Journal of Modern Languages
Vol. 30, No.2 (2020)
https://doi.org/10.22452/jml.vol30no2.1

Rhetorical preferences in L2 writings: A contrastive analysis of metadiscourse markers

Yuichiro Kobayashi
kobayashi0721@gmail.com
Nihon University, Japan

Abstract

This study investigates variations in the L2 rhetoric across different L1 groups. To do this, the author compared the metadiscourse in English essays composed by 11 Asian learner groups. The multivariate statistical method, namely, the heat map with hierarchical clustering, was applied to clarify differences in metadiscourse among these groups. The results suggest that the groups can be categorized, according to the frequency patterns of metadiscourse markers, into four different clusters: (a) East Asian groups (viz., China, Japan, Korea, and Taiwan), (b) Southeast Asian groups (viz., Indonesia, the Philippines, and Thailand), (c) West Asian groups (viz., Pakistan and the United Arab Emirates), and (d) outer circle users with Chinese backgrounds (viz., Hong Kong and Singapore). In addition, the frequencies of self-mentions and boosters contribute greatly to the clustering of four writer groups.

Keywords**:** learner corpus research, contrastive analysis, metadiscourse markers

1. Introduction

Language corpora have enabled linguists to identify a set of linguistic features that distinguishes various registers. Using computational techniques, Biber (1988) clarified differences in language usage among multiple spoken and written registers. Based on the findings from these corpus-based analyses, Biber, Johansson, Leech, Conrad, and Finegan (1999) published *Longman Grammar of Spoken and Written English*, which exhibits a comprehensive list of grammatical features, characteristic of various registers of English (e.g., conversation, fiction, news, and academic prose) for language users. Moreover, Biber (2006) described a wide range of lexico-grammatical, syntactic, and discourse features in academic and non-academic speech and writing, which university students encounter in their campus life.

Corpus-based studies have also offered various insights to researchers and teachers of English. For example, corpus analyses provide useful information to compile academic word lists, which contain a common core vocabulary across a range of academic disciplines (Coxhead, 2000, 2011). In addition, the comparison of text samples from different academic genres help linguists to identify a variety of genre-specific features, including lexico-grammatical characteristics (Conrad, 2001), textual and discoursal resources (Bruce, 2010), and metadiscourse (Hyland, 2005).

Since the emergence of learner corpus research in the 1990s, researchers have investigated second and foreign language (L2) learners' written and spoken productions, to gain useful knowledge for second language acquisition research and foreign language teaching. The availability of learner corpora enables linguists to analyze a vast amount of descriptive data of interlanguage performances. This has led to contrastive interlanguage analysis, which intends to examine the first language (L1) transfer on L2 production (Osborne, 2015). For instance, Murakami (2013) demonstrated the L1 influence on L2 acquisition order of grammatical morphemes, by analyzing a corpus consisting of over 3,000 essays across seven L1 groups. Similarly, Leacock, Chodorow, Gamon, and Tetreault (2014) showed that the native language of learners significantly affects the likelihood of article error. This was done by contrasting four learner groups whose L1s have articles (viz., French, German, Greek, and Spanish), with three other groups whose L1s do not have articles (viz., Chinese, Japanese, Korean, and Russian).

Learner corpus research has also unveiled the relationship between L2 performance and language family relationship among L1s. For example, Nagata and Whittaker (2013) succeeded

in reconstructing an Indo-European family tree from the syntactic patterns in English texts, by statistically classifying European learners of English from 11 different countries into three branches of the Indo-European family (viz., Italic, Germanic, and Slavic). This proved that European learners' English language usage varies greatly from that of Asian English learners. Moreover, corpus-based studies on L1 transfer have exhibited that L2 learners' texts can be classified using automated lexical and syntactic indices (Jarvis and Crossley, 2012). By identifying lexical and syntactic features that can distinguish different L1 groups, learner corpus research can contribute to the studies on native language identification (Paquot and Jarvis, 2015).

A learner's L1 can strongly affect the discourse as well as vocabulary and syntax in their L2 production. Contrastive rhetorical studies have demonstrated that the discourse characteristics of L2 writing clearly reflects the rhetorical preferences in the learner's L1 (Conner, 1996). To be specific, rhetorical preferences in L1 have an influence on various aspects of L2, such as on paragraph development (Bickner & Peyasantiwong, 1988), discourse development (Reid, 1992), and metadiscourse (Crismore, Markkanen, & Steffensen, 1993). Among these aspects, metadiscourse attracts the most attention in current corpus-based contrastive rhetorical studies.

2. Theoretical framework

2.1. Metadiscourse

In current linguistic research, especially in corpus linguistics and applied linguistics, the most common framework of metadiscourse is undoubtedly Ken Hyland's theory (e.g., Hyland, 1998, 2000, 2005; Hyland and Tse, 2004). His theory is based on a broad range of research findings, accumulated across the history of discourse analysis. In an early study, Harris (1959) coined the term *metadiscourse,* referring to the writer's efforts to guide a reader's perception of a text. Thereafter, Williams (1981) 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). Based on text analysis, he advanced a taxonomy of metadiscourse, consisting of two large categories: interactional resources and interpersonal

resources. Hyland organizes interactive metadiscourse into five major categories: transitions, frame markers, endophoric markers, evidentials, and code glosses. Meanwhile, the interactional metadiscourse is sorted into the following five main categories with specific functions: hedges, boosters, attitude markers, self-mentions, and engagement markers. Functions and examples of each category are 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*
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*

(Hyland 2005, p. 49)

Over the past three decades, Hyland's framework of metadiscourse has been applied when studying various texts, such as company annual reports (Hyland, 1998), undergraduate textbooks (Hyland 2000), and research articles (Hu and Cao, 2015). In addition to these applications for L1 texts, the framework can be developed to explore the metadiscourse in L2 texts, including undergraduate students' writings (Ädel, 2006) and postgraduate dissertations (Hyland and Tse, 2004).

