

書き手の習熟度と母語が第 2 言語ライティングに与える影響

小林雄一郎
日本学術振興会

阿部真理子
中央大学

成田真澄
東京国際大学

本研究の目的は、英語学習者のライティングにおける書き手の習熟度と母語が第 2 言語ライティングに与える影響を調査することである。具体的には、東アジア圏（香港、韓国、台湾、日本）の大学生によって書かれた英作文コーパスにおける 58 種類の言語項目（語彙、品詞、統語、談話など）を分析対象とした。その結果、香港人学習者と日本人学習者に関しては、習熟度による影響よりも母語による影響の方が大きいことが示唆された。また、韓国人学習者と台湾人学習者に関しては、母語による影響よりも習熟度による影響の方が大きいことが分かった。

The influence of one's first language and English proficiency level on English language writing

Yuichiro Kobayashi
Japan Society for the
Promotion of Science

Mariko Abe
Chuo University

Masumi Narita
Tokyo International
University

The purpose of this study is to investigate the influence of one's first language and English proficiency on the writing in English as a foreign language. The researchers compared the written production of East Asian learners of English (in Hong Kong, Korea, Taiwan, and Japan) to examine the use of 58 linguistic features such as vocabulary, parts-of-speech, grammar, and discourse particles. The results suggest that there is more influence of first language on the output of learners from Hong Kong and Japan; in contrast, there is more influence of proficiency levels on the output of learners from Korea and Taiwan.

1. はじめに

近年、急速にグローバル化が進んでおり、東アジアのような非英語圏においても、「英語で書くこと」の重要性が高まっている。それにとともに、英語教育の分野でも、ライティング教育の見直しが強求められている。

しかしながら、従来の第 2 言語 (L2) ライティング研究では、英語圏の国々で第 2 言語 (ESL) として学ぶ留学生を対象とした研究が主流で、東アジアなどで外国語としての英語 (EFL) を学ぶ学習者を対象とした研究は圧倒的に少なかった。ESL 環境と EFL 環境の間には、(1) 教室外で英語に触れる機会の差による言語力の違い、(2) 英語を使って専門的な内容を学ぶのか、語彙や文法といった言語そのものを学ぶのかというニーズの違い、が存在する (Ortega, 2004)。従って、ESL ライティング研究の知見を EFL ライティング研究にそのまま取り入れることはできない。

さらに、同じ EFL 環境でも、それぞれの国や地域で実情が大きく異なり、特にヨーロッパ圏と東アジア圏では、その違いが顕著である。東アジア諸国では、(1) 母語と英語の間の言語的距離が

離れている、(2) 英語ライティングに割かれる時間が伝統的に少ない、(3) そもそも母語によるライティング指導のカリキュラムが充実しているとは言い難い、などの共通する課題がある (大井, 2013)。このような状況において、東アジア圏の学習者の言語的特徴や言語発達過程を解明することは急務である。

2. 研究の目的

近年、習熟度に関する様々な指標が提案されており、それらを用いた学習者のパフォーマンスの評価がなされている。しかしながら、同じ習熟度の学習者であったとしても、母語によってパフォーマンスが異なる可能性がある。

本研究の目的は、東アジア圏の大学生によって書かれた英作文を対象に、語彙、品詞、統語、談話などの様々な観点から、学習者の習熟度と母語が第 2 言語ライティングに与える影響を調査することである。また、どのような言語項目が大きな影響を及ぼすのかを特定することも目的としている。

3. 分析の手順

3.1. 分析データ

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進められている。そして、異なる言語的背景を持つ学習者の言語、あるいは母語話者と学習者の言語を比較する対照中間言語分析 (Granger, 1996) が盛んに行われている。

対照中間言語分析の多くは、学習者のデータと母語話者のデータを比較するものである (Gilquin, 2008)。だが、それに対して、母語話者を比較基準とすれば、学習者の言語それ自体を観察できず、比較の誤謬に陥るという批判もある (Bley-Vroman, 1983)。また、そもそも「母語話者」という概念が曖昧であり、母語話者の言語使用にも多くの問題が見られることから、母語話者を絶対的基準として学習者の言語を評価することは適切ではない (Leech, 1998)。従って、本研究では、母語話者のデータを比較対象として用いない。

