Journal of Pan-Pacific Association of Applied Linguistics, 25(2), 87-104.
https://doi.org/10.25256/PAAL.25.2.5

Identifying L2 Developmental Indices while Controlling for L1 Effects: A Multilevel Ordinal Logistic Regression Analysis[*]

Yuichiro Kobayashi[**]

Kobayashi, Y. (2021). Identifying L2 developmental indices while controlling for L1 effects: A multilevel ordinal logistic regression analysis. ***Journal of Pan-Pacific Association of Applied Linguistics, 25*****(2), 87-104.**

This study aimed to identify second language (L2) developmental indices while controlling for the effects of first language (L1). More specifically, this study investigated the differences in the use of metadiscourse markers among learners from different L1 backgrounds. The following research questions were explored: (1) Which metadiscourse markers can be used as developmental indices to predict learners' proficiency levels? (2) How strongly do learners' L1s impact their L2 development? To answer these questions, multilevel ordinal logistic regression analysis was performed comparing the metadiscourse in English essays written by three Asian learner groups (i.e., Japanese, Thai, and Taiwanese). The results suggest that the frequencies of transitions and self-mentions are the best predictors of language development of metadiscourse categories. Additionally, this study showed that the multilevel regression model can measure the significance of each metadiscourse category more accurately than the single-level model. Therefore, we must consider the influence of learners' L1s for precise detection of the linguistic features that predict their proficiency levels. Through this application of multilevel ordinal regression analysis to learner corpus research, this study illustrated the effectiveness of multilevel analysis for tracking the language acquisition process.

Keywords: learner corpus research, developmental indices, L1 effects, metadiscourse markers, multilevel analysis

---

[*] This work was supported by Grants-in-Aid for Scientific Research Grant Number 21K00660. A part of this study was reported at the Japan Association for Language Education and Technology (LET), Kansai Chapter, Methodology Special Interest Group (SIG) on July 10th, 2021.

[**] **Yuichiro Kobayashi**, Tenured Lecturer, College of Industrial Technology, Nihon University

1 Introduction

The advent of computer learner corpora has enabled language researchers to track the language acquisition process (Meunier, 2015). Additionally, the development of natural language processing technology has enabled the automatic annotation of a wide range of linguistic features, including parts of speech, syntax, semantics, and discourse (Newman & Cox, 2020), and automated error detection techniques can identify several types of grammatical errors in learner corpora (Leacock et al., 2014). Such rich linguistic information allows us to seek developmental indices that can predict learners' proficiency levels from various angles (Verspoor et al., 2021). Finding the best predictors of the quality of learner language has been a challenge for second language acquisition (SLA) studies.

Recently, learner corpus research (LCR) has begun to recognize the effectiveness of multifactorial regression analysis through a series of enlightening papers by Stefan Th. Gries (Gries, 2015a; Gries & Adelman, 2014; Gries & Deshors, 2014, 2021; Gries & Wulff, 2013; Wulff & Gries, 2015). This method can consider multiple predictors (e.g., the frequencies of linguistic features, learners' linguistic backgrounds) in an analysis, explore interactions among variables, and generate predictions of how a response (e.g., proficiency levels, test scores) will behave. Furthermore, it can evaluate the strength of association between the predictors and response in the context of statistical hypothesis testing (via *t*-tests, Wald tests). Thus, it allows us to investigate multiple factors involved in language acquisition with greater sophistication than traditional SLA studies based on mono-factorial tests.

Another new methodological trend in LCR is the use of multilevel models, also called mixed-effect models, random effects models, hierarchical linear models, or nested data models (Gries, 2015b, 2021). Multilevel models are particularly suitable for research designs in which data have a nested structure (e.g., having individuals in different learner groups). These models have been used in educational and social research for several decades (Goldstein, 1987; O'Connell & McCoach, 2008), and are being increasingly used as an alternative to conventional analyses in SLA research (Cunnings, 2012). By performing multilevel analysis, we can detect developmental indices of the second language (L2) while taking various factors, such as the influence of first language (L1) or task differences, into account.

