Journal of Pan-Pacific Association of Applied Linguistics, 21(2), 41-54.
https://doi.org/10.25256/PAAL.21.2.3

Developmental Patterns of Metadiscourse in Second Language Writing*

Yuichiro Kobayashi
Nihon University

Kobayashi, Y. (2017). Developmental patterns of metadiscourse in second language writing. *Journal of Pan-Pacific Association of Applied Linguistics, 21*(2), 41-54.

The present study aimed to profile the developmental patterns of discourse in second language (L2) writings among different first language (L1) groups. Applying the list of metadiscourse markers proposed by Ken Hyland to learner language, this study investigates variation of metadiscourse across proficiency levels, as well as across L1 backgrounds. Using the International Corpus Network of Asian Learners of English, the present study compared the frequencies of metadiscourse markers used in the writings among different learner groups. The results suggest that the six learner groups that were compared have diverse frequency change patterns of metadiscourse features across proficiency levels. To be specific, Japanese learners' heavy reliance on self-mentions and boosters is remarkably antithetical to Thai learners' preference of engagement markers and hedges. Moreover, B2 and higher level learners in China and Taiwan exhibited greater numbers of evidential patterns (*hereafter* evidentials) than learners in other groups. These differences can be attributable to their L1 rhetorical strategy, not to their lexical and grammatical competence. Therefore, we should consider the idiosyncrasies in metadiscourse of each L1 group when assessing L2 learners based on their language performance.

Keywords: contrastive rhetoric, metadiscourse, learner corpus

1 Introduction

The understanding of language learners' developmental patterns is one of the central issues in second language acquisition (SLA) research. Now that a wide variety of computer learner corpora are available, SLA researchers can use them to focus afresh on descriptive facets of interlanguage processes and identify in an increasingly meticulous manner the characteristics of learner language at different developmental stages (Tono, 2013). For example, the largest-scale project describing interlanguages, the English Profile

* An earlier version of the present study was reported at the 9th International Corpus Linguistics Conference held at the University of Birmingham on July 26th, 2017. This work was supported by Grants-in-Aid for Scientific Research Grant Numbers 26770205 and 17K13511.

Programme, aims to uncover what learners of English *can* and *cannot* do with the language at each of the Common European Framework of Reference for Languages (CEFR) levels (Hawkins & Filipović, 2012). By analyzing a large amount of learner performance data, the project is trying to discover the distribution of correct use and distribution and features of misuse, overuse, and underuse of linguistic features of English, in particular criterial features that can be used to differentiate CEFR levels.

Developmental patterns in a second language (L2) can vary across learners' first languages (L1s) due to the linguistic and cultural features of different L1 communities. Murakami (2013) demonstrated the L1 influence on L2 acquisition order of grammatical morphemes by investigating a corpus consisting of more than 3,000 essays across seven L1 groups. In addition, Leacock, Chodorow, Gamon and Tetreault (2014) showed that the native language of learners significantly affects the likelihood of article error by contrasting four learner groups whose L1s have articles with three other groups whose L1s do not. Given these L1 effects upon L2 performance, we should consider L1-specific criterial features when profiling the language use of learner groups from different L1 backgrounds.

These L1-derived differences in L2 use also reflect back and show the distances among the learners' L1s. Comparing the syntactic features characteristic of L2 English writings across L1s, Nagata and Whittaker (2013) succeeded in reconstructing an Indo-European family tree from non-native English texts. Using natural language processing and statistical modeling techniques, they automatically classified European learners of English from 11 different countries into three branches the Indo-European family: Italian, Germanic, and Slavic. Moreover, they proved that English language use of European learners as a whole is greatly dissimilar from that of Asian learners of English. Therefore, L1 influence is not a negligible factor in L2 writing research.

A learner's L1 can affect discourse as well as syntax in their L2 written productions. Contrastive rhetorical studies demonstrate that the discourse characteristics of L2 writing clearly reflect rhetorical preferences in the learner's L1 (Conner, 1996). In line with the tradition of contrastive rhetoric, Kobayashi (2016) investigated differences in rhetorical preferences in L2 writings by learners of different L1 groups, comparing the use of metadiscourse markers in L2 essays written by six learner groups in Asia. Performing multivariate statistical analyses, Kobayashi found a substantial difference in metadiscourse between East and Southeast Asian groups and identified discourse devices used to distinguish different L1 groups.