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) における香港人学習者 (HKG)、韓国人学習者 (KOR)、台湾人学習者 (TWN)、日本人学習者 (JPN) のデータ (2,000 人) である。また、このデータでは、全ての学習者に CEFR (Common European Framework of Reference for Languages) に基づく 5 段階の習熟度情報が付与されている (表 1)。

表 1 分析データの概要 (人数)

	HKG	KOR	TWN	JPN	Total
A2	2	150	58	308	518
B1_1	60	122	174	358	714
B1_2	104	176	122	98	500
B2	30	116	44	34	224
C1	4	36	2	2	44
Total	200	600	400	800	2,000

ICNALE における英作文のトピックは、「大学生のアルバイトへの賛否」と「レストラン全面禁煙への賛否」の 2 種類である。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。

3.2. 分析対象とする言語項目

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、58 項目を分析対象とする (詳細は、本稿末尾の Appendix を参照)。語彙、品詞、文法、談話など、様々なカテゴリーに属する言語項目の使用実態を明らかにすることで、少数の言語的特徴のみを対象とすることが多かった先行研

究と比べて、学習者のライティングを多角的に評価することが可能になる。

3.3. 分析手法

まず、対応分析を用いて、学習者グループ間の関係を概観する。対応分析は、分割表において、行の項目と列の項目の相関が最大になるように行と列の両方を並び替えて、関連性が強いもの同士をまとめていく手法である (Greenacre, 2007)。

次に、階層型クラスター分析を用いて、学習者グループを少数のグループに分類する。階層型クラスター分析は、個体間の類似度あるいは非類似度に基づいて、最も似ている個体から順次に集めてクラスターを作っていく手法である (Gordon, 1999)。

そして、クラスカル・ウォリス検定の効果量 η^2 を用いて、クラスターの分類に寄与した言語項目を特定する。クラスカル・ウォリス検定は、順位に基づくノンパラメトリックの検定で、「対応がない 3 つ以上の条件の中央値が全て等しい」という帰無仮説を検定する手法である (Hollander and Wolfe, 1973)。

最後に、箱ひげ図 (Tukey, 1977) を用いて、各クラスターにおける言語項目の頻度分布の違いを調べる。

4. 結果と考察

4.1. 学習者グループの分類

以下の図 1 は、表 1 における 20 (4 つの国と地域 \times 5 つの習熟度) の学習者グループをケースとし、58 種類の言語項目を変数とする対応分析を行った結果である。

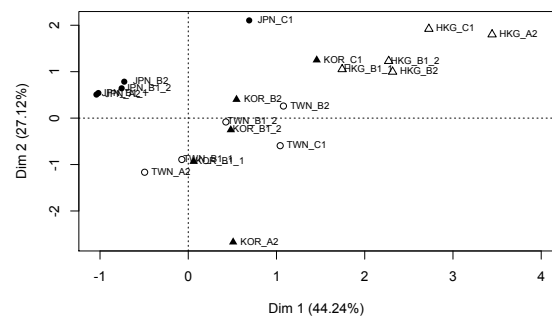


図 1 対応分析 (ケースの布置図)

図 1 を見ると、左上に JPN が分布し、右上に HKG が分布し、中央から下にかけて KOR と TWN が分布している。分析者による解釈の恣意性を軽減するために、対応分析における第 1 次元と第 2 次元の得点に基づき、階層型クラスター分析を行った。その際、Gower 距離 (Gower, 1971) と Ward

法 (Ward, 1963) を用いた。図 2 がその結果である。また、表 2 は、図 2 における分類結果をまとめたものである。

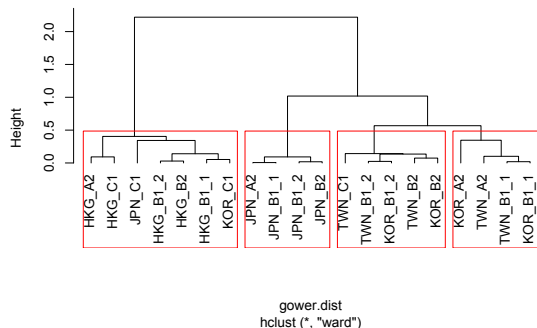


図 2 階層型クラスター分析 (ケースの分類)

