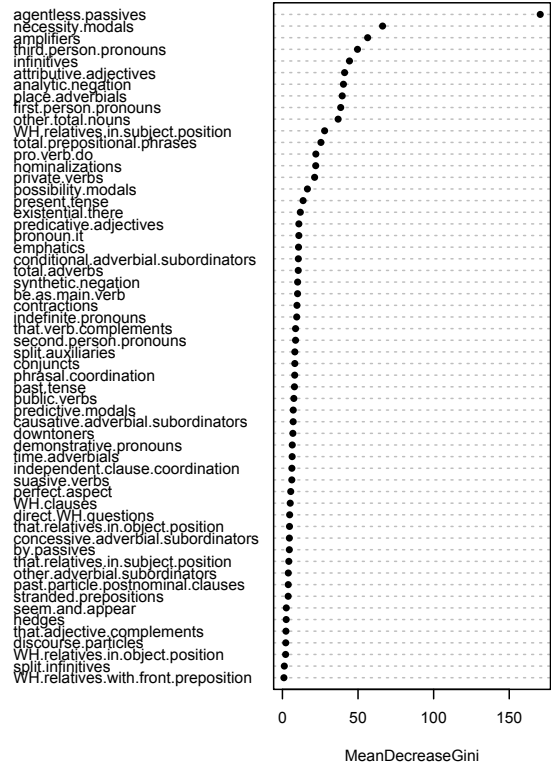


図 4 トピックの弁別における言語項目の重要度

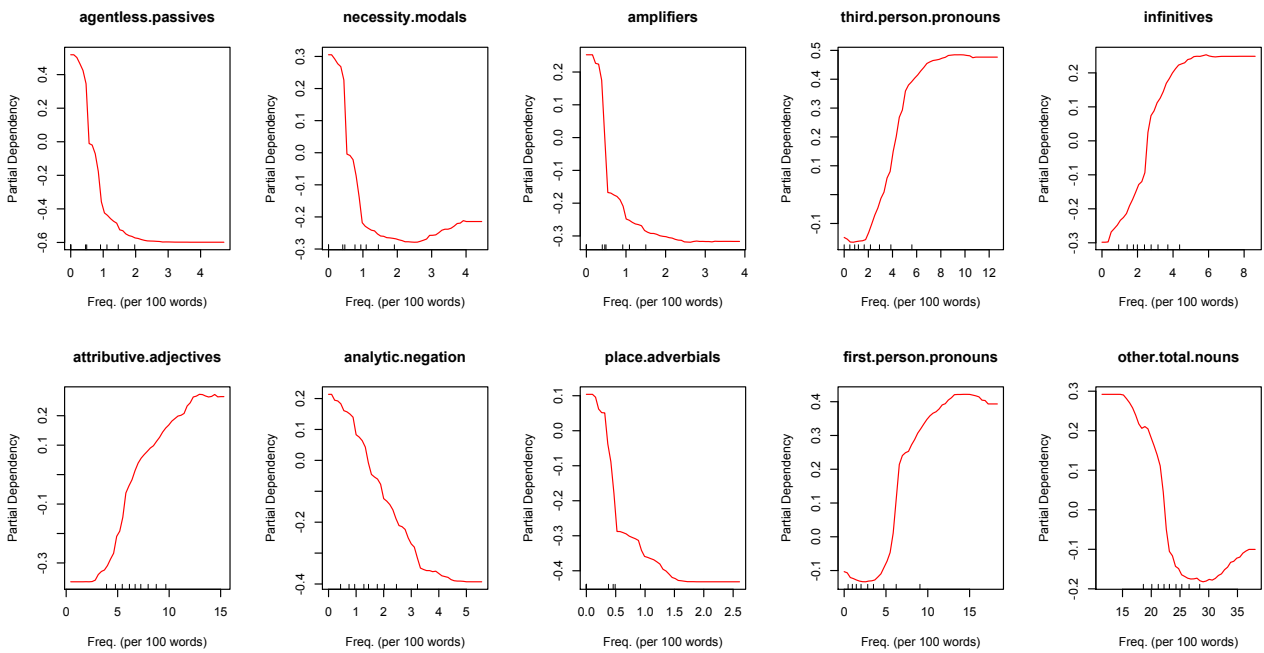


図 5 それぞれの言語項目とトピックの関係 (重要度上位 10 項目)