This study aimed to identify L2 developmental indices by controlling L1 effects using multilevel regression analysis. More specifically, this study examined the differences in the use of metadiscourse markers among learners from different L1 backgrounds. The following research questions were explored: (1) Which metadiscourse markers can be used as developmental indices to predict learners' proficiency levels? (2) How strongly do learners' L1s impact their L2 development? While attempting to address these research questions, this study also explores the effectiveness of multilevel analysis that can ascertain the strength of L1 effects more accurately than conventional methods in LCR.

2 Theoretical Framework

2.1 Contrastive interlanguage analysis

The most widely used framework in LCR is contrastive interlanguage analysis (CIA), which mainly compares L2 learners from different L1 backgrounds (Granger, 2015). Its comparative design has enabled researchers to uncover a broad range of linguistic features distinctive of interlanguage and evaluate their degree of generalizability across learner groups. As an example of CIA, Murakami (2013) compared the order of accuracy of L2 English grammatical morphemes in seven L1 groups (viz., Japanese, Korean, Spanish, Russian, Turkish, German, and French), and reported that the groups whose L1s do not obligatorily mark the morpheme tend to have a lower level of accuracy with respect to the morphemes than those whose L1s mark it. Nagata and Whittaker (2013) also compared the syntactic patterns in English texts written by learners from 11 European countries (viz., Bulgarian, Czech, Dutch, French, German, Italian, Norwegian, Polish, Russian, Spanish, and Swedish), and concluded that these learner groups can be statistically classified into three branches of the Indo-European family (viz., Italic, Germanic, and Slavic).

The CIA framework can be used for comparing L2 learners at different proficiency levels. Investigating advanced learners as well as novice learners, CIA can address the gap in SLA research focusing on the early stages of acquisition (Granger, 2015). Specifically, Abe (2014) identified a set of linguistic features that can differentiate among L2 oral proficiency groups by checking the frequency change patterns of 58 linguistic features across seven oral proficiency levels. Abe (2019) also investigated the accuracy rates of lexical and grammatical features used by Japanese learners with varying levels of proficiency in English and showed that the acquisition process is different for written and spoken languages.

Moreover, an increasing body of LCR has analyzed L1 backgrounds and proficiency levels simultaneously. For example, Römer and Berger (2019) examined the frequency patterns of English verb-argument constructions in writing produced by German and Spanish learners at different proficiency levels using basic statistics and visualization. Ionin and Díez-Bedmar (2021) also compared the use of articles by Russian and Spanish learners of English with B1 and B2 CEFR (Common European Framework of Reference for Languages) levels through the Kruskal-Wallis test followed by Mann-Whitney *U* tests with Bonferroni correction. However, multilevel analysis can more efficiently analyze such nested data (i.e., learners with different L1 backgrounds and proficiency levels), and more precisely estimate the effects of L1s and proficiency on learners' performance, as explained in the next section.

2.2 Multilevel analysis

Multilevel analysis is a methodology for the analysis of data with complex patterns of variability, focusing on nested sources of such variability (e.g., students in schools, employees in companies, residents in cities). This method is highly suitable for corpus-linguistic studies because corpora often have nested structures (Gries, 2015b, 2021). Specifically, many learner corpora have a hierarchical structure in which learners are nested into files, which are nested into proficiency levels, which are nested into L1 backgrounds, which in turn are nested into production modes (e.g., writing, speaking). However, most corpus-linguistic studies routinely ignore such nested structure, as well as the usefulness of multilevel analysis. Ignoring multilevel data structures leads to multiple potential errors (Heck & Thomas, 2020; Snijders & Bosker, 2012). First, it can result in a shift in meaning. A variable aggregated at the macro-level refers to the macro-units (e.g., L1 backgrounds), not directly to the micro-units (e.g., individual learners). Second, disregard of nested structures can lead to the ecological fallacy. A correlation between macro-level variables cannot be used to make assertions about micro-level relations. Moreover, unawareness of the hierarchical structure leads to statistically incorrect estimates. The aggregation of the micro-level units to the macro-level results in an overestimation of the degrees of freedom and underestimation of standard errors of the data. Consequently, it increases the likelihood of making Type I errors (i.e., falsely rejecting the null hypothesis in statistical tests). Therefore, when analyzing data with nested structure, we must precisely measure the effects of micro-level variables by statistically controlling (excluding) those of macro-level variables.