2.2. Learner corpus studies on metadiscourse

Hyland's taxonomy of metadiscourse has had a great impact on contrastive interlanguage analysis, which compares language users from different language backgrounds or language proficiency levels. Using the metadiscourse schema, Hong and Cao (2014) compared English language essays written by Chinese, Spanish, and Polish learners of English, and found significant differences among the three learner groups in the use of interactional metadiscourse markers. Tan and Eng (2014) examined the metadiscourse by novice and advanced English learners from Malaysia in writing, and showed that both learner groups exhibited a greater preference for the use of interactional over interactive resources. In contrast, Attarn (2014) investigated metadiscourse resources of research articles written by Iranian learners and native speakers of English and demonstrated that both writer groups used interactive features more commonly than interactional features. Furthermore, Akbas (2014) contrasted English and Turkish texts, written by Turkish writers, suggesting that Turkish writers were following their native language and culture at certain points, even when they were writing in a foreign language.

In line with the tradition of contrastive interlanguage analysis, comparing 16 learner groups in the International Corpus of Learner's English (ICLE), Kobayashi (2010) found that there are differences in the use of metadiscourse markers in English essays, between European and Asian learners, and between learners whose L1s are Latin-based languages, and those from Slavic countries. Kobayashi (2016a) also analyzed the frequency patterns of metadiscourse markers in six sub-corpora from the International Corpus Network of Asian Learners of English (ICNALE), and clarified the substantial gap in metadiscourse between East Asian learner groups and Southeast Asian learner groups. These two studies succeeded in detecting the patterns that can distinguish different learner groups and classify learners from different language backgrounds. A

further comparison of various learner groups would contribute to the findings of these preceding studies, offering a more global perspective on L2 metadiscourse.

3. Research design

3.1. Purpose of the study

The present study aimed to investigate differences in rhetorical preferences of L2 writings among various L1 groups. More specifically, this study compares the frequencies of metadiscourse markers, identifying rhetorical devices that can be used in distinguishing learner groups. The research questions (RQ) that drive this article are as follows:

RQ 1: How can L2 learners be classified in terms of their use of metadiscourse markers?
RQ 2: Which metadiscourse features can distinguish different learner groups?

By pursuing RQ 1, this study can contribute to contrastive interlanguage analysis that investigates the relationship between L2 performance and language family relationship among L1s. In addition, the answer to RQ 2 can contribute to studies on L1 transfer and native language identification. As many studies show that teaching metadiscourse resources can enhance the writing skills of learners (Asadi, 2018; Cheng and Steffensen, 1996; Taghizadeh and Tajabadi, 2013; Vahid Dastjerdi and Shirzad, 2010), the findings of this study can assist language teachers in instructing their students from different L1 backgrounds.

3.2. Corpus data

The present study draws on the written component of the International Corpus Network of Asian Learners of English (ICNALE-Written), which contains 1.3 million words of 5,600 argumentative essay samples, written by college students in ten Asian countries and regions (viz., China, Hong Kong, Indonesia, Japan, Korea, Pakistan, the Philippines, Singapore, Taiwan, and Thailand) (Ishikawa, 2013), and the additional data set of the corpus, which contains 50,000 words of 200 essay samples, written by students from the United Arab Emirates (ICNALE-UAE). From the viewpoint of World Englishes (Kachru, 1992), the 11 learner groups compared in this study can generally be classified into two groups: the outer circle (viz., Hong Kong,

Pakistan, the Philippines, and Singapore) and the expanding circle (viz., China, Indonesia, Japan, Korea, Taiwan, Thailand, and UAE) of English users. Considering that the use of metadiscourse markers differs for L2 proficiency levels (Kobayashi, 2017), the subset analyzed in this study includes only writers with a B1 CEFR level. The writing conditions were also strictly controlled for the comparison of learner groups. All writers were required to write 200 to 300 words, using Microsoft Word or a similar word processor, within 20 to 40 minutes. The use of dictionaries or other reference tools was prohibited. All essays in the subset were written in response to a single prompt, namely: "It is important for college students to have a part-time job" (Ishikawa, 2013, p. 97). Table 2 shows the size of the 11 learner groups, as compared in this study.

Table 2: Corpus size of the 11 learner groups

	Participants	Words
China (CHN)	337	83,896
Hong Kong (HKG)	82	20,023
Indonesia (IDN)	165	39,085
Japan (JPN)	228	51,778
Korea (KOR)	149	34,126
Pakistan (PAK)	179	42,462
the Philippines (PHL)	187	47,158
Singapore (SIN)	200	50,571
Taiwan (TWN)	148	35,294
Thailand (THA)	279	64,166
the United Arab Emirates (UAE)	40	10,017
Total	1,994	478,576

3.3 Data analysis

The present study counted frequencies of ten functional categories of metadiscourse markers in L2 writings from the 11 learner groups using the Perl program developed by the author. The program can automatically annotate multiple texts and aggregate the raw and relative frequencies of metadiscourse markers and categories, as defined in Hyland (2005). The text sample, annotated using the program was as follows:

(1) **{** I **}_SEM {** disagree **}_ATM** that it is **{** important **}_ATM** for college students to have a part time job. **{** Firstly **}_FRM**, the university workload is very heavy **{** and **}_TRA** involves a lot of self-learning. [...]