表 2 学習者グループの分類結果

第 1 クラスター	全ての香港人学習者 C1 レベルの韓国人学習者 C1 レベルの日本人学習者
第 2 クラスター	A2~B1_2 レベルの日本人学習者
第 3 クラスター	B1_2~B2 レベルの韓国人学習者 B1_2~C1 レベルの台湾人学習者
第 4 クラスター	A2~B1_1 レベルの韓国人学習者 A2~B1_1 レベルの台湾人学習者

表 2 を見ると、大まかに第 1 クラスターが HKG、第 2 クラスターが JPN、第 3 クラスターと第 4 クラスターが KOR と TWN となっており、東アジア圏の学習者による英作文では、習熟度による影響よりも母語による影響の方が大きいことが示唆されている。その一方で、KOR と TWN に関しては、母語ではなく、習熟度での分類がなされている。

KOR と JPN の C1 レベルが HKG と同じクラスターに入っているのは興味深い。Abe, Kobayashi and Narita (2013) が指摘しているように、HKG は、他のアジア圏の学習者と比べて、nominalizations を高い頻度で用いて、first person pronouns をあまり用いない傾向がある。この言語使用のパターンは、他の国における上級の学習者と共通するものである。初級や中級の学習者がこのようなパターンを持っていることについて、香港におけるアカデミック・ライティングの教育効果による可能性が示唆されている (成田, 2013)。

4.2. 分類に寄与した言語項目の特定

本節では、前節の階層型クラスター分析の結果を踏まえて、クラスターの分類に寄与した言語項目の特定を行う。

図 3 は、クラスカル・ウォリス検定を用いて、4 つのクラスター間における中央値の差を検定

し、効果量 η^2 の大きい順にプロットした結果である。なお、効果量 η^2 は、 χ^2 値 / (サンプルサイズ-1) という方法 (Morse, 1999) で求めている。図 3 において、 η^2 が 0.05 以上の言語項目は、first person pronouns, private verbs, *seem* and *appear*, nominalizations, pro-verb *do*, *be* as main verb, demonstrative pronouns, present tense の 8 項目、そして、0.03 以上 0.05 未満の言語項目は、independent clause coordination, contractions, concessive adverbial subordinators, indefinite pronouns, WH-relatives in subject position, split infinitives, emphatics, second person pronouns, WH-clauses, conjuncts, predictive modals, past participle postnominal clauses の 12 項目である。以下、これら 20 種類の言語項目に光を当てる。

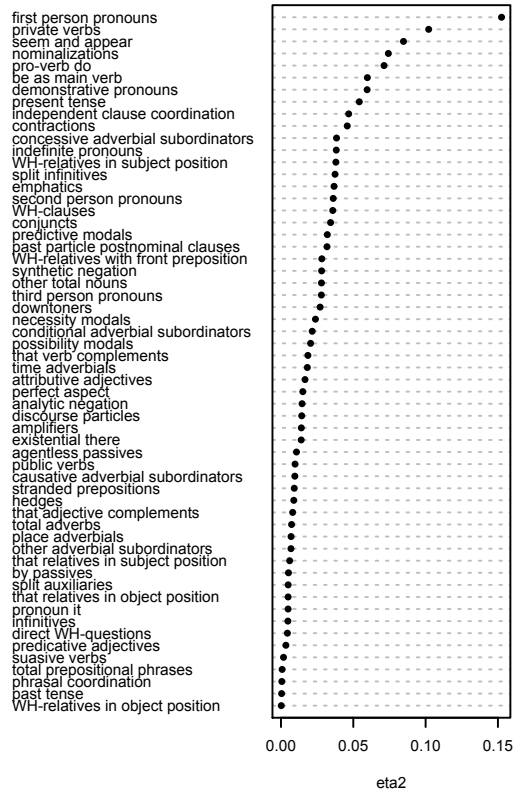


図 3 クラスカル・ウォリス検定の効果量 (変数の寄与度)

図 4 は、 η^2 が 0.03 を超える 20 種類の言語項目に関して、箱ひげ図を描いたものである (横軸の CL1~CL4 は、表 2 の分類結果と対応している)。