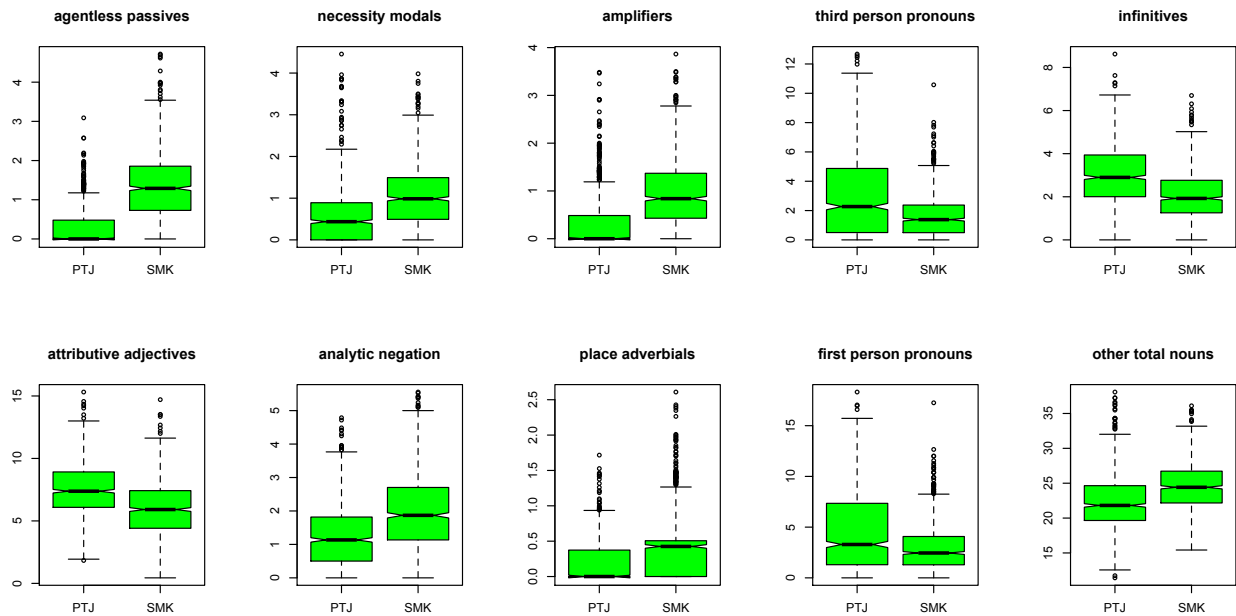


図6 それぞれのトピックにおける言語項目の頻度分布（重要度上位10項目）

図5では、縦軸の上方向に近づくほどライティングがPTJである確率が増し、下方向に近づくほどSMKである確率が増す。従って、図中の線が右肩上がりの言語項目はPTJで高頻度な項目であり、右肩下がりの言語項目はSMKで高頻度な項目である。このことは、図6の箱ひげ図からも確認することができる。

重要度上位10項目のうち、third person pronouns, infinitives, attributive adjectives, first person pronounsの4項目はPTJに高頻度であり、infinitivesとattributive adjectivesはPTJの課題文で用いられている。また、third person pronounsとfirst person pronounsに関しては、*we can* (e.g. *we can learn how to, we can make friends, we can realize the importance of money*) や *I think* (e.g. *I think that it is important for, I think part-time jobs are, I learned a lot if things*) のような表現が多く見られた。

一方、agentless passives, necessity modals, amplifiers, analytic negation, place adverbials, other total nounsの6項目はSMKに高頻度であり、agentless passives, necessity modals, amplifiers, other total nounsはSMKの課題文で用いられている（なお、other total nounsは、両方の課題文で用いられているが、SMKのライティングに高頻度で現れている）。また、analytic negationやplace adverbialsに関しては、*people who do not smoke, not allowed to smoke, inside buildings, inside the restaurant* のような表現が多く見られた。

PTJとSMKのいずれにおいても、高頻度で現れている言語項目は、課題文で用いられている項目やトピックと密接に関わるような項目である。まず、重要度が最も高

い受動態 (agentless passives) の頻度などは、テキストのリーダビリティの算出にも用いられるものである (e.g. Stockmeyer 2009)。このことは、トピックの違いがライティングの評価に影響を及ぼす可能性を示唆している。