Following the frequency counts, this study compared frequencies using the multivariate statistical method, called *heat map with hierarchical clustering* (Kobayashi, 2016b). This method is a powerful technique for visualizing multivariate data, including large frequency tables for corpus analysis, where the graphical representation obtained provides a statistical summary of complex frequency patterns as well as the original frequency information contained within the data. Using the multivariate analysis, linguists can analyze the relationship between certain linguistic features and texts in a more sophisticated fashion than simple statistical tests, such as the analysis of variance or chi-square test. In this paper, the complete linkage method and Euclidean distances (Anderberg, 1973) were used to cluster the metadiscourse categories and learner groups.

As the use of metadiscourse markers depends on the context, it is difficult to automatically annotate them with 100 percent accuracy, especially in L2 texts. Frequent and obvious annotation errors were manually modified; however some cases that were difficult to judge remain. This is a methodological limitation of this study.

4. Results and discussion

4.1. Frequency counts and comparisons of metadiscourse categories

The present study begins through tabulating the frequencies of ten metadiscourse categories in the writings of 11 learner groups. Tables 3 and 4 list the relative frequencies (per 100 words), and the standardized scores of categories used by each writer group. The standardized scores (i.e., *z* scores), being the value in parentheses, indicate the number of standard deviations each frequency value deviates from the mean of the data set. Therefore, positive and negative scores represent frequencies greater and less than the mean, respectively.

Table 3: The relative frequencies (per 100 words) and standardized scores of functional categories for metadiscourse markers (interactive resources)

	TRA	FRM	END	EVI	COD
CHN	4.45	1.05	0.01	0.02	0.55
	(-1.01)	(0.48)	(-0.56)	(0.62)	(-1.76)
HKG	4.32	0.86	0.02	0.01	0.74
	(-1.28)	(-0.47)	(-0.36)	(0.15)	(-0.60)
IDN	4.83	0.81	0.07	0.02	0.96
	(-0.22)	(-0.74)	(0.83)	(0.48)	(0.76)
JPN	4.77	1.33	0.00	0.01	0.95
	(-0.34)	(1.87)	(-0.75)	(-0.89)	(0.67)
KOR	5.11	1.15	0.03	0.03	0.61
	(0.35)	(0.97)	(-0.22)	(1.77)	(-1.37)
PAK	5.36	0.85	0.00	0.03	0.76
	(0.87)	(-0.54)	(2.61)	(1.38)	(-0.46)
PHL	5.25	0.70	0.06	0.01	0.80
	(0.64)	(-1.29)	(0.42)	(-0.34)	(-0.23)
SIN	4.76	0.72	0.01	0.01	0.89
	(-0.36)	(-1.19)	(-0.49)	(-0.87)	(0.30)
THA	4.73	0.87	0.03	0.01	1.01
	(-0.43)	(-0.43)	(-0.08)	(-0.48)	(1.03)
TWN	4.68	1.05	0.01	0.01	0.87
	(-0.53)	(0.46)	(-0.61)	(-0.26)	(0.19)
UAE	6.06	1.13	0.00	0.00	1.08
	(2.30)	(0.87)	(-0.80)	(-1.54)	(1.48)

Table 4: The relative frequencies (per 100 words) and standardized scores of functional categories for metadiscourse markers (interactional resources)

	HED	BOO	ATM	ENG	SEM
CHN	1.05	1.66	0.83	2.84	4.53
	(-0.23)	(0.53)	(-0.39)	(0.26)	(0.97)
HKG	1.43	1.07	0.96	2.04	1.68
	(0.76)	(-0.68)	(0.07)	(-0.87)	(-0.74)
IDN	0.81	1.35	0.89	2.47	2.80
	(-0.86)	(-0.09)	(-0.18)	(-0.26)	(-0.07)
JPN	0.85	2.26	1.43	2.32	6.39
	(-0.75)	(1.76)	(1.67)	(-0.47)	(2.09)
KOR	1.13	1.70	0.93	2.85	3.95
	(-0.01)	(0.62)	(-0.04)	(0.28)	(0.62)
PAK	0.65	0.51	0.33	1.84	1.08
	(-1.28)	(-1.82)	(-2.10)	(-1.15)	(-1.10)
PHL	1.16	1.26	0.69	3.10	2.32
	(0.07)	(-0.28)	(-0.87)	(0.63)	(-0.35)
SIN	2.06	0.92	1.07	1.87	1.55
	(2.43)	(-0.99)	(0.43)	(-1.11)	(-0.82)
THA	1.13	1.67	0.86	4.28	2.65
	(-0.03)	(0.55)	(-0.28)	(2.31)	(-0.16)
TWN	1.31	1.85	1.11	3.12	4.04
	(0.44)	(0.92)	(0.57)	(0.66)	(0.68)
UAE	0.93	1.15	1.27	2.45	1.05
	(-0.55)	(-0.51)	(1.12)	(-0.29)	(-1.12)

Figures 1 and 2 visualize the differences in frequencies of metadiscourse categories among learner groups, using the standardized scores shown in Tables 3 and 4. These figures can provide an intuitive understanding of the metadiscourse characteristics of each learner group. More specifically, learners from the United Arab Emirates use transitions much more frequently than other learner groups, while learners from China and Hong Kong use the feature less frequently

than other groups. Additionally, writers from China, Japan, Korea, Taiwan, and the United Arab Emirates use frame markers more frequently than average, whereas the other six writer groups use the device less frequently than average. However, it is difficult to detect all meaningful patterns underlying the diagrams. To deal with this problem, the author implemented the multivariate statistical method, namely, the heat map with hierarchical clustering, which can offer a more understanding of the notable associations among learner groups and metadiscourse features.