The metadiscursive framework used in the present study was established by Hyland (2005) and is the most widely accepted in the field of corpus-based discourse analysis. The application of Hyland's framework to learner corpus research also enables us to describe variation across proficiency levels. Furthermore, contrastive analysis of L2 metadiscourse

development leads to the further identification of L1-specific criterial features pertaining to discourse.

2 Metadiscourse

Hyland's framework for metadiscourse is based on a wide range of research findings accumulated over the history of discourse analysis. An early study, Harris (1959) coined the term metadiscourse in order to refer to the writer's effort to guide a reader's perception of a text. Williams (1981) then categorized written metadiscourse into three types: (a) *hedges and emphatics,* (b) *sequencers and topicalizers*, and (c) *narrators and attributors*. Vande Kopple (1985) and Crismore (1989) further developed the concept, and revised the categories of metadiscourse. Hyland (2005), reflecting a general trend, defined metadiscourse as "the cover term for the self-reflective expressions used to negotiate interactional meanings in a text, assisting the writer (or speaker) to express a viewpoint and engage with readers as members of a particular community" (p. 37). On the basis of the above definition, he advanced a taxonomy of metadiscourse consisting of two large categories, interactional resources and interpersonal resources, as shown in Table 1.

Table 1. Hyland's Taxonomy of Metadiscourse

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 finally,*
Frame markers (FRM)	Refer to discourse acts, sequences, or text stages	*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*

Employing the taxonomy shown in Table 1 for the study of graduate student writing in six different fields, Hyland and Tse (2004) demonstrated that all ten functional categories of metadiscourse are more frequently used in doctoral dissertations than in master's theses, and that the use of metadiscourse categories varies across academic disciplines.

Hyland's taxonomy of metadiscourse has had a great impact on learner corpus research. Based on the metadiscourse schema, Hong and Cao (2014) compared descriptive and argumentative English essays written by Chinese, Spanish, and Polish learners of English, and showed statistically significant differences among the three learner groups in the use of interactional metadiscourse. Tan and Eng (2014) examined the use of metadiscourse in writing by novice and advanced English learners in Malaysia, and found that with the aim of building rapport with their readers, both learner groups exhibited a greater preference for the use of interactional than interactive resources. In contrast, Attarn (2014) investigated the use of two types of metadiscourse resources, in research articles written by Iranian learners and native speakers of English, and demonstrated that, with the purpose of making explicitly the relationships between independent discourse units, both groups used interactive features more commonly than interactional. However, these and other previous learner-corpus-based discourse analyses have involved comparing native and nonnative groups, and have given little attention to L2 discourse development owing to the lack of large learner corpora coded with proficiency levels.

3 Research Design

3.1 Purpose of the study

In order to address a lacuna in previous scholarship, the present study aimed to profile the developmental patterns of metadiscourse in L2 essays written by learners from different L1 groups. More specifically, the study applies the list of metadiscourse markers proposed by Hyland (2005) to learner language in order to investigate variation across both L1 backgrounds and proficiency levels. The findings should contribute to the identification of L1-specific criterial features regarding English discourse.

3.2 Corpus data

The present study draws on the International Corpus Network of Asian Learners of English (ICNALE) (Ishikawa, 2013), the largest Asian composition database in English. The data investigated here are a subset from this database, including L2 writings from six L1 groups (Chinese, Indonesian, Japanese, Korean, Taiwanese, and Thai). From the standpoint of World

Englishes, these groups fall within the "expanding circle" of English-speakers (Kachru, 1992). In this corpus, all learners were classified into three levels based on the CEFR: A2 (Waystage), B1 (Threshold), or B2+ (Vantage or higher). The writing conditions were rigorously controlled for the contrastive analysis of these groups. All compositions in the subset were written in response to a single topic, namely "It is important for college students to have a part-time job" (Ishikawa, 2013, p. 97). Table 2 provides basic information on the subset.

Table 2. Basic Information on Sample

	CEFR A2 level		CEFR B1 level		CEFR B2+ level	
	N	Total words	*N*	Total words	*N*	Total words
Chinese (CHN)	50	22,640	337	83,896	13	3,360
Indonesian (IDN)	32	7,385	165	39,085	3	791
Japanese (JPN)	154	34,959	23	51,778	18	4,281
Korean (KOR)	75	16,804	149	34,126	76	18,664
Taiwanese (TWN)	29	6,491	148	35,294	23	5,856
Thai (THA)	119	26,866	279	64,166	2	514
Total	459	115,145	1,101	308,345	135	33,466