With the awareness of the above issues, several corpus linguists have conducted pioneering studies with multilevel analyses. For instance, Gries and Deshors (2015) exemplified the advantages of hierarchical mixed-effects modeling by analyzing 17 lexical verbs in the dative alternation in the speech and writing of five English learner groups. Murakami (2016) also showed the usefulness of generalized linear mixed-effects models and generalized additive mixed models for tracking the development of the L2 accuracy of English grammatical morphemes. Additionally, through a series of linear mixed-effects models, Kyle et al. (2020) examined the developmental trajectories of English learners across two academic years with regard to syntactic complexity and sophistication. Furthermore, Paquot et al. (2021) investigated the development of phraseological complexity in the writing of French learners of English with a mixed-effect model that controlled for the potential effects of topics and prompts. However, the numbers and types of multilevel models used in corpus-linguistic research are limited. In particular, there are few examples of multilevel modeling of multiple categorical response variables.

Since categorical variables, such as L1 backgrounds and proficiency levels, are frequently compared in LCR, categorical regression models can be a powerful tool to explore L1 effects on L2 performance. For example, multinomial logit models can be used to compare three or more nominal variables, including L1 backgrounds or writing/speaking topics, and ordinal logit models can be employed to analyze three or more ordered variables, including proficiency levels and years of study. Furthermore, these categorical models can be extended to analyze data with multilevel structures (Agresti, 2012, 2018). Thus, multilevel categorical models are very helpful in statistically controlling for the potential effects of various factors, such as L1s, proficiency levels, and writing/speaking topics, on learners' language use (Michel et al., 2019; Saito et al., 2020).

3 Research Design

3.1 Corpus data

The present study draws on the written component of the International Corpus Network of Asian Learners of English (ICNALE; Ishikawa, 2011), which contains 1.3 million words from argumentative essays written by 2,800 college students in Asia. The writing conditions were rigorously controlled for the comparison of learners from different backgrounds. Writers were required to use word processing software to compose essays without using dictionaries or other reference materials, and were asked to use an electronic spell-checker before submitting the essay. The essays were assigned as extra homework for their English classes. All compositions in the subset were written in response to a single topic: "It is important for college students to have a part time job." The data analyzed in the present study made up a subset of this corpus comprising the written compositions of 1,000 learners in Japan (JPN), Thailand (THA), and Taiwan (TWN). Their Common European Framework of Reference (CEFR) levels were assessed as A2, B1.1, B1.2, and B2. Table 1 provides an overview of the data analyzed in this study. In this table, the four levels of the CEFR are translated into numbers from 1 to 4, with 1 being the most novice and 4 being the most advanced.

Table 1. Overview of the Data Analyzed in This Study

CEFR level	Numbers	JPN	THA	TWN	Total
1	learners (words)	154 (34,959)	119 (26,866)	29 (6,491)	302 (68,316)
2	learners (words)	179 (40,463)	179 (40,899)	87 (20,658)	445 (102,020)
3	learners (words)	49 (11,315)	100 (23,267)	61 (14,636)	210 (49,218)
4	learners (words)	18 (4,281)	2 (514)	23 (5,856)	43 (10,651)
total	learners (words)	400 (91,081)	400 (91,546)	200 (47,641)	1,000 (230,205)

3.2 Linguistic features

This study applied Hyland's list of metadiscourse markers (Hyland, 2005) to explore L1 effects on L2 writing. Examining a broad range of research findings accumulated in the past years with respect to metadiscourse analysis, Hyland (2005) defined metadiscourse as "a 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 a member of a particular community" (p. 37). Based on this definition, he advanced a taxonomy of metadiscourse, consisting of two large categories: interactional and interpersonal resources. Interactive metadiscourse is organized into five main categories with specific functions: transitions, frame markers, endophoric markers, evidentials, and code glosses. Meanwhile, interactional metadiscourse is sorted into the following five major categories: hedges, boosters, attitude markers, engagement markers, and self-mentions. The functions and examples of each category are summarized in Table 2.