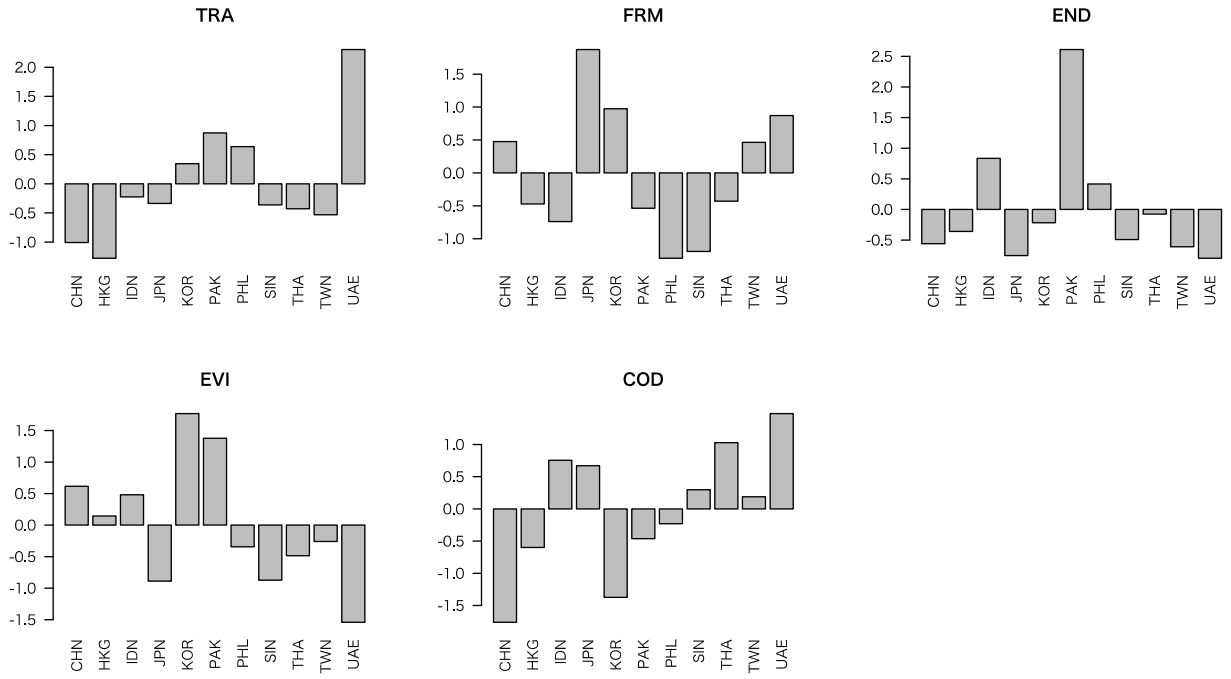

Figure 1: Differences in standardized scores of metadiscourse categories (interactive resources)

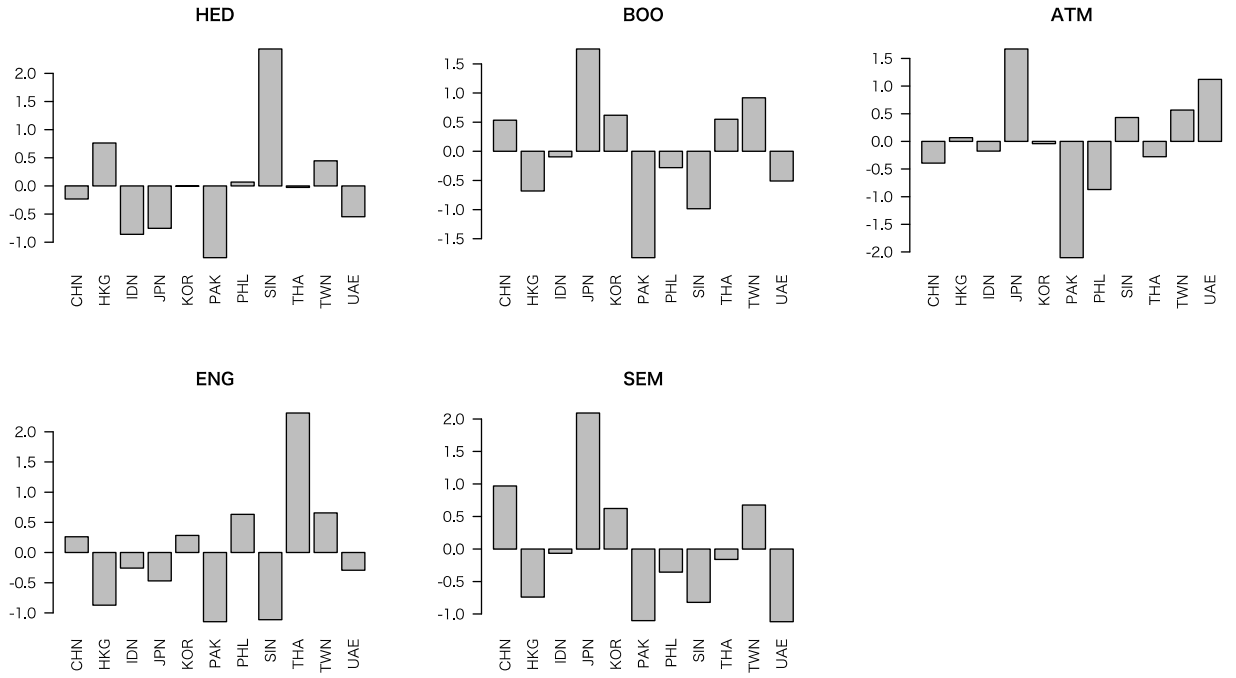

Figure 2: Differences in standardized scores of metadiscourse categories (interactional resources)

4.2. Heat map with hierarchical clustering

The heat map with hierarchical clustering is a useful statistical method for a bird's-eye view of the complex data structure. The method can display the results of text clustering (e.g., learner groups) and linguistic features (e.g., metadiscourse categories). It can also generate the heat map from the permutated frequency table, in two-dimensional space, at the same time. The results of clustering are visualized as tree-like categorizations, where small groups of highly similar items are included within much larger groups of less similar items (Oakes, 1998). In the heat map, a comparison is drawn between texts, with more frequent linguistic features represented by darker cells, and less frequent features denoted by lighter cells.