コーパス言語学や計量文献学の分野では、テキストのトピックの影響を受けないように品詞 *n*-gram を用いて、構文や談話の分析を行うことがある (e.g. Aarts & Granger 1998, Wiersma *et al.* 2011)。しかしながら、重要度上位10項目には、受動態のみならず、副詞 (amplifiers, place adverbials), 助動詞 (necessity modals), 人称代名詞 (third person pronouns, first person pronouns), 不定詞 (infinitives), 形容詞 (attributive adjectives), 否定表現 (analytic negation), 名詞 (other total nouns) が含まれている。このことは、課題文やトピックの影響がライティングで用いられている品詞や構文にも及ぶことを表している。本研究で用いている ICNALE の構築にあたっては、「語彙分布におけるトピックの影響が大まかに統制されている」(Ishikawa 2009) と言われているが、トピックの影響を色濃く受けるのは語彙分布だけにとどまらない。

では、あるトピックで書かれたライティングと異なるトピックで書かれたライティングの間には、全体として、どの程度の言語使用の差が存在するのであろうか。ランダムフォレストでは、近接性尺度に基づく多次元尺度法を用いることで、個体間の関係を視覚化することが可能である。図7は、58種類の言語項目の使用パターンに基づいて、2,000本のライティングを分類した結果である。

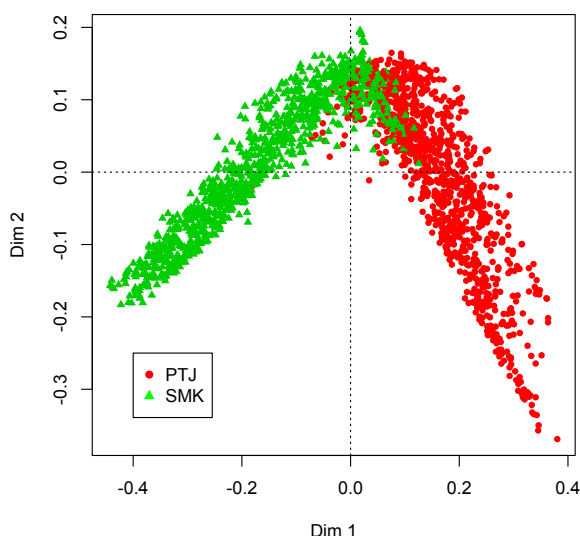


図7 個々のライティングの分類結果（教師なし分類）

図7では、PTJのライティングを●で表し、SMKのライティングを▲で表している。この図を見ると、異なるトピックで書かれたライティングでは、言語使用が大きく異なることがよく分かる。

そして、表3は、out-of-bag (OOB) による誤判別率の推定に基づく個体の分類結果である。

表3 個々のライティングの分類結果（教師あり分類）

	PTJ	SMK	class error
PTJ	943	57	0.057
SMK	119	881	0.119

表3の混同行列を見ると、2,000本のライティングのうち、1,824本が正しく分類されている（正判別率は、91.2%）。この結果は、多次元尺度法の結果と同様に、異なる課題文に基づき、異なるトピックで書かれたライティングにおいては、言語使用の大きな差異が存在するというを示している。

## 5. おわりに

本研究では、異なるトピックで書かれたライティングを対象として、語彙、品詞、統語、談話などに関する言語使用の差を調査してきた。そして、*t* 検定、決定木、ランダムフォレストなどの結果から、異なるトピックで書かれたライティングでは、言語使用が大きく異なることが明らかにされた。以下、上記の分析結果を踏まえて、学習者コーパスの構築と分析に関する示唆を述べる。

まず、学習者コーパスの構築にとって、トピックの影響を統制することは大きな課題である。母語話者のライ

ティングと比較して、学習者のライティングでは、課題文にある表現がそのまま使われることもある。従って、課題文の設定にあたっては、文の長さや使用語彙の難しさだけでなく、文法項目や談話表現についても細心の注意を払うべきである。言語研究では、特定の言語項目のみを分析することも多く、分析対象とされる言語項目と課題文で用いられている言語項目が相関関係を持っている場合などは、分析結果が歪められている可能性が高い。Ishikawa (2011) が述べているように、多様なトピックで書かれたライティングを学習者コーパスに含めると、個々のライティングやサブコーパスの比較が難しくなる。しかしながら、少数のトピックに限定するのであれば、何故それらのトピックに限定したのか、そして、それらのトピックに限定したことが学習者の言語使用にどのような影響を与えるのか、などの点に配慮する必要がある。学習者に馴染みのあるトピックを用いた方がライティングにおける流暢さを確保しやすいという指摘 (e.g. Miyazaki 2007) もあるが、多様な言語研究のための資料という観点に限って言えば、そのような理由だけでトピックを選択することは危険である（無論、教育現場で特定の言語項目を指導、もしくは評価するためのライティング課題であれば、その限りではない）。