Table 2. Hyland's Taxonomy of Metadiscourse Markers

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)

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

Furthermore, the frequency patterns of metadiscourse markers can be used to compare learners' L1 backgrounds and proficiency levels (Kobayashi, 2016, 2017, 2020).

3.3 Data analysis

The present study computed the frequencies of 10 functional categories of metadiscourse markers in L2 writings from the three learner groups using a Perl program developed by the author. The program can automatically annotate multiple texts and aggregate the raw and relative frequencies of metadiscourse categories, as defined by Hyland (2005). The obvious annotation errors were manually modified, although some cases were difficult to judge. This was a technical limitation of the present study.

Following the frequency counts, single-level and multilevel ordinal logistic regression analysis were conducted to detect the metadiscourse markers that can distinguish between learners' proficiency levels. Ordinal logistic regression analysis is a typical method for analyzing three or more ordered categorical responses (Agresti, 2010), and it is more robust for data containing outliers than other typical methods, such as ordinal probit regression analysis (Liao, 1994). In this study, the effectiveness of multilevel ordinal logistic regression analysis has been demonstrated by comparing it with single-level analysis. To the best of the author's knowledge, there are no learner corpus studies using multilevel ordinal logit models.

All statistical analyses in this study were conducted using R, a free software environment for statistical computing and graphics (R Core Team, 2020). The *ordinal* package (Christensen, 2018) was used to perform ordinal logistic regression analyses, and the *effects* package (Fox & Weisberg, 2020) was used to draw the effect plots. For other R techniques related to multilevel analysis, this study mainly referred to Gałecki and Burzykowski (2015), Finch et al. (2019), and Roback and Legler (2020).

4 Results and Discussion

4.1 Frequency counts

This study began by counting the frequencies of metadiscourse categories in the writings of all individual learners. Figure 1 shows the differences in the relative frequencies (per 100 words) of the nine categories among the three learner groups. Endophoric markers did not occur in the data analyzed in this study.

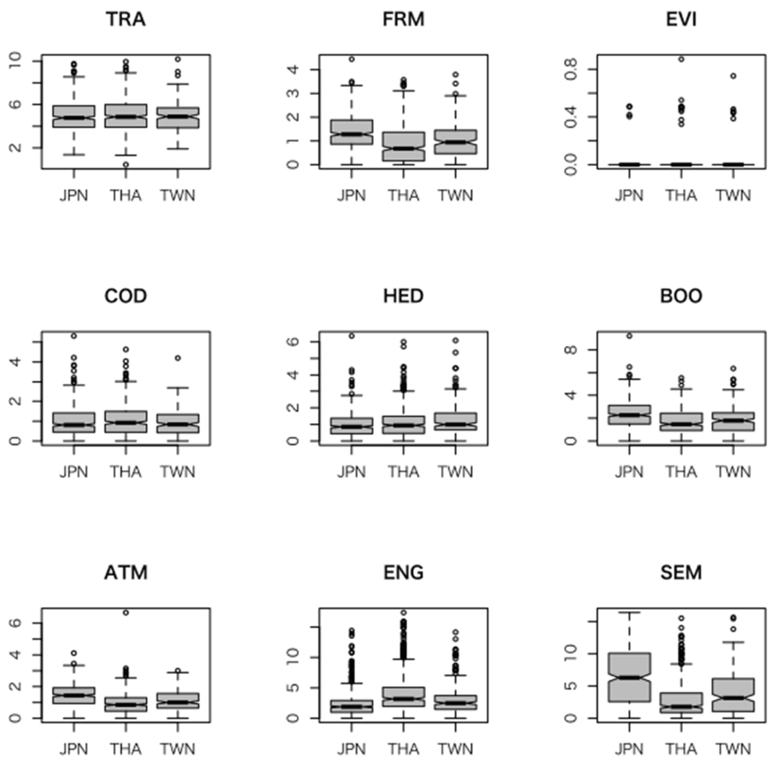

Figure 1. Differences in the frequencies of metadiscourse categories among the three learner groups

These box plots show the differences in several metadiscourse categories among learner groups. However, they refer only to the group level, which aggregates learner-level information. Therefore, regression models utilizing learner-level information are more appropriate for investigating L1 effects on learners' performance.