Figure 3 shows the results of the heat map with hierarchical clustering in this study. The key to interpreting tree-like diagrams in the figure are to focus on the height at which any two objects are joined together—in other words, when the length of horizontal lines corresponds to the size of the difference between two groups or clusters.

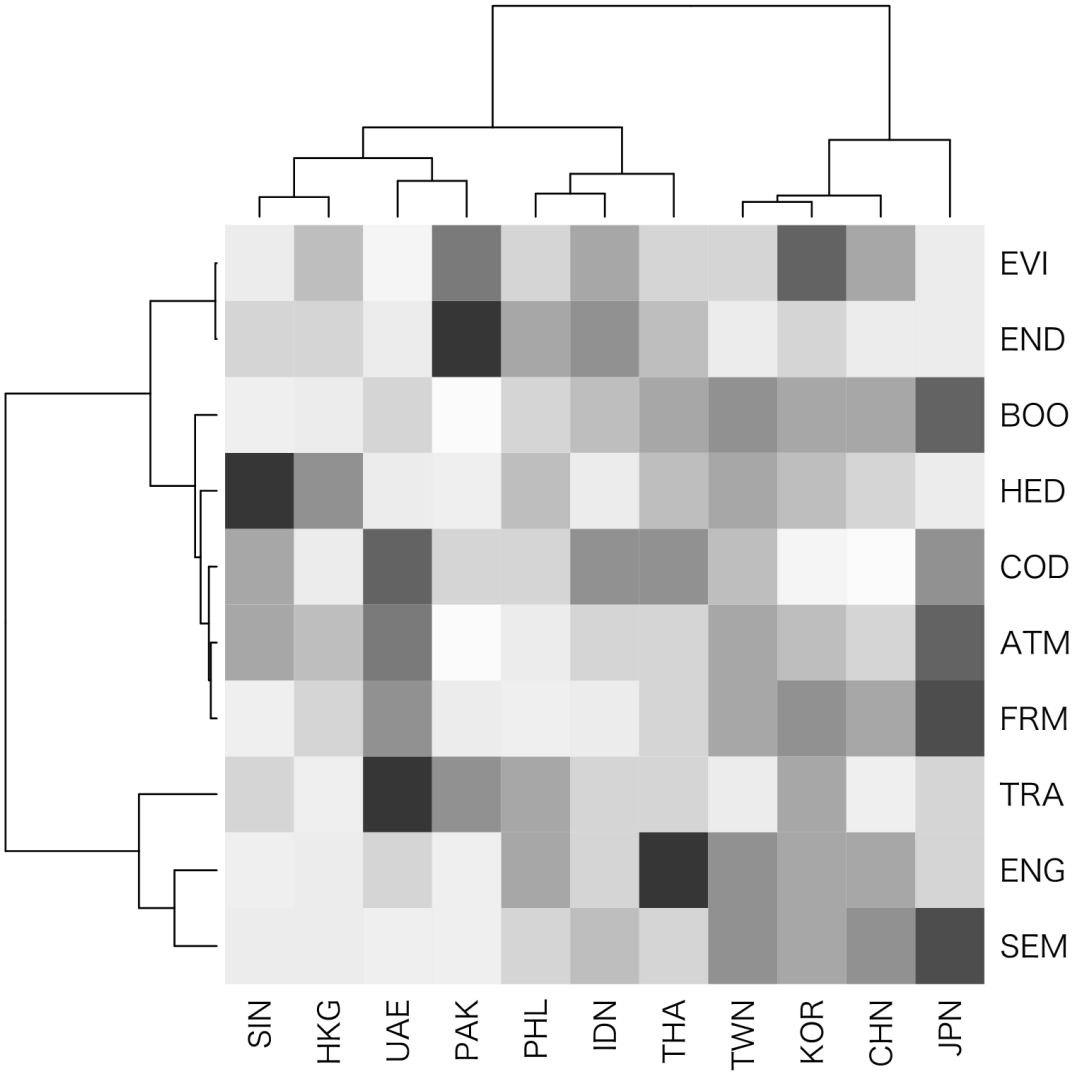

Figure 3: Heat map with hierarchical clustering of learner groups and metadiscourse categories

The clustering results suggest that learner groups can be categorized into four different clusters: (a) East Asian groups (viz., China, Japan, Korea, and Taiwan), (b) Southeast Asian groups (viz., Indonesia, the Philippines, and Thailand), (c) West Asian groups (viz., Pakistan and the United Arab Emirates), and (d) outer circle users with Chinese backgrounds (viz., Hong Kong and Singapore). As reported in Kobayashi (2016a), there is a substantial difference in metadiscourse between East Asian learners and Southeast Asian learners. In addition, learners from West Asian countries are similar to those from Southeast Asian countries, which are relatively close geographically. On the other hand, learners from Hong Kong and Singapore show similarities when compared to those from West Asian countries, but not to East nor

Southeast Asian countries, which are geographically closer. The characteristics of the four clusters of metadiscourse will be examined in detail in the following sections.