図 4 を見ると、第 1 クラスターでは、nominalizations, conjuncts, predictive modals の頻度が高く、first person pronouns, *be* as main verbs, present tense, contractions, second person pronouns の頻度が低い。Biber (1988) によれば、nominalizations は書き言葉に顕著な言語項目であり、first person pronouns, second person pronouns,

contractions は話し言葉に顕著な言語項目である。また、Biber, Johansson, Leech, Conrad and Finegan (1999) は, conjuncts がアカデミック・ライティングにおいて高い頻度で用いられることを示している。そして, Kobayashi (2013) は, *be* as main verbs と present tense が話し言葉と, 初級の学習者による書き言葉に顕著な表現であると報告し

ている。つまり, このクラスターに属する学習者の英作文には, 書き言葉, 特にアカデミック・ライティングに相応しい言語使用のパターンが見られる。以下の例文 (1) は, C1 レベルの HKG による英作文である (Biber の定義に基づく nominalizations, conjuncts, predictive modals の強調は引用者による)。

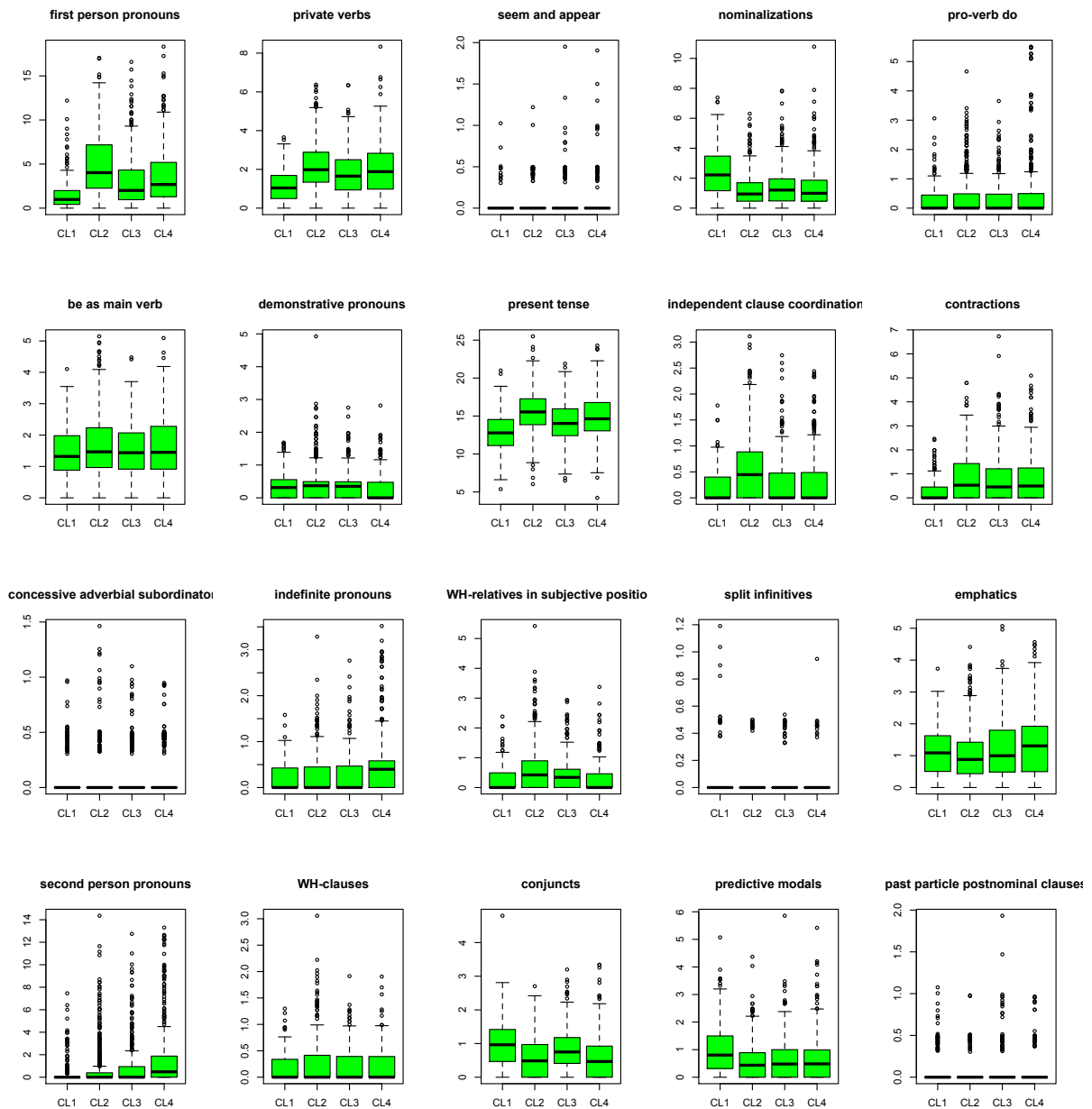


図4 箱ひげ図 (η^2 が 0.03 以上の言語項目)