4.2 Ordinal logistic regression analysis

This study applied ordinal logistic regression analysis to predict learners' proficiency levels (i.e., response variables) using the frequencies of metadiscourse categories (i.e., predictor variables). Given the potential risk of multicollinearity among predictor variables, a stepwise variable selection using Akaike's information criterion (AIC) was implemented before performing ordinal logistic regression analysis. The presence of multicollinearity results in smaller *t*-values, larger coefficients of determination, and regression coefficients with positive or negative signs opposite their expected values. All these adverse effects can lead to incorrect estimation. As a result of stepwise variable selection, three variables were selected: transitions, self-mentions, and boosters.

Table 3 lists the intraclass correlation coefficients (ICCs) for the three variables selected by the stepwise method. ICCs greater than 0.05 suggest that multilevel analysis should be applied due to the high similarity within groups (Hox et al., 2018). As the ICCs of self-mentions and boosters were 0.25 and 0.11, respectively, substantial differences are indicated in the use of metadiscourse markers among the three L1 groups.

Table 3. Intraclass Correlation Coefficients for the Three Variables Selected by the Stepwise Method

Variables	ICCs
TRA	0.00
SEM	0.25
BOO	0.11

Given the sizes of the ICCs (Table 3), multilevel ordinal logistic analysis was carried out with a random intercept model that predicts learners' proficiency levels on the basis of the frequencies of metadiscourse categories. In this model, the three learner groups had group-specific regression lines with different intercepts but the same slope. The results of the model were compared with those of the single-level model to reveal L1 effects on L2 writing. Table 4 summarizes the results of the single-level and multilevel ordinal logistic regression analysis.

Table 4. Results of Single-Level and Multilevel Ordinal Logistic Regression Analysis

	Single-level model		Multilevel model	
formula	CEFR ~ TRA + SEM + BOO		CEFR ~ TRA + SEM + BOO + (1 | L1)	
TRA	−0.08	*	−0.08	*
	(0.04)		(0.04)	
SEM	−0.05	***	−0.04	*
	(0.02)		(0.02)	
BOO	−0.11	*	−0.10	
	(0.05)		(0.05)	
1|2	−1.72	***	−1.78	***
	(0.24)		(0.37)	
2|3	0.24		0.25	
	(0.23)		(0.37)	
3|4	2.28	***	2.34	***
AIC	2356.04		2319.45	
log-likelihood	−1172.02		−1152.72	
N	1000		1000	
groups			3	
variance (intercept)			0.24	

Note. The row "formula" shows the setting of predictor and response variables (and random effects) for each regression model; "TRA," "SEM," and "BOO" represent (a) its regression coefficients and (b) standard errors written in parentheses; "1|2," "2|3," and "3|4" mean the threshold coefficients; "AIC," "log-likelihood," and "*N*" indicate the AIC values, log-likelihood values, and number of objects, respectively; "groups" and "variance (intercept)" refer to the number of learner groups and the variance of random intercepts in the multilevel model, respectively. The asterisks next to the numbers denote the significance levels (*$p < .05$, ***$p < .001$) of the coefficients.

Table 4 shows that, although the results of both analyses were similar, they differed at several points. In particular, the significance levels of self-mentions and boosters were lower in the multilevel model than in the single-level model. This means that the single-level model overestimates their significance by ignoring the multilevel structure in the analyzed data. In other words, the use of metadiscourse markers differs with respect to learners' L1 backgrounds. Therefore, we must consider the influence of learners' L1s to correctly identify the linguistic features that predict their proficiency levels.

The multilevel model is superior to the single-level model in terms of its explanatory power. As Table 4 shows, the AIC of the multilevel model (2319.45) is lower than that of the single-level model (2356.04). When the AIC values of two models differ by more than 10, there is no reason to choose the model with the higher value (Burnham & Anderson, 2004). Furthermore, the result of the likelihood ratio test comparing the two models indicates a significant difference in their explanatory power (LR = 27.71, df = 1, ***p < .001).