4.3. Characteristics of learner groups

4.3.1. East Asian learners

As can be seen in Figure 3, East Asian learners differ largely from other learner groups in terms of metadiscourse. As the heat map in Figure 3 indicates, learners from East Asia used self-mentions, frame markers, and boosters more frequently than learners from other corners of Asia. In particular, Japanese learners are the most frequent users of these metadiscourse features. In general, L2 writers whose L1s are reader-responsible languages tend to use metadiscourse markers less frequently (Kim & Lim, 2013). However, Japanese learners who have reader-responsible language backgrounds were the most frequent users out of the above four categories. The most salient feature of Japanese learners is self-mentions as follows:

> (2) **I** agree with the statement. **I** think there are two points for the reason **I** agree. First, to have a part time job when **we** are college students can be a preparation for being a member of society. **I** have had a part time job for four years. **My** part time job is teaching junior high school students Math and English. **I** learned many important things. For example, how to talk with boss, how to deal with personal information, how to associate with **my** colleague. (JPN)

As Biber, Johanson, Leech, Conrad, and Finegan (1999) illustrate, first person pronouns are linguistic features that characterize spoken language. This means that Japanese learners' writing has a spoken-like nature, which is informally called a *chatty* style (Gilquin & Paquot, 2008).

Another notable feature of East Asian learners is the word *think* as a booster. As Aijmer (2002) points out, the frequent usage of the word is greatly influenced by spoken language.

> (3) **I think** every college student should have a least one part time job in his college life. (TWN)
> (4) So in my opinion, **I think** it is important for college students to have a part time job. (CHN)

Furthermore, East Asian writers prefer to explicitly show logical structure in their essays using frame markers. The frequent use of this metadiscourse feature may be a consequence of "superficial attention" (Intaraprawat & Steffensen, 1995, p. 271) to logical forms, and results in "artificial, mechanical prose" (Zamel, 1983, p. 27). Frequent usage of connectors does not necessarily improve the cohesive quality of a text, since semantic relations between main clauses do not have to be explicitly marked (Altenberg & Tapper, 1998).

(5) **First**, it is their personal reason. I felt university student is so much time. But they didn't use their time. They consume time on play. [...] **Second**, university student is spending a lot of money. [...] **Third**, it is opportunity to Social experience in advance. (KOR)

As shown in Figure 3, East Asian groups uses many metadiscourse categories more frequently than other learner groups. Considering that high-proficient writers efficiently use fewer metadiscourse markers than low-proficient writers (Bax, Nakamura, & Waller, 2019), their written compositions may leave room for improvement from the viewpoint of academic writing.

4.3.2. Southeast Asian learners

As shown in Figure 3, in contrast to East Asian learners, Southeast Asian learners have fewer distinctive metadiscourse features in their writings. The cells in the columns assigned to Southeast Asian writers are generally denoted by lighter colors, which means that most of the metadiscourse categories are used less frequently than those of other learner groups. Interestingly enough, this cluster includes the outer circle users of English (viz., the Philippines) as well as the expanding circle users (viz., Indonesia and Thailand). The clustering result suggests that L2 metadiscourse can be affected by the geographic (and possibly cultural) proximity more strongly than the difference between outer circle and expanding circle environments.

An exceptionally salient feature of Southeast Asian groups is the significant use of engagement markers, especially second person pronouns, in Thai learners' essays. Engagement markers involve readers as discourse participants, whereas self-mentions foreground the presence of the author in the text. Regarding the use of personal pronouns, Thai learners are

opposed to Japanese learners, who use first person pronouns in a *chatty* style. However, the use of second person pronouns is also part of the linguistic features distinctive of an informal writing style (Petch-Tyson, 1998). Thai learners repeatedly used second person pronouns in combination with interrogatives, which are characteristic of spoken language.

(6) Do **you** think it is necessary for college students? (THA)
(7) Do **you** want some further supporting reasons? (THA)

Additionally, a minor characteristic of Southeast Asian groups are code glosses in essays, common among Thai and Indonesian learners. They provide supporting examples using phrases like *such as* and *for example*.

(8) If they don't have a part time job, they may do other bad things **such as** using drugs, fighting each other, playing online games and wasting their time. (THA)
(9) **For example**, I used to work with McDonald as a cashier when I was a student in high school in Bangkok. (THA)
(10) Students can use the facilities of the company, **such as** computers or the Internet service [...] (IDN)
(11) **For example**, we can be a waiter or a waitress in a restaurant or maybe a private teacher for elementary students and junior high students. (IDN)

As these expressions are characteristic of written language (Biber et al., 1999), Thai learners who frequently use second person pronouns displayed both written and spoken linguistic features in their written discourse. However, the consistent use of a cohesive tone is required for successful writing (Petch-Tyson, 1998).

4.3.3. West Asian learners

Two learner groups from West Asia, namely Pakistan and the United Arab Emirates, shared a low frequency of four interactional resources (viz., hedges, boosters, engagement markers, and self-mentions). Furthermore, previous research has found that, statistically, Persian writers underuse these resources in comparison to native English writers (Karimi, Maleki, and Farnia,

2017). Therefore, the low frequency of interactional resources can be regarded as a similarity among West Asian learners' English writings.

On the other hand, as shown in Figure 1, learners from the United Arab Emirates and Pakistan used transitions more frequently than those from other parts of Asia. They utilized multi-word connectors, including *as a result*, *even though*, and *in addition*, as well as one-word conjunctions, such as *because*, *however*, and *therefore*.