3.3 Data analysis

Starting from the presumption that a "unique matrix of frequencies of various linguistic forms" characterizes every interlanguage (Krzeszowski, 1990, p. 212), this study compared the frequencies of metadiscourse markers used by different learner groups through correspondence analysis (Greenacre, 2016). An advantage of this method lies in its high replicability, which emerges because there are few options in the calculation process compared to other clustering techniques like factor analysis (Nakamura, 1995). Another advantage is that the method graphically represents the relationships between learner groups and metadiscourse characteristics in a two-dimensional scatter plot (Glynn, 2014). It also provides a statistical summary of the characteristics of variation, and therefore can be utilized as a first step to consider which metadiscourse features should be investigated in more detail. After identifying analysis points for further investigation, the present study tracked frequency change patterns of metadiscourse features across proficiency levels in each L1 group.

4 Results and Discussion

This study started by tabulating the frequencies of ten metadiscourse categories in the writings of all learner groups shown in Table 2. Table 3 lists the adjusted frequencies (per 10,000 words) of categories used by each group. This frequency matrix is too large to manually find all meaningful patterns underlying it; to deal with the problem, the author implemented correspondence analysis, which can help us gain an intuitive understanding of the notable associations among learner groups and metadiscourse features.

Table 3. Adjusted Frequencies of Ten Metadiscourse Categories

		TRA	FRM	END	EVI	COD
CHN	A2	450.60	149.61	0.00	0.89	49.58
	B1	444.45	120.12	1.08	1.92	54.85
	B2+	437.13	110.78	0.00	5.99	47.90
IDN	A2	500.74	86.62	4.06	0.00	85.26
	B1	482.91	93.66	7.44	1.80	96.22
	B2+	471.34	127.39	0.00	0.00	114.65
JPN	A2	489.20	128.75	0.00	0.29	86.89
	B1	477.34	137.66	0.19	0.58	94.81
	B2+	429.48	100.92	0.00	0.00	105.61
KOR	A2	561.70	94.41	0.60	0.00	58.56
	B1	510.63	124.35	2.65	2.94	61.15
	B2+	438.31	107.43	0.54	2.15	99.37
THA	A2	522.18	91.56	4.84	2.23	100.86
	B1	472.89	96.17	3.28	0.94	100.71
	B2+	564.20	77.82	0.00	0.00	116.73
TWN	A2	481.93	98.86	0.00	0.00	83.41
	B1	467.72	119.56	0.85	1.14	86.83
	B2+	497.95	104.38	1.71	5.13	94.11

Table 3. Adjusted Frequencies of Ten Metadiscourse Categories (Continued)

		HED	BOO	ATM	ENG	SEM
	A2	130.13	184.14	110.66	222.20	393.94
CHN	B1	147.31	165.99	82.88	241.09	452.71
	B2+	161.68	143.71	65.87	176.65	338.32
	A2	123.16	146.16	67.67	166.46	358.64
IDN	B1	100.33	135.48	89.55	227.86	280.20
	B2+	114.65	76.43	50.96	292.99	356.69
	A2	147.11	243.74	137.07	195.57	647.78
JPN	B1	135.53	225.68	143.47	181.67	639.43
	B2+	199.48	239.38	152.55	183.06	469.37
	A2	128.47	172.69	106.96	328.65	322.08
KOR	B1	135.81	170.21	93.78	262.81	395.10
	B2+	156.85	136.97	92.39	189.07	248.70
	A2	157.44	168.97	82.25	350.23	308.17
THA	B1	163.57	166.86	86.63	377.19	264.59
	B2+	252.92	136.19	77.82	642.02	116.73
	A2	142.11	197.71	111.21	264.13	472.66
TWN	B1	165.96	184.75	111.02	276.13	403.95
	B2+	131.76	121.49	118.07	253.25	248.12

Figure 1 shows the results of correspondence analysis in the two most significant dimensions, which account for 78.78% of total variation in the frequency matrix. The positioning of each group and feature on the horizontal axis (Dim 1) deserves particularly close examination due to its large contribution ratio (69.50%). The coordinates on the scatter plot reflect the interrelationships between all learner groups, the relative similarity between the ten metadiscourse categories, and the association patterns between the groups and the categories. The most prominent feature of this diagram is that self-mentions (SEM) and engagement markers (ENG) are distributed on opposite sides of the horizontal axis; under the theory of writer-reader

visibility, first- and second-person pronouns are used to express the personal feelings of author and to interact with readers, respectively (Petch-Tyson, 1998). Another notable feature is the positioning of hedges (HED) and boosters (BOO) on the axis; these two metadiscourse devices represent the writer's level of commitment to a proposition (Hyland, 2005). Turning to the relationship between these four types of interactional resources and learner groups, self-mentions and boosters are especially characteristic of Japanese learners (JPN) and engagement markers and hedges of Thai learners (THA).