(1) Someone believed that having a part-time job for a university student is important, but I cannot totally agree. Certainly, life and study in university is quite different from one's previous **education** experience as more individual **motivation** is required and less **rigidity** to

decide one's own learning **activities**. Having a part-time job **will** be a good choice as students can earn their tuition fee and gain some experiences which can make their resumes more impressive and attractive. **However**, for the students who want to be involved in academic

research, having a part time job may not be that essential and even at certain circumstances, can be counterproductive. I **shall** concede that they can gain some experience in team work, time **management** and **communication** skills. But these can also be acquired by doing research projects in the laboratory. **Furthermore**, working part-time can have negative influence on their academic performance as people's energy is not inexhaustible and there **will** be inevitable time conflicts. After all, their results in academic are far more important than working experience. In addition, at the university level, most students are not expected as experienced or equipped with apt skills. Most of the time, they are not likely to be on the appropriate positions where they **will** be after **graduation**. So a part-time job **won't** be efficient to tell them what their future careers look like. (HK_C1)

その一方、第2クラスターを見ると、first person pronouns, private verbs, present tense, independent clause coordination の頻度が高く、nominalizations, emphatics, conjuncts の頻度が低い。つまり、このクラスターは、第1クラスターと対照的に、書き言葉よりも話し言葉に近い特徴を多く持っているということになる。学習者が一般的に first person pronouns を過剰使用することは知られているが (Petch-Tyson, 1998)、とりわけ JPN は高い頻度で用いる。そして、JPN は、private verbs の *think* とともに、*I think* というコロケーションを多用する (Ishikawa, 2009)。なお、第2クラスターに属する学習者による *I* の使用頻度と *think* の使用頻度に関する Spearman の順位相関係数 (Spearman, 1904) は 0.38 であり、無相関検定 (Hollander and Wolfe, 1973) の結果、0.1%水準の有意差が見られた。さらに、independent clause coordination に関しても、JPN による過剰使用が先行研究で報告されている (Kobayashi, 2009)。以下の例文 (2) は、A2 レベルの JPN による英作文である (Biber の定義に基づく first person pronouns, private verbs, present tense, independent clause coordination の強調は引用者による)。

(2) **I** agree with this topic that smoking should be banned at all the restaurants in the country. There are two reasons why **I think** so. First, smoking is bad for our health, especially lung. Smoking is known as the major cause of lung cancer. Smoking also causes many heart or bronchus diseases. If all the restaurants in the country would prohibit smoking, the number of smokers, or the number of smoking times might be decreasing, **I think**. Second, some of the

nonsmoker spaces are insignificant. Many restaurants in Japan have nonsmoking space. But actually, sometimes smoke from tobacco comes to nonsmoking space from smoking space. It is very bad for nonsmoker because most of them are not like smoking smell. **I'm** nonsmoker, **and I** hate smoking smoke too. **And** generally speaking, it is more danger around smokers who have a smoke than smoker themselves. So they hate tobaccos smoke, all the more they are eating. However, for smokers, it is stressful that they cannot smoke, **and** maybe it is bad for their health. But smokers should stop smoking even if they are eating. So **I'm** in favor of this topic. **I** wish it would come true some day, **and** other public institution should be banned too, **I think**. (JPN_A2)

例文 (2) は、文末の *I think*、文頭の *And, But, So* など、かなり話し言葉に近い特徴を持っている。そして、JPN が英作文で *I think* という表現を過剰使用するのは、必ずしも英語力だけが原因ではない。小林 (2011) は、本研究の分析データと同一の学習者グループによって同一のテーマで書かれた日本語作文を解析している。その結果、JPN は、彼らの母語である日本語で作文を書く場合においても、「私」や「思う」のような表現を多用することが明らかにされている。以下の例文 (3) と (4) は、その一例である。