(12) **In addition**, students will learn how to deal with customers or how to solve the problems. (UAE)

(13) It is important for university students to have a part-time job **because** they will have more experience and will be independent. (UAE)

In case of the positions of conjunctions, these groups sometimes redundantly used multiple transitions in a single sentence as follows:

(14) **Also** students do not have enough time to relax far away from studying and working environment **and also** they do not have time to spend with family or even to have friends. (UAE)

In connection with the redundant use, most transitions occurred in the middle of sentences in their essays. Learners from the United Arab Emirates are, in this regard, sharply contrasted to East Asian learners, including Korean and Japanese learners, who prefer to use conjunctions in the sentence-initial positions (Kobayashi, 2016a). On top of this, writers from the United Arab Emirates frequently applied code glosses, attitude markers, and frame markers, in addition to transitions.

Figures 1 and 3 indicate that the standardized scores of evidentials and endophoric markers in Pakistani learners' essays are relatively high. However, because of its low frequency, as shown in Table 3, these two categories will not be investigated further in this paper.

4.3.4. Outer circle users with Chinese backgrounds

The most prominent feature of outer circle users with Chinese backgrounds was the high frequency of hedges, which are one of the most significant rhetorical devices in academic writing. Good writers can use hedges to strengthen their argument as well as weaken the claim in their discourse (Meyer, 1997).

> (15) First, a student who works while studying **may** suffer poorer grades, due to less study time. Second, he **would** have lesser time to engage in extra curricular activities, which are vital for his future job prospects. (SIN)
>
> (16) Students **could** also learn to be independent by earning some money as their allowance rather than taking from their parents. I **would** strongly recommend university students to have a part time job. (HKG)

As Figures 2 and 3 indicate, the most frequent users of hedges are the writers from Singapore, and the second most frequent users are those from Hong Kong. Furthermore, the third and fifth most frequent users are Taiwanese and Chinese writers, respectively. In light of this ranking, the high frequency of hedges is possibly affected by their Chinese backgrounds rather than by the distinction between outer and expanding circles of English language. In Chinese-speaking countries, hedging devices play an important role in arguments, because they can concurrently perform several discourse functions due to their ambiguous nature (Hinkel, 2002).

Hedging is one of the metadiscourse features that has been most extensively researched across various linguistic fields. It is also known to be linked to multiple factors, such as writer's language and cultural background (Itakura, 2013; Tan and Chan, 2008; Yang, 2013), academic disciplines (Adbi, 2002; Hyland and Tse, 2004), and publishing contexts (Hu and Cao, 2011). Therefore, the frequency and usage of hedging devices can change, depending on the writer's proficiency, writing tasks, and essay topics. However, the results of this study could serve as a benchmark for future research on hedges in L2 writings because the writing conditions and learners' proficiency levels were rigorously controlled for the comparison.

Another noteworthy aspect of writings composed by learners from Singapore and Hong Kong was the infrequent use of self-mentions and boosters. Figure 3 clearly shows that the frequencies of these two metadiscourse resources most significantly contribute to the clustering

of writer groups. To put it more plainly, frequencies gradually decreased from the rightmost group (viz., East Asian learners) to the leftmost group (viz., outer circle users with Chinese backgrounds) within the figure. Considering this pattern, together with the frequency of hedges mentioned above, writers from Singapore and Hong Kong are diametrically opposed to East Asian learners in the viewpoint of the degree of writer's certainty in proposition and presence within the text.

5. Conclusion

The aim of the present study was to investigate differences of rhetorical preferences in L2 writings, and identify metadiscourse features that can distinguish different L1 groups. The results indicate that learner groups can be classified according to the frequency patterns of metadiscourse markers into four different clusters: (a) East Asian groups (viz., China, Japan, Korea, and Taiwan), (b) Southeast Asian groups (viz., Indonesia, the Philippines, and Thailand), (c) West Asian groups (viz., Pakistan and the United Arab Emirates), and (d) outer circle users with Chinese backgrounds (viz., Hong Kong and Singapore). Each writer group displayed the specific characteristics of metadiscourse, which offers suggestions for improving L2 learners' writings. Specifically, L2 writers can enhance the quality of their essays by utilizing their less frequently used metadiscourse markers and by replacing the linguistic features they repeatedly use with alternative expressions.

In addition to the methodological problem mentioned above, this study has some other notable limitations. First, the target learners were limited to Asian learners at a particular proficiency level. Second, this study focused on metadiscourse markers, and hence has not considered other aspects of learners' performances, including the proper use and misuse of vocabulary and grammar. Third, the influence of teaching materials and instruction methods should be considered in order to understand variation in L2 metadiscourse across different L1 groups on a deeper level. However, despite these limitations, the findings of this study can help language teachers instruct their students from various L1 backgrounds as well as assist language learners in becoming aware of common rhetorical patterns in academic writing.

Acknowledgments

Earlier versions of the present study were reported at the CamTESOL 2019 Conference, and the Language in Focus 2019 Conference. This work was supported by Grants-in-Aid for Scientific Research Grant Numbers 17K13511.

References

Abdi, R. (2002). Interpersonal metadiscourse: An indicator of interaction and identity. *Discourse Studies, 4*(2), 139–145.

Ädel, A. (2006). *Metadiscourse in L1 and L2 English.* Amsterdam: John Benjamins.

Aijmer, K. (2002). Modality in advanced Swedish learners' written interlanguage. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 55–76). Amsterdam: John Benjamins.

Akbas, E. (2014). Are they discussing in the same way? Interactional metadiscourse in Turkish writers' texts. In A. Łyda, & K. Warchał (Eds.), *Occupying niches: Interculturality, cross-culturality and aculturality in academic research* (pp. 119–133). London: Springer.

Altenberg, B., & Tapper, M. (1998). The use of adverbial connectors in advanced Swedish learners' written English. In S. Granger (Ed.), *Learner English on computer* (pp. 80–93). London: Longman.

Anderberg, M. R. (1973). *Cluster analysis for applications.* New York: Academic Press.