4.3 Significant variables

Table 4 indicates that the coefficients of transitions and self-mentions were significant in the multilevel model. Figures 2 and 3 show the relationship between the two significant linguistic features and learners' proficiency levels by marginalizing (averaging) out the effects of other variables. The horizontal axes in these plots indicate the relative frequency of the particular linguistic feature (per 100 words). The vertical axes show the probabilities of the response variables.

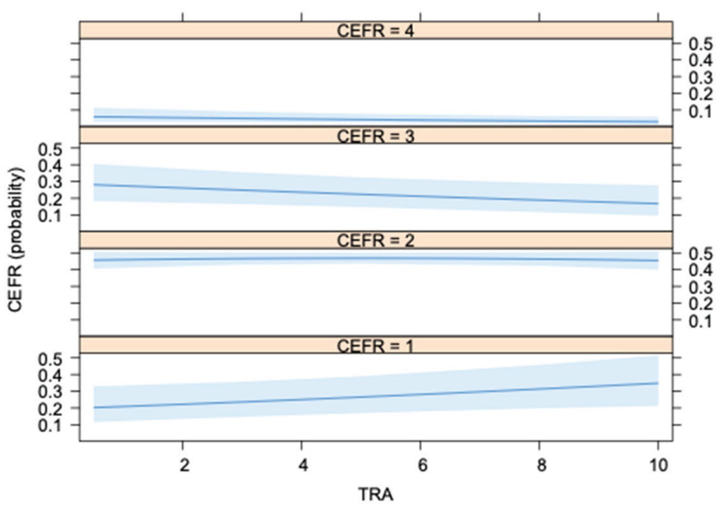

Figure 2. Effect plot for transitions

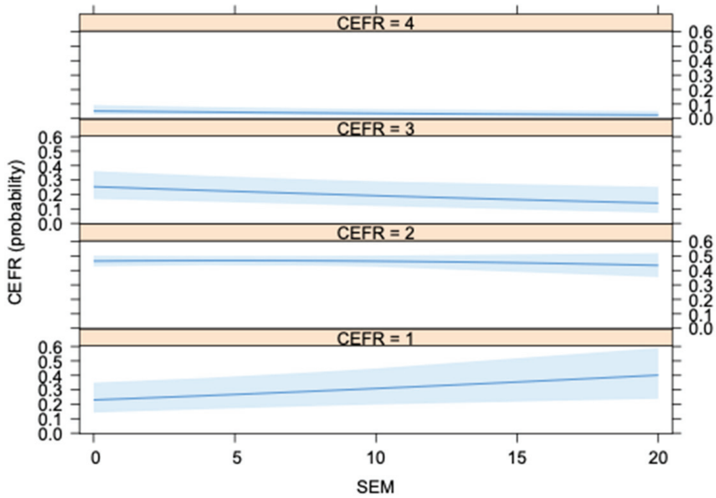

Figure 3. Effect plot for self-mentions

Figures 2 and 3 clearly demonstrate that a higher frequency of transitions and self-mentions characterizes lower proficiency levels. Specifically, learners with CEFR levels 1 and 2 used these two linguistic features more frequently than those with levels 3 and 4. As also pointed out in previous studies (Kobayashi, 2009, 2017), transitions and self-mentions are characteristics of novice English learners. As learners become more proficient, the relative frequencies of the two metadiscourse devices decrease because of the development of the sentence structure. In other words, linking words and first-person pronouns can predict the quality of L2 writing as a developmental index.

5 Conclusion

The goal of the present study was to investigate L2 developmental indices while controlling for L1 effects using a multilevel ordinal logistic regression analysis. The results suggest that the frequencies of transitions and self-mentions are the best predictors of language development of the metadiscourse categories defined by Hyland (2005). Additionally, this study shows that the multilevel model can measure the significance of predictor variables more accurately than the single-level model. This means that we must consider the effect of learners' L1s for correct identification of the linguistic features that predict their proficiency levels.

