L2ライティングにおけるメタ談話標識の頻度を用いたL1グループの分類

小林, 雄一郎

(Citation)
Learner Corpus Studies in Asia and the World,5:43-48

(Issue Date)
2020-12-21

(Resource Type)
departmental bulletin paper

(Version)
Version of Record

(JaLCDOI)
https://doi.org/10.24546/81012488

(URL)
https://hdl.handle.net/20.500.14094/81012488

Learner Corpus Studies in Asia and the World, vol. 5 (Papers from LCSAW2020) ISSN: 2435-2632
Published by School of Language & Communication, Kobe University, Japan

L2 ライティングにおけるメタ談話標識の頻度を用いた L1 グループの分類 [1)]

小林　雄一郎
日本大学

概要

本研究では，アジア人英語学習者の英語ライティングにおけるメタ談話標識の頻度を用いて，6 つの L1 グループ（中国，インドネシア，日本，韓国，台湾，タイ）を分類する。そして，機能カテゴリーごとの使用傾向を分析することで，東アジア圏の学習者と東南アジア圏の学習者による言語使用の違いを明らかにする。

キーワード

学習者コーパス，ライティング，メタ談話標識

1. はじめに

第二言語（L2）における母語（L1）の影響は，第二言語習得分野における長年のトピックであり，学習者コーパス研究でも多くの先行研究がある。たとえば，Murakami (2013) は，7 つの L1 グループを比較して，形態素の習得過程における L1 の影響を明らかにしている。また，Leacock, Chodorow, Gamon and Tetreault (2014) は，7 つの L1 グループを比較し，L1 に冠詞があるグループと冠詞がないグループでは冠詞使用のエラー率に差があると報告している。

そして，L1 の言語的距離が L2 の言語的距離となるのではないか，という指摘もある。これに関して，Nagata and Whittaker (2013) は，11 の L1 グループを比較し，L2 ライティングにおける統語的特徴に注目した。その結果，同じヨーロッパ人学習者でも，インド・ヨーロッパ語族における語派による違いがあるという結論にいたっている。また，小林 (2010) は，17 の L1 グループを比較し，L2 ライティングにおける談話的特徴に注目した。その結果，ヨーロッパ人学習者とアジア人学習者に大きな差があり，ヨーロッパ人学習者の中でもゲルマン語系とロマンス語系による違いがあると報告している。このように，複

数の先行研究において，L2 における L1 の影響が示唆されている。

このような流れを受けて，本研究では，アジア人学習者の英語ライティングにおけるメタ談話標識の使用傾向を分析し，L1 グループの違いによる言語使用の違いを明らかにする。

2. コーパス

本研究の分析データには，International Corpus Network of Asian Learners of English (ICNALE) (Ishikawa, 2013) に含まれる L2 ライティングを用いた。具体的には，中国，インドネシア，日本，韓国，台湾，タイの学習者を分析対象とする。また，part-time job という単一のトピック，B1 という単一の習熟度のデータのみを分析する。表 1 は，分析データの概要をまとめたものである。

表 1
分析データの概要

	Participants	Total words
China (CHN)	337	83980
Indonesia (IDN)	165	39096
Japanese (JPN)	228	51780
Korea (KOR)	149	34175
Taiwan (TWN)	148	35294
Thailand (THA)	279	64186

3. 分析方法

3.1 メタ談話標識

本研究で用いる分析項目は，メタ談話標識と呼ばれる談話表現の一種である。メタ談話は，テキストの命題内容に直接影響を与えないが，首尾一貫した論理的な文章を構成したり，読み手が命題内容や文章の展開，読み手に対する書き手の立場を理解したりする際に役立つ。そして，メタ談話標識は，メタ談話における様々な機能を担う表現のことで，アカデミックライティングの研究などで，分析対象とされることが多い。メタ談話標識の定義には様々なものがあるが，ここでは，Hyland (2005) のリストを用いる。このリストは，それ以前のメタ談話標識に関する先行研究を網羅的に検討して作成されたものである。

Hyland のメタ談話標識のリストには，10 種類の機能カテゴリーに分類される約 500 の表現が収録されている。表 2 は，その 10 種類の機能カテゴリーをまとめたものである。

表 2

Hyland のメタ談話標識のリストにおける 10 種類の機能カテゴリー

Category	Function	Examples
Interactive resources	Help to guide reader through the text	
Transitions (TRA)	Express semantic relation between main clauses	*in addition, but, thus, and*
Frame markers (FRM)	Refer to discourse acts, sequences, or text stages	*finally, to conclude, my purpose here is to*
Endophoric markers (END)	Refer to information in other parts of the text	*noted above, see Fig, in section 2*
Evidentials (EVI)	Refer to source of information from other texts	*according to X, (Y, 1990), Z states*
Code glosses (COD)	Help readers grasp functions of ideational material	*namely, e.g., such as, in other words*
Interactional resources	Involve the reader in the argument	
Hedges (HED)	Without writer's full commitment to proposition	*might, perhaps, possible, about*
Boosters (BOO)	Emphasize force or writer's certainty in proposition	*in fact, definitely, it is clear that*
Attitude markers (ATM)	Express writer's attitude to proposition	*unfortunately, I agree, surprisingly*
Engagement markers (ENG)	Explicitly refer to or build relationship with reader	*consider, note that, you can see that*
Self-mentions (SEM)	Explicit reference to author(s)	*I, we, my, our*