Asadi, A. (2018). Enhancing writing skills of English learners through metadiscourse resources. *MEXTESOL Journal, 42*(3), 1–19.

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.

Bax, S., Nakamura, F., & Waller, D. (2019). Researching L2 writers' use of metadiscourse markers at intermediate and advanced levels. *System, 83,* 79–95.

Biber, D. (1988). *Variation across speech and writing.* Cambridge: Cambridge University Press.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers.* Amsterdam: John Benjamins.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English.* London: Longman.

Bickner, R., & Peyasantiwong, P. (1988). Cultural variation in reflective writing. In A. C. Purves (Ed.), *Writing across languages and cultures: Issues in contrastive rhetoric* (pp. 160–176). Newbury Park: Sage.

Bruce, I. (2010). Textual and discoursal resources used in the essay genre in sociology and English. *Journal of English for Academic Purposes, 9*(3), 153–166.

Cheng, X., & Steffensen, M. S. (1996). Metadiscourse: A technique for improving student writing. *Research in the Teaching of English, 30*(2), 149–181.

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

Conrad, S. (2001). Variation among disciplinary texts: A comparison of textbooks and journal articles in biology and history. In S. Conrad & D. Biber (Eds.), *Multi-dimensional Studies of Register Variation in English* (pp. 94–107). Harlow: Pearson Education.

Coxhead, A. (2000). A new academic wordlist. *TESOL Quarterly, 34*(2), 213–238.

Coxhead, A. (2011). The academic word list 10 years on: Research and teaching implications. *TESOL Quarterly, 45*(2), 355–362.

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

Crismore, A., Markkanen, R., & Steffensen, M. (1993). Metadiscourse in persuasive writing: A study of texts written by American and Finnish university students. *Written Communication, 10(1),* 39–71.

Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction, 1*(1), 41–61.

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

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.

Hu, G., & Cao, F. (2011). Hedging and boosting in abstracts of applied linguistics articles: A comparative study of English-and Chinese-medium journals. *Journal of pragmatics, 43*(11), 2795–2809.

Hu, G., & Cao, F. (2015). Disciplinary and paradigmatic influences on interactional metadiscourse in research articles. *English for Specific Purposes, 39,* 12–25.

Hyland, K. (1998). Exploring corporate rhetoric: Metadiscourse in the CEO's letter. *The Journal of Business Communication, 35,* 224–246.

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing.* London: Longman.

Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics, 25, 1*56–177.

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

Intaraprawat, P., & Steffensen, M. (1995). The use of metadiscourse in good and poor ESL essays. *Journal of Second Language Writing, 4,* 253–272.

Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world, 1* (pp. 91–118). Kobe: Kobe University.

Itakura, H. (2013). Hedging praise in English and Japanese book reviews. *Journal of Pragmatics, 45*(1), 131–148.

Jarvis, S., & Crossley, S. A. (2012). *Approaching language transfer through text classification: Explorations in the detection-based approach.* New York: Multilingual Matters.

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

Karimi, K., Maleki, M., & Farnia, N. (2017). Metadiscourse markers in the abstract sections of Persian and English law articles. *International Journal of Foreign Language Teaching & Research, 5*(18), 69–83.

Kim, L. C., & Lim, J. M. (2013). Metadiscourse in English and Chinese research article introductions. *Discourse Studies, 15*(2), 129–146.

Kobayashi, Y. (2010). A multivariate discourse analysis of non-native Englishes. *Proceedings of Jinmoncom Symposium 2010* (pp. 41–48). Tokyo: IPSJ SIG Computers and the Humanities.

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.

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

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

Meyer, P. G. (1997). Hedging strategies in written academic discourse: Strengthening the argument by weakening the claim. In R. Markkanen & H. Shroder (Eds.), *Hedging and discourse: Approaches to the analysis of a pragmatic phenomenon in academic texts* (pp. 21–41). Berlin: Walter de Gruyter.

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.

Oakes, M. P. (1998). *Statistics for corpus linguistics.* Edinburgh: Edinburgh University Press.

Osborne, J. (2015). Transfer and learner corpus research. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 333–356). Cambridge: Cambridge University Press.

Paquot, M., and Jarvis, S. (2015). Learner corpora and native language identification. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 605–628). Cambridge: Cambridge University Press.

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

Reid, J. (1992). A computer text analysis of four cohesion devices in English discourse by native and nonnative writers. *Journal of Second Language Writing, 1*(2), 79–108.

Taghizadeh, M. & Tajabadi, F. (2013). Metadiscourse in essay writing: An EFL case. *International Research Journal of Applied and Basic Sciences, 4*(7), 1658–1662.

Tan, H., & Chan, S. H. (2008). The writing of research articles: How do native and non-native writers differ in the use of hedges? In R. Jacobsen (Ed.), *The pulse of a Malaysian university: Ethno and sociolinguistic issues and the TESOL dimension* (pp. 163–178). New York: Peter Lang.

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

Vahid Dastjerdi, H., & Shirzad, M. (2010). The impact of explicit instruction of metadiscourse markers on EFL learners' writing performance. *The Journal of Teaching Language Skills, 29*(2), 155–174.

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.

Yang, Y. (2013). Exploring linguistic and cultural variations in the use of hedges in English and Chinese scientific discourse. *Journal of Pragmatics, 50*(1), 23–36.

Zamel, V. (1983). Teaching those missing links in writing. *ELT Journal, 37,* 22–29.

About the Author

Yuichiro Kobayashi is a full-time lecturer of the College of Industrial Engineering at Nihon University, Japan. He received his doctoral degree in language and culture from Osaka University. His research interests are in corpus linguistics and natural language processing. He is currently working on contrastive interlanguage analysis using learner corpora.