In addition to the potential annotation errors mentioned above, this study has some limitations. First, as each text was fairly short, some metadiscourse categories were little or not used at all. This type of problem persists if the writings of novice and intermediate learners are analyzed. Second, the target learners were limited to learners from three Asian countries. It would be desirable to investigate a wider range of L1 backgrounds to gain a broader understanding of L1 effects on L2 development. Third, other linguistic features can be instrumental in contrasting multiple learner groups. For instance, syntactic complexity and grammatical accuracy highlight the influence of L1 from other perspectives than metadiscourse. Moreover, the statistical model used in this study might be too simple to afford a full understanding of the mechanism of language acquisition. To model complex phenomena in the real world, Bayesian multilevel analysis could be more useful because of its greater modeling flexibility, inclusion of prior distributions, and more informative results (Bürkner & Vuorre, 2019; Liddell & Kruschke, 2018). Nevertheless, the present study illustrates the usefulness of multilevel categorical regression models for LCR. As multilevel analysis can be used for longitudinal studies as well as cross-sectional studies (Kyle et al., 2020; Paquot et al., 2021), the methodology of this study can be applied to track the language acquisition process of individual learners.

References

Abe, M. (2014). Frequency change patterns across proficiency levels in Japanese EFL learner speech. *Journal of Applied Language Studies, 8*(3), 85–96.

Abe, M. (2019). Comparing errors across an L2 spoken and written error-tagged Japanese EFL learner corpus. In S. Götz & J. Mukherjee (Eds.), *Learner corpora and language teaching* (pp. 157–174). John Benjamins.

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

Agresti, A. (2010). *Analysis of ordinal categorical data* (2nd ed.). Wiley.

Agresti, A. (2012). *Categorical data analysis* (3rd ed.). Wiley.

Agresti, A. (2018). *An introduction to categorical data analysis* (3rd ed.). Wiley.

Bürkner, P., & Vuorre, M. (2019). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science, 2*(1), 77–101.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33,* 261–304.

Christensen, R. H. B. (2018). Cumulative link models for ordinal regression with the R package *ordinal* [PDF file]. Retrieved from https://cran.r-project.org/package=ordinal/vignettes/clm_article.pdf

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research, 28*(3), 369–382.

Finch, W. H., Bolin, J. E., & Kelley, K. (2019). *Multilevel modeling using R* (2nd ed.). CRC Press.

Fox, J., & Weisberg, S. (2020). Regression methods supported by the *effects* package [PDF file]. Retrieved from https://cran.r-project.org/package=effects/vignettes/methods-supported-by-effects.pdf

Gałecki, A., & Burzykowski, T. (2015). *Linear mixed-effects models using R: A step-by-step approach.* Springer.

Goldstein, H. (1987). *Multilevel models in educational and social research.* Griffin.

Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research, 1*(1), 7–24.

Gries, S. Th. (2015a). Statistics for learner corpus research. In S. Granger, G. Gilquin & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 159–182). Cambridge University Press.

Gries, S. Th. (2015b). The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora, 10*(1), 95–125.

Gries, S. Th. (2021). (Generalized linear) mixed-effects modeling: A learner corpus example. *Language Learning, 71*(3), 757–798.

Gries, S. Th., & Adelman, A. S. (2014). Subject realization in Japanese conversation by native and non-native speakers: Exemplifying a new paradigm for learner corpus research. In J. Romero-Trillo (Ed.), *Yearbook of corpus linguistics and pragmatics 2014: New empirical and theoretical paradigms* (pp. 35–54). Springer.

Gries, S. Th., & Deshors, S. C. (2014). Using regressions to explore deviations between corpus data and a standard/target: Two suggestions. *Corpora, 9*(1), 109–136.

Gries, S. Th., & Deshors, S. C. (2015). EFL and/vs. ESL? A multi-level regression modeling perspective on bridging the paradigm gap. *International Journal of Learner Corpus Research, 1*(1), 130–159.

Gries, S. Th., & Deshors, S. C. (2021). Statistical analyses of learner corpus data. In N. Tracy-Ventura & M. Paquot (Eds.), *The Routledge handbook of second language acquisition and corpora* (pp. 119–132). Routledge.