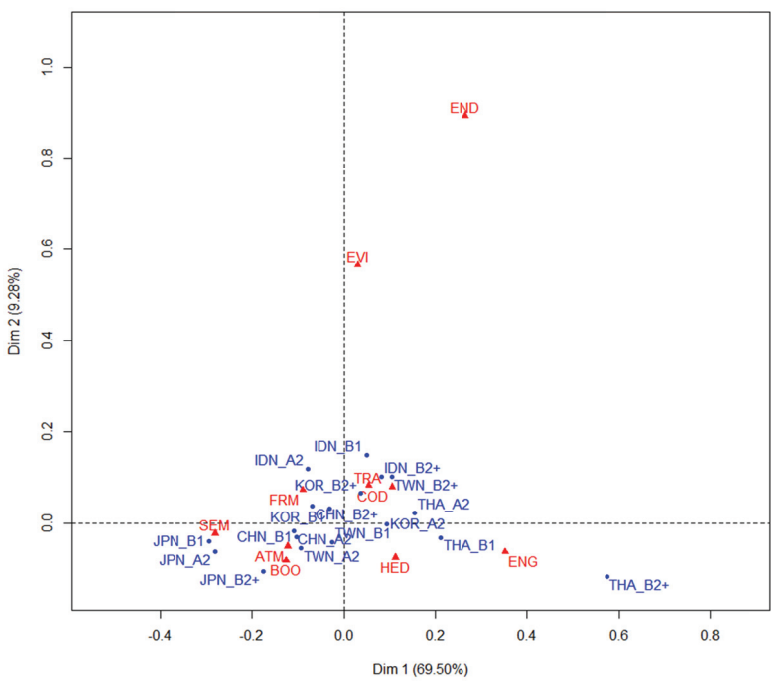

Figure 1. Scatter plot displaying the results of correspondence analysis

The next step was to meticulously investigate the metadiscourse devices that can be used as L1-specific criterial features. Given the results of correspondence analysis, self-mentions and engagement markers are the first linguistic features to be considered. Figure 2 shows the patterns of change in frequency of self-mentions, which especially characterize Japanese learners' writings. Although self-mentions decline sharply in their essays between B1 and B2+ level, they use this metadiscourse feature more frequently than any other learner group. In contrast, its frequency in Thai learners' essays is constantly lower than among the other groups. In addition, five of six groups show falling frequency of self-mentions between B1 and B2+ level, only in Indonesian learners does frequency rise.

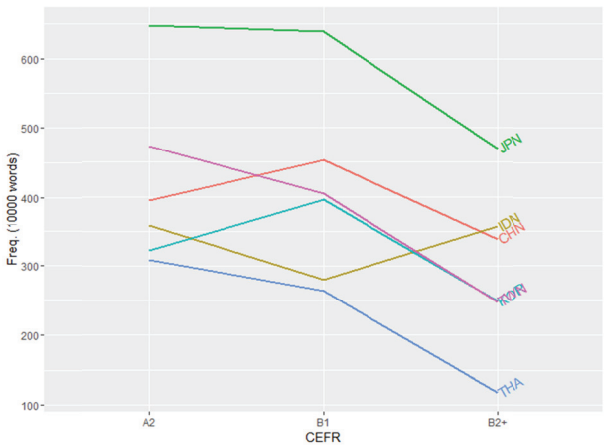

Figure 2. Frequency change patterns of self-mentions

Figure 3 summarizes frequency change patterns for engagement markers. In contrast to their low use of self-mentions, Thai learners use a great number of engagement markers, whereas Japanese learners are the lightest users of this resource. Unlike the four East Asian groups (Chinese, Japanese, Korean, and Taiwanese), the two Southeast Asian groups (Indonesian and Thai) show rises in frequency of engagement markers with increased proficiency.