3.2 クラスタリング

6 つの L1 グループ（表 1）から 10 種類の機能カテゴリー（表 2）ごとの使用頻度（正規化頻度）を求めたら，ヒートマップつきのクラスター分析（Kobayashi, 2016b）を用いて，L1 グループのクラスタリングを行う。なお，クラスター分析には，標準化ユークリッド距離とウォード法を用いる。

4. 分析結果

表 3 は，6 つの L1 グループから 10 種類の機能カテゴリーごとの使用頻度を集計した結果である。表中の値は，それぞれ列合計で正規化されている。

表 3

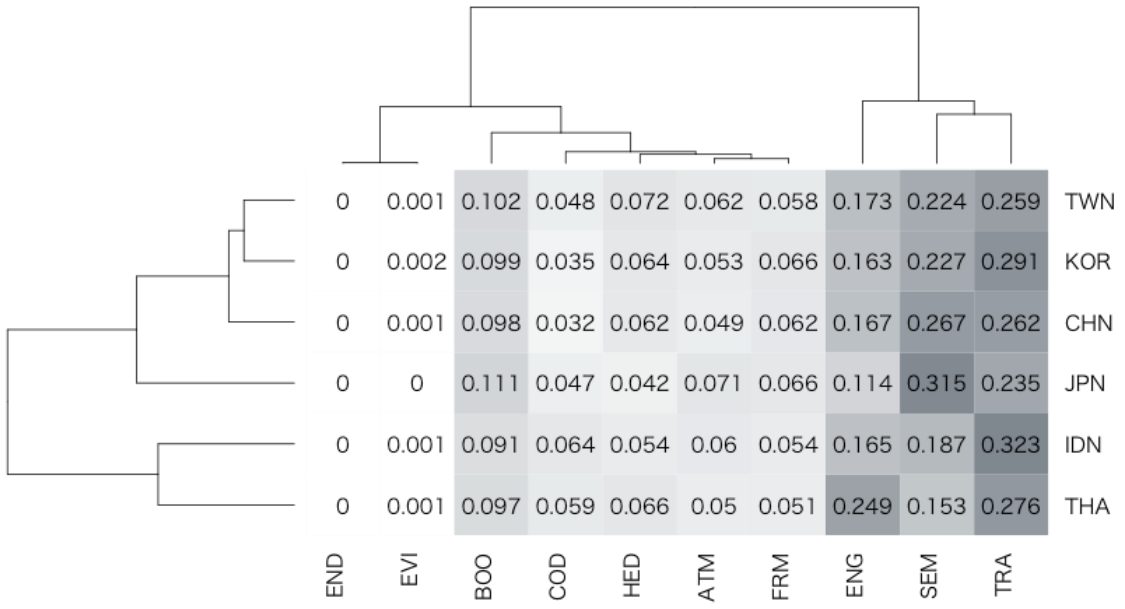
L1 グループから 10 種類の機能カテゴリーごとの使用頻度を集計した結果

	CHN	IDN	JPN	KOR	TWN	THA
TRA	0.262	0.323	0.235	0.291	0.259	0.276
FRM	0.062	0.054	0.066	0.066	0.058	0.051
END	0.000	0.000	0.000	0.000	0.000	0.000
EVI	0.001	0.001	0.000	0.002	0.001	0.001
COD	0.032	0.064	0.047	0.035	0.048	0.059
HED	0.062	0.054	0.042	0.064	0.072	0.066
BOO	0.098	0.091	0.111	0.099	0.102	0.097
ATM	0.049	0.060	0.071	0.053	0.062	0.050
ENG	0.167	0.165	0.114	0.163	0.173	0.249
SEM	0.267	0.187	0.315	0.227	0.224	0.153

そして，図 1 は，集計した頻度を用いて，6 つの L1 グループをクラスタリングした結果である。この図のヒートマップでは，セルの頻度が高いほど，濃いグレーで示している。