次に、学習者コーパスの分析を行うにあたっては、トピックや課題文が言語使用に与える影響を事前に調査する必要がある。例えば、本研究で行われているように、できるだけ多様な言語項目を対象とするトピック間の比較を行うことが有効である。また、*n*-gram などを用いて、課題文で用いられている表現とライティングで用いられている表現の重複率を求めることも考えられる。その上で、分析対象とするコーパスのデザインや言語項目の数によっては、混合効果モデルなどを用いたマルチレベルの分析を行うことも検討すべきである。

## 謝辞

本研究の成果の一部は、科学研究費補助金（基盤研究B）「第二言語ライティング研究の現代的課題と解決のための将来構想—東アジアからの発信」（代表：大井恭子）（2012-2015年度）、科学研究費補助金（特別研究員奨励費（PD 実験））「パターン認識と自然言語処理の技術を用いた習熟度判定」（代表：小林雄一郎）（2012-2014年度）によるものである。

## 参考文献

- Aarts, J., & Granger, S. (1998). Tag sequences in learner corpora: A key to interlanguage grammar and discourse. In S. Granger (Ed.), *Learner English on computer* (pp. 132-141). London: Longman.

- Abe, M., Kobayashi, Y., & Narita, M. (2013). Using multivariate statistical techniques to analyze the writing of East Asian learners of English. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world – Vol.1* (pp. 55-65). Kobe: School of Language and Communication, Kobe University.
- Atkins, S., & Clear, J. (1992). Corpus design criteria. *Literary and Linguistic Computing*, 7(1), 1-16.
- Biber, D. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language*, 62, 384-414.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-23.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. New York: Chapman and Hall.
- Brook, J., & Hirst, G. (2013). Native language detection with 'cheap' learner corpora. In S. Granger, G. Guilquin & F. Meunier (Eds.), *Twenty years of learner corpus research: Looking back, moving ahead* (pp. 37-47). Louvain-la-Neuve: Presses universitaires de Louvain.
- Brossell, G., & Hoetker Ash, B. (1984). An experiment with the wording of essay topics. *College Composition and Communication*, 35(4), 423-425.
- Carlson, S., Bridgeman, B., Camp, R., & Waanders, J. (1985). Relationship of admission test scores to writing performance of native and nonnative speakers of English. *TOEFL Research Reports*, 19. Princeton: Educational Testing Service.
- Cohen, J. (1988). *Statistical power analysis for the behavioral science*. 2nd ed. Hillsdale: Lawrence Erlbaum.
- Granger, S. (1998). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on computer* (pp. 3-18). London: Longman.
- Granger, S. (2013). The passives in learner English: Corpus insights and implications for pedagogical grammar. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world – Vol.1* (pp. 5-15). Kobe: School of Language and Communication, Kobe University.
- Hamp-Lyons, L. (1990). Second language writing: Assessment issues. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 69-87). New York: Cambridge University Press.
- Hinkel, E. (2009). The effects of essay topics on modal verb uses in L1 and L2 academic writing. *Journal of Pragmatics*, 41, 667-683.
- Ishikawa, S. (2009). Phraseology overused and underused by Japanese learners of English: A contrastive interlanguage analysis. In K. Yagi & T. Kanzaki (Eds.), *Phraseology, corpus linguistics and lexicography: Papers from Phraseology 2009 in Japan* (pp. 87-98). Hyogo: Kwansai Gakuin University Press.
- Ishikawa, S. (2011). A new horizon in learner corpus studies: The aim of the ICNALE project. In G. Weir, S. Ishikawa, & K. Poonpon (Eds.), *Corpora and language technologies in teaching, learning, and research* (pp. 3-11). Glasgow: University of Strathclyde Press.
- Kroll, B., & Reid, J. (1994). Guidelines for designing writing prompts: Clarifications, caveats and cautions. *Journal of Second Language Writing*, 3(3), 231-255.
- Miyazaki, K. (2007). Exploring the influence of prompts and topics on EFL essay writing. *The Bulletin of the Professional License Training Center, Tokai University*, 6, 13-25.
- Pravec, N. A. (2002). Survey of learner corpora. *ICAME Journal*, 26, 81-114.
- Reid, J. (1990). Responding to different topic types: A quantitative analysis from a contrastive rhetoric perspective. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 191-210). New York: Cambridge University Press.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Spain, M. (1993). The effect of prompt in essay examinations. In D. Douglas & C. Chapelle (Eds.), *A new decade of language testing research* (pp. 98-122). Alexandria: TESOL.
- Stockmeyer, N. O. (2009). Using Microsoft Word's readability program. *Michigan Bar Journal*, 88(1), 46-47.
- Tono, Y. (2007). The roles of oral L2 learner corpora in language teaching: The case of the NICT JLE Corpus. In M. C. Compoy & M. J. Luzón (Eds.), *Spoken corpora in applied linguistics* (pp. 163-179). Bern: Peter Lang.
- Wiersma, W., Nerbonne, J., & Lattamus, T. (2011). Automatically extracting typical syntactic differences from corpora. *Literary and Linguistic Computing*, 26(1), 107-124.