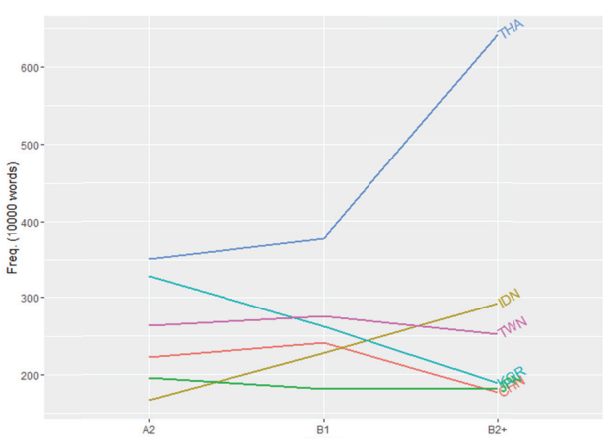

Figure 3. Frequency change patterns of engagement markers

As mentioned above, boosters and hedges are also instrumental in discriminating these L1 groups. Figure 4 illustrates changes in the frequency of boosters across proficiency levels and L1 groups, and shows that boosters

are particularly prominent in Japanese learners, at all developmental stages. While Japanese boosters have a relatively flat frequency pattern across proficiency, other learner groups show decreasing frequency changes between B1 and B2+. In addition, across proficiency levels, Indonesian learners incorporate the metadiscourse resource less frequently than other groups.

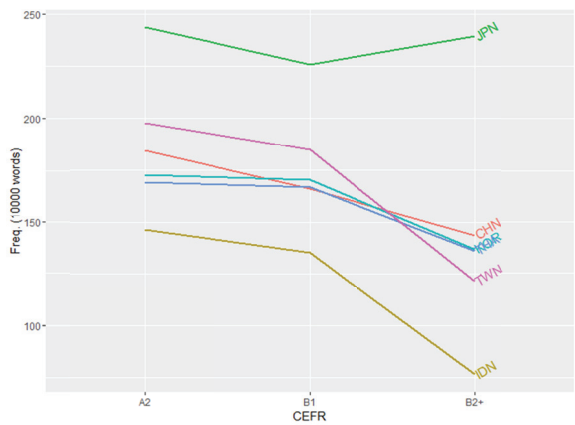

Figure 4. Frequency change patterns of boosters

Figure 5 visualizes the variation in the frequency of hedges across proficiency and L1 backgrounds, indicating marked growth in their frequency between B1 and B2+ Thai and Japanese learners. Chinese and Korean learners display a gradual increase in the metadiscourse device according to developmental stage; Indonesian learners use the fewest hedges (as well as the fewest boosters).

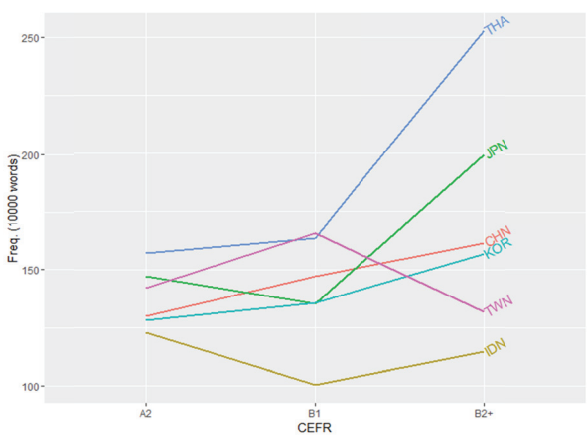

Figure 5. Frequency change patterns of hedges

Finally, one further metadiscourse feature is useful to distinguish learners by L1 background. Figure 6 indicates that frequency change patterns of evidentials, showing a considerable rise between B1 level- and B2+ learners from Chinese and Taiwanese L1 backgrounds. From the perspective of contrastive rhetoric, this high frequency of references to other texts may be ascribable to Chinese practice of rhetoric, in which a writer's claim is often supported with quotations from classical literature (Hinkel, 2002). In contrast, the Thai learners display a dramatic falling pattern, while the other three groups show fluctuations.