(3) **私**は大学生にとってアルバイトはさほど重要ではないと考えている。少なくとも最優先すべき事ではないであろう。確かにアルバイトをする事によって少しでも社会を見る事が出来たり、**自ら**お金を稼ぐ大変さを学ぶという意義はあるし、**私**も大学生になってから現在に至るまで欠かさずアルバイトを続けている。しかし、**私**は最低限を稼ぐに過ぎず後は別の事に時間を割いている。(cjejus_ptj_010.txt) (小林, 2011)

(4) 全面禁煙化 [原文ママ] 喫煙できるかのどちらかではなく、両者にとって快適な時間を提供することがレストランには必要だと**思います**。2つ目について、喫煙者が喫煙可能なレストランを選んで利用することがあります。また、煙草税は税収の大きな部分を担っており、政府としても煙草税を廃止しようとは**思っていません**。このように、煙草の背景には多額の利益が隠れています。以上の2つより、私は全てのレストランで禁煙を行うべきではないと**思います**。(cjejus_smk_020.txt) (小林, 2011)

最後に、第3クラスターと第4クラスターについて見てみよう。これらのクラスターに属する学習者グループは、図1の原点付近に分布していることから分かるように、第1クラスターと第2クラスターほどに際立った特徴を持っているとは言えない。しかし、図4を見ると、特に第4クラスターにおいて、private verbs, indefinite pronouns, emphatics, second person pronounsが高い頻度で現れている。以下の例文(5)と(6)は、それぞれA2レベルのKORとB1_1レベルのTWNによる英作文である(Biberの定義に基づくindefinite pronounsとemphaticsの強調は引用者による)。

(5) I think part time job is important. First, it makes me to feel importance of saving money. Before I entered the university, I already had experienced part time job. I worked in the factory in 5 days, in a resting place in 3 days and in restaurant in 3 days. I don't remember exactly due to work very shortly. But it was very tired and terrible. Whenever parents gave allowance, I had done anything to do. I bought clothes, ate out and went out to play with my friends. I didn't pay attention to the money before part time jobs did. I **really** knew how hard it was to earn the money. I thanked for parents after working. It is not bad that earning the tuition for myself. But one of bad points isn't time for me. Second, I felt **a lot** of things. At the end of each working, I thought that I **really** study hard. The other word, I **really** don't want work because part time job is too tired physically. But it makes me feel lots of things. For things that I want to do, study is very important. And I feel that there aren't easy things in the world. The little things need too much effort. (KOR_A2)

(6) I think it is an important. If you have been working some where, you may know **someone**, and have some personal networks. **Someone** is not agree have part-time job is good thing; you can learn **more** profession knowledge in the same time. When you are young, you have a good memory, so you can use this advantage to do **something** that has significance. In **someone's** life, they need to do **something** for living, because they don't have a good background to support them. And they need work for them or they family. In this situation, the children may to face some difficult thing, such as bad guys, drug attractive. They can be a sturdy guys or depraved guys. In different life, you have choice to select what you want to do.

If you want test what you can do, what guys you can be, how do get you want to have, you just to do, but don't forget why you study in the college? Why don't you have full-job to work? You need think clearly. Otherwise how about give sometime to think yourself, and ask **someone** about your ideals. And you will have a different suggests, just to do what you want, but don't do a bad thing. (TWN_B1_1)

まず、引用(5)を見ると、emphaticsのreallyが繰り返し用いられている。また、Biber(1988)の定義するemphaticsには含まれないものの、very(very shortly, very tired, very important)やtoo(too tired, too much short)のように大げさな表現を好んで使っている。次に、引用(6)を見ると、indefinite pronounsのsomeoneやsomethingを多用している。また、それ以外にも、some where, some, sometimeなどの表現や助動詞のmayを用いるなど、曖昧な言い回しを好んでいる。これらの2つの引用に共通しているのは、いずれの書き手も物事をはっきりと書かない点である。このような表現がKORやTWNの母語の影響によるものなのかどうか、それとも教育効果などによるものなのかを判断するには、母語話者の協力に基づく詳細な分析が必要である。