Gries, S. Th., & Wulff, S., (2013). The genitive alternation in Chinese and German ESL learners: Towards a multifactorial notion of context in learner corpus research. *International Journal of Corpus Linguistics, 18*(3), 327–356.

Heck, R. H., & Thomas, S. L. (2020). *An introduction to multilevel modeling techniques: MLM and SEM approaches* (4th ed.). Routledge.

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge.

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.* Longman.

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

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

Ionin, T., & Díez-Bedmar, M. D. (2021). Article use in Russian and Spanish learner writing at CEFR B1 and B2 levels: Effects of proficiency, native language, and specificity. In B. L. Bruyn & M. Paquot (Eds.), *Learner corpus research meets second language acquisition* (pp. 10–38). Cambridge 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). University of Strathclyde Press.

Kobayashi, Y. (2009). Profiling metadiscourse markers in native and non-native English. *Lexicon, 39,* 1–17.

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.

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

Kobayashi, Y. (2020). Rhetorical preferences in L2 writings: A contrastive analysis of metadiscourse markers. *Journal of Modern Languages, 30*(2), 1–24.

Kyle, K., Crossley, S., & Verspoor, M. (2020). Measuring longitudinal writing development using indices of syntactic complexity and sophistication. *Studies in Second Language Acquisition, 43*(4), 781–812.

Leacock, C., Chodorow, M., Gamon, M., & Tetreault, J. (2014). *Automated grammatical error detection for language learners* (2nd ed.). Morgan and Claypool.

Liao, T. F. (1994). *Interpreting probability models: Logit, probit, and other generalized linear models.* Sage.

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Psychology, 79,* 328–348.

Meunier, F. (2015). Developmental patterns in learner corpora. In S. Granger, G. Gilquin & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 379–400). Cambridge University Press.

Michel, M., Kormos, J., Brunfaut, T., & Ratajczak, M. (2019). The role of working memory in young second language learners' written performances. *Journal of Second Language Writing, 45,* 31–45.

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). Presses universitaires de Louvain.

Murakami, A. (2016). Modeling systematicity and individuality in nonlinear second language development: The case of English grammatical morphemes. *Language Learning, 66*(4), 834–871.

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.

Newman, J., & Cox, C. (2020). Corpus annotation. In M. Paquot & S. Th. Gries (Eds.), *A practical handbook of corpus linguistics* (pp. 25–48). Springer.

O'Connell, A. A., & McCoach, D. B. (Eds.) (2008). *Multilevel modeling of educational data.* Information Age Publishing.

Paquot, M., Naets, H., Gries, S. Th. (2021). Using syntactic co-occurrences to trace phraseological complexity development in learner writing: Verb + object structures in LONGDALE. In B. L. Bruyn & M. Paquot (Eds.), *Learner corpus research meets second language acquisition* (pp. 122–147). Cambridge University Press.

R Core Team (2020). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Online. http://www.r-project.org/

Roback, P., & Legler, J. (2020). *Beyond multiple linear regression: Applied generalized linear models and multilevel models in R.* CRC Press.

Römer, U., & Berger, C. M. (2019). Observing the emergence of constructional knowledge: Verb patterns in German and Spanish learners of English at different proficiency levels. *Studies in Second Language Acquisition, 41*(5), 1089–1110.

Saito, K., Macmillan, K., Mai, T., Suzukida, Y., Sun, H., Magne, V., Ilkan, M., & Murakami, A. (2020). Developing, analyzing and sharing multivariate datasets: Individual differences in L2 learning revisited. *Annual Review of Applied Linguistics, 40,* 9–25.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage.

Verspoor, M., Lowie, W., & Wieling, M. (2021). L2 developmental measures from a dynamic perspective. In B. L. Bruyn & M. Paquot (Eds.), *Learner corpus research meets second language acquisition* (pp. 172–190). Cambridge University Press.

Wulff, S., & Gries, S. Th. (2015). Prenominal adjective order preferences in Chinese and German L2 English: A multifactorial corpus study. *Linguistic Approaches to Bilingualism, 5*(1), 122–150.

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

Received: September 14, 2021
Revised: October 26, 2021
Accepted: November 15, 2021