## Appendix: 分析対象とする言語項目

### A. Tense and aspect markers

1. past tense
2. perfect aspect
3. present tense

## **B. Place and time adverbials**

4. place adverbials (e.g. *across, behind, inside*)
5. time adverbials (e.g. *early, recently, soon*)

## **C. Pronouns and pro-verbs**

6. first person pronouns
7. second person pronouns
8. third person pronouns (excluding *it*)
9. pronoun *it*
10. demonstrative pronouns (*that, this, these, those*)
11. indefinite pronouns (e.g. *anybody, nothing, someone*)
12. pro-verb *do* (e.g. *the cat did it*)

## **D. Questions**

13. direct WH-questions

## **E. Nominal forms**

14. nominalizations (ending in *-tion, -ment, -ness, -ity*)
15. total other nouns (except for nominalizations)

## **F. Passives**

16. agentless passives
17. *by*-passives

## **G. Stative forms**

18. *be* as main verb
19. existential *there* (e.g. *there are several explanations...*)

## **H. Subordination**

### **H1. Complementation**

20. *that* verb complements (e.g. *I said that he went*)
21. *that* adjective complements (e.g. *I'm glad that you like it*)
22. WH-clauses (e.g. *I believed what he told me*)
23. infinitives (*to*-clause)

### **H2. Participial forms**

24. past participial postnominal (reduced relative) clauses (e.g. *the solution produced by this process*)

### **H3. Relatives**

25. *that* relatives in subject position (e.g. *the dog that bit me*)
26. *that* relatives in object position (e.g. *the dog that I saw*)
27. WH relatives in subject position (e.g. *the man who likes popcorn*)
28. WH relatives in object position (e.g. *the man who Sally likes*)
29. WH relatives with fronted preposition (e.g. *the manner in which he was told*)

### **H4. Adverbial clauses**

30. causative adverbial subordinators: *because*
31. concessive adverbial subordinators: *although, though*
32. conditional adverbial subordinators: *if, unless*
33. other adverbial subordinators: having multiple functions (e.g. *since, while, whereas*)

## **I. Prepositional phrases, adjectives, and adverbs**

34. total prepositional phrases
35. attributive adjectives (e.g. *the big horse*)
36. predicative adjectives (e.g. *the horse is big*)
37. total adverbs (except conjuncts, hedges, emphatics, discourse particles, downtoners, amplifiers)

## **J. Lexical classes**

38. conjuncts (e.g. *consequently, furthermore, however*)
39. downtoners (e.g. *barely, nearly, slightly*)
40. hedges (e.g. *at about, something like, almost*)
41. amplifiers (e.g. *absolutely, extremely, perfectly*)
42. emphatics (e.g. *a lot, for sure, really*)

43. discourse particles (e.g. sentence initial *well, now, anyway*)

## **K. Modals**

44. possibility modals (*can, may, might, could*)
45. necessity modals (*ought, should, must*)
46. predictive modals (*will, would, shall*)

## **L. Specialized verb classes**

47. public verbs (e.g. *acknowledge, admit, agree*)
48. private verbs (e.g. *anticipate, assume, believe*)
49. suasive verbs (e.g. *agree, arrange, ask*)
50. *seem* and *appear*

## **M. Reduced forms and dispreferred structures**

51. contractions
52. stranded prepositions (e.g. *the candidate that I was thinking of*)
53. split infinitives (e.g. *he wants to convincingly prove that...*)
54. split auxiliaries (e.g. *they are objectively shown to...*)

## **N. Coordination**

55. phrasal coordination (e.g. *NOUN and NOUN, ADJ and ADJ*)
56. independent clause coordination (clause initial *and*) (e.g. *It was my birthday and I was excited.*)

## **O. Negation**

57. synthetic negation (e.g. *no answer is good enough for Jones*)
58. analytic negation: *not* (e.g. *that's not likely*)

\* 本研究で分析対象としない 9 項目に関しては, Abe, Kobayashi & Narita (2013) を参照されたい。