5. おわりに

本研究では、東アジア圏の大学生によって書かれた英作文を対象に、語彙、品詞、統語、談話などの観点から、学習者の習熟度と母語が第2言語ライティングに与える影響を考察してきた。そして、対応分析と階層型クラスター分析の結果から、HKGとJPNに関しては、習熟度による影響よりも母語による影響の方が大きいことが分かった。また、KORとTWNに関しては、母語による影響よりも習熟度による影響の方が大きいことが示唆された。また、HKGの英作文にはnominalizations, conjuncts, predictive modals, JPNの英作文にはfirst person pronouns, private verbs, present tense, independent clause coordination,そして、KORやTWNの英作文にはprivate verbs, indefinite pronouns, emphatics, second person pronounsが顕著であることが明らかにされた。

今後は、アジア圏における他の国や地域の学習者データも含めた分析を行う予定である。そして、そのような分析を通して、(1)アジア圏の学習者に望ましいカリキュラム、(2)効果的な教授法と教材、(3)求められるライティング教育支援環境などを実現するための基礎資料を提供することを目指す。英語を外国語として学ぶアジア全体に共通する課題を明らかにし、新たなライティング教育のモデルを提案することができれば、その意義は大きい。

謝辞

本研究の成果の一部は、科学研究費補助金（特別研究員奨励費（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

- 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.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.
- Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: The case of systematicity. *Langugae Learning*, 33(1), 1-17.
- Gilquin, G. (2008). Combining contrastive and interlanguage analysis to apprehend transfer: Detection, explanation, evaluation. In G. Gilquin, S. Papp & M. B. Diez-Bedmar (Eds.), *Linking up contrastive and learner corpus research* (pp. 3-33). Amsterdam: Rodopi.
- Gordon, A. D. (1999). *Classification*. 2nd ed. London: Chapman and Hall.
- Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 27(4), 857-871.
- Granger, S. (1996). From CA to CIA and back: An integrated contrastive approach to bilingual and learner computerized corpora. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Language in contrast: Text-based cross-linguistic studies* (pp. 37-51). Lund: Lund University Press.
- Greenacre, M. (2007). *Correspondence analysis in practice*. 2nd ed. London: Chapman and Hall.
- Hollander, M., & Wolfe, D. A. (1973). *Nonparametric statistical methods*. New York: John Wiley and Sons.
- 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.
- Kobayashi, Y. (2009). Profiling metadiscourse markers in native and non-native English. *Lexicon*, 39, 1-17.
- 小林雄一郎 (2011). 「コンパラブル・コーパスを用いた英作文と日本語作文の使用語彙に関する対照研究」 野瀬昌彦 (編) 『日本語と X 語の対照一言語を対照することでわかること』 (pp. 37-47). 名古屋: 三恵社.
- Kobayashi, Y. (2013). A comparison of spoken and written learner corpora: Analyzing developmental patterns of vocabulary used by Japanese EFL learners. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *Twenty years of learner corpus research: Looking back, moving ahead* (pp. 277-287). Louvain-la-Neuve: Presses universitaires de Louvain.
- Leech, G. (1998). Preface. In S. Granger (Ed.), *Learner English on computer* (pp. xiv-xx). Harlow: Addison Wesley Longman.
- Morse, D. T. (1999). Minsize2: A computer program for determining effect size and minimum sample size for statistical significance for univariate, multivariate, and nonparametric tests. *Educational and Psychological Measurement*, 59(3), 518-531.
- 成田真澄 (2013). 「香港出張報告」 大井恭子 (編) 『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』 (pp. 94-101). 平成 24~27 年度文部科学省科学研究費補助金基盤研究(B) 成果報告書.
- 大井恭子 (編) (2013). 『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』 平成 24~27 年度文部科学省科学研究費補助金基盤研究(B) 成果報告書.
- Ortega, L. (2004). L2 writing research in EFL contexts: Some challenges and opportunities for EFL researchers. *Applied Linguistic Association of Korea Newsletter, Spring*, 1-10.
- Petch-Tyson, S. (1998). Writer / reader visibility in EFL written discourse. In S. Granger (Ed.), *Learner English on computer* (pp. 107-118). Harlow: Addison Wesley Longman.
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1), 72-101.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading: Addison-Wesley.

Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of American Statistical Association*, 58(301), 236-244.

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、紙面の都合で割愛するが、詳しくは、Abe, Kobayashi and Narita (2013) を参照されたい。