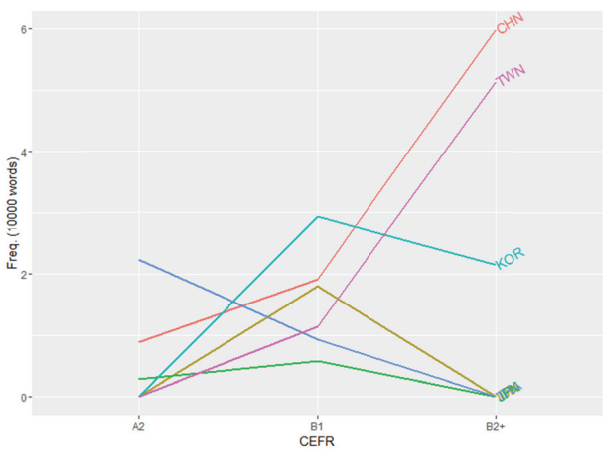

Figure 6. Frequency change patterns of evidentials

5 Conclusion

The purpose of the present study was to describe the developmental patterns of metadiscourse in L2 English essays written by Asian learners from different L1 backgrounds. The results suggest that the six learner groups compared have diverse frequency change patterns of metadiscourse features that also vary across proficiency levels. In particular, Japanese learners' heavy reliance on self-mentions and boosters is remarkably antithetical to Thai learners' preference for engagement markers and hedges. Moreover, B2+ learners in China and Taiwan exhibited greater numbers of evidentials than learners in other groups—a difference attributable not to their lexicogrammatical competence but to an identified L1 rhetorical strategy. As this implies, we should consider the L1 metadiscursive idiosyncrasies of the L1 groups to which L2 learners belong when assessing their language performance. More detailed contrastive analysis of metadiscourse in future research may suggest whether and how learners' rhetoric is affected by cultural factors prevalent in their L1 communities. Nevertheless, the present

study makes an initial contribution to our understanding of the differential nature and characteristics of L2 metadiscourse across L1 groups and proficiency levels.

References

Attarn, A. (2014). Study of metadiscourse in ESP articles: A comparison of English articles written by Iranian and English native speakers. *International Journal of Learning, Teaching and Educational Research*, *5*(1), 63-71.

Conner, U. (1996). Contrastive rhetoric: *Cross-cultural aspects of second-language writing*. Cambridge: Cambridge University Press.

Crismore, A. (1989). *Talking with readers: Metadiscourse as rhetoric act*. New York: Peter Lang.

Glynn, D. (2014). Correspondence analysis: An exploratory technique for identifying usage patterns. In D. Glynn & J. A. Robinson (Eds.), *Corpus methods in cognitive semantics: Quantitative studies in polysemy and synonymy* (pp. 443-485). Amsterdam: John Benjamins.

Greenacre, M. (2016). *Correspondence analysis in practice*. Third edition. Boca Raton: Chapman & Hall.

Harris, Z. (1959). The transformational model of language structure. *Anthropological Linguistics*, *1*(1), 27-29.

Hawkins, J. A., & Filipović, L. (2012). *Criterial features in L2 English: Specifying the reference levels of the Common European Framework*. Cambridge: Cambridge University Press.

Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features*. Mahwah: Lawrence Erlbaum Associates.

Hong, H., & Cao, F. (2014). Interactional metadiscourse in young EFL learner writing: A corpus-based study. *International Journal of Corpus Linguistics*, *19*(2), 201-224.

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

Hyland, K., & Tse, P. (2004). *Metadiscourse in academic writing: A reappraisal. Applied Linguistics*, *25*, 156-177.

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.

Kachru, B. (1992). *The other tongue: English across cultures*. Illinois: University of Illinois Press.

Kobayashi, Y. (2016). 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

Krzeszowski, T. (1990). *Contrasting languages: The scope of contrastive linguistics*. Berlin: Mouton de Gruyter.

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*, pp. 1137-1147.

Nakamura, J. (1995). Text typology and corpus: A critical review of Biber's methodology. *English Corpus Studies*, *2*, 75-90.

Petch-Tyson, S. (1998). Writer/reader visibility in EFL written discourse. In S. Granger (Ed.), *Learner English on computer* (pp. 107-118). London: Longman.

Tan, H., & Eng, W. B. (2014). Metadiscourse use in the persuasive writing of Malaysian undergraduate students. *English Language Teaching, 7*(7), 26-39.

Tono, Y. (2013). Criterial feature extraction using parallel learner corpora and machine learning. In A. Díaz-Negrillo, N. Ballier & P. Thompson (Eds.), *Automatic treatment and analysis of learner corpus data* (pp. 169-204). Amsterdam: John Benjamins.

Vande Kopple, W. (1985). Some explanatory discourse on metadiscourse. *College Composition and Communication*, *36*, 82-93.

Williams, J. (1981). *Style: Ten lessons in clarity and grace*. Boston: Scott Foresman.

Yuichiro Kobayashi
College of Industrial Technology
Nihon University
2-11-1, Shinei, Narashino, Chiba, 275-8576, Japan
Phone: 047-474-2824
E-mail: kobayashi0721@gmail.com

Received: September 24, 2017
Revised: December 4, 2017
Accepted: December 5, 2017