図 1

L1 グループのクラスタリング結果

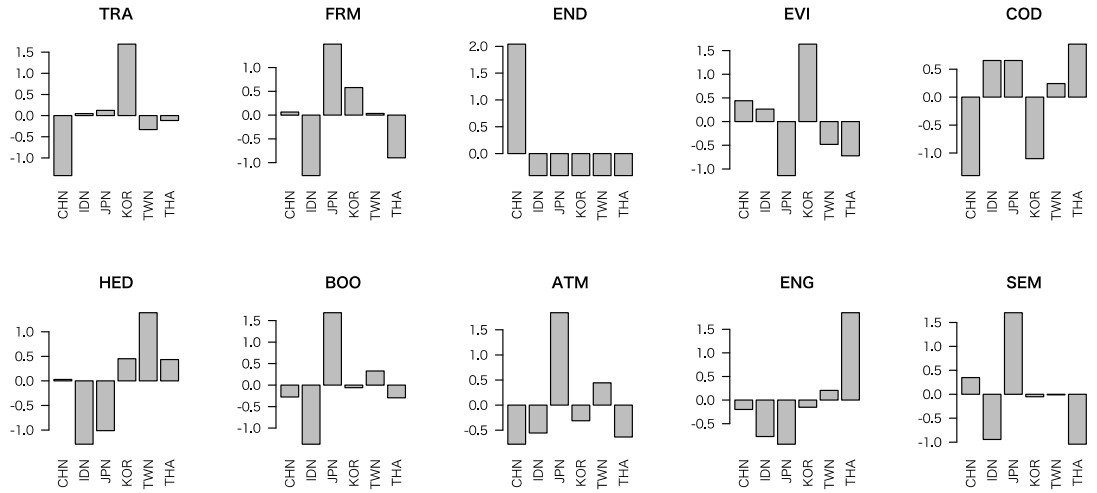


この図を見ると，東アジアの学習者（台湾，韓国，中国，日本）と東南アジアの学習者（インドネシア，タイ）が大きく 2 つに分かれている。言い換えれば，これら 2 つの地域

の学習者では，ライティングにおける 10 種類の機能カテゴリーの使用パターンが大きく異なるということが示されている。

図 2 は，6 つの L1 グループにおける 10 種類の機能カテゴリーの *z* 得点を可視化したものである。この図では，0 よりも大きい値を持っているグループが 6 グループの平均よりも高い頻度でそのカテゴリーの表現を使用しており，0 よりも小さい値を持っているグループが 6 グループの平均よりも低い頻度でそのカテゴリーの表現を使用しているということが示されている。

図 2
L1 グループごとの 10 種類の機能カテゴリーの *z* 得点



図中で最も目を引く点は，日本人学習者とタイ人学習者のコントラストである。具体的には，日本人学習者が SEM，つまり 1 人称代名詞を多用するのに対して，タイ人学習者は ENG，つまり 2 人称代名詞を多用している。言い換えれば、日本人学習者のライティングでは書き手の存在が前景化されているのに対して，タイ人学習者のライティングでは「あなたはどう思いますか」や「どのように対処しますか」といった問いかけが見られる (Kobayashi, 2016a)。また，日本人学習者が BOO，つまり強調表現を多用するのに対して，タイ人学習者は HED，つまり緩衝表現を多用している。この点においても，これら 2 つのグループのライティングでは，書き手のスタンスの取り方が異なっている。

5. おわりに

本稿では，アジア人英語学習者の英語ライティングにおけるメタ談話標識の頻度を用いて，6 つの L1 グループをクラスタリングした。その結果，アジアの学習者と東南アジアの学習者が大きく 2 つに分けられることが明らかになった。今後の課題としては，(1) より多

くの L1 グループを分析対象とすること，(2) 今回分析しなかった習熟度の学習者やライティングのトピックも分析すること，(3) 機能カテゴリーに含まれる個々のメタ談話標識の使用例を分析すること，などが考えられる。

注

[1)] 本稿は，Kobayashi (2016a) の一部を加筆・修正したものである。分析結果の詳細などについては，そちらを参照されたい。

引用文献

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing.* New York: Continuum.

Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. *Learner Corpus Studies in Asia and the World, 1*, 91-118.

小林雄一郎 (2010). 「多変量解析による世界の英語学習者の談話分析」 『人文科学とコンピュータシンポジウム論文集―人文工学の可能性』(pp. 41-48). 東京: 情報処理学会.

Kobayashi, Y. (2016a). Investigating metadiscourse markers in Asian Englishes: A corpus-based approach. *Language in Focus: International Journal of Studies in Applied Linguistics and ELT, 2*(1), 19-35

Kobayashi, Y. (2016b). Heat map with hierarchical clustering: Multivariate visualization method for corpus-based language studies. *NINJAL Research Papers, 11,* 25-36.

Leacock, C., Chodorow, M., Gamon, M., & Tetreault, J. (2014). *Automated grammatical error detection for language learners.* Second edition. San Rafael: Morgan and Claypool Publishers.

Murakami, A. (2013). Cross-linguistic influence on the accuracy order of L2 English grammatical morphemes. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *Twenty years of learner corpus research: Looking back, moving ahead* (pp. 325-334). Louvain-la-Neuve: Presses universitaires de Louvain.

Nagata, R., & Whittaker, E. (2013). Reconstructing an Indo-European family tree from non-native English texts. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics,* 1137-1147.