ReCALL (2019), 31: 2, 135–149
doi:10.1017/S0958344018000162

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Advancing CALL research via data-mining techniques: Unearthing hidden groups of learners in a corpus-based L2 vocabulary learning experiment

Hansol Lee
University of California, Irvine, United States
Korea Military Academy, Republic of Korea (hansol3@uci.edu)

Mark Warschauer
University of California, Irvine, United States (markw@uci.edu)

Jang Ho Lee
Chung-Ang University, Republic of Korea (jangholee@cau.ac.kr)

Abstract
In this study, we used a data-mining approach to identify hidden groups in a corpus-based second-language (L2) vocabulary experiment. After a vocabulary pre-test, a total of 132 participants performed three online reading tasks (in random orders) equipped with the following glossary types: (1) concordance lines and definitions of target lexical items, (2) concordance lines of target lexical items, and (3) no glossary information. Although the results of a previous study based on variable-centred analysis (i.e. multiple regression analysis) revealed that more glossary information could lead to better learning outcomes (Lee, Warschauer & Lee, 2017), using a model-based clustering technique in the present study allowed us to unearth learner types not identified in the previous analysis. Instead of the performance pattern found in the previous study (more glossary led to higher gains), we identified one learner group who exhibited their ability to make successful use of concordance lines (and thus are optimized for data-driven learning, or DDL; Johns, 1991), and another group who showed limited L2 vocabulary learning when exposed to concordance lines only. Further, our results revealed that L2 proficiency intersects with vocabulary gains of different learner types in complex ways. Therefore, using this technique in computer-assisted language learning (CALL) research to understand differential effects of accommodations can help us better identify hidden learner types and provide personalized CALL instruction.

Keywords: corpus use; data-driven learning (DDL); data-mining approach; glossary; person-centred analysis/approach; second-language vocabulary learning

1. Background

Most quantitative analyses can be categorized into two main types: variable-centred and person-centred (see Bergman & Magnusson, 1997). The former is used to examine associations between variables, whereas the latter is used to identify groups of individuals with similar values across variables. To illustrate this using the main topic of the present study as an example, a variable-centred analysis would investigate whether providing different types of glossary information is associated with second language (L2) vocabulary learning, whereas a person-centred analysis

Cite this article: Lee, H. Warschauer, M. & Lee, J. H. (2019) Advancing CALL research via data-mining techniques: Unearthing hidden groups of learners in a corpus-based L2 vocabulary learning experiment. *ReCALL* 31(2): 135–149, doi:10.1017/S0958344018000162

© European Association for Computer Assisted Language Learning 2018

https://doi.org/10.1017/S0958344018000162 Published online by Cambridge University Press

would explore whether there exist different learner types when various types of glossary information are provided.

The use of person-centred analysis in the field of L2 research dates back to the early 1980s. It has been adopted to identify hidden patterns or groups among L2 learners in terms of L2 aptitude (e.g. Hummel & French, 2016; Skehan, 1986), L2 motivation (e.g. Csizér & Dörnyei, 2005; Papi & Teimouri, 2014), and L2 learning approaches or strategies (e.g. Yamamori, Isoda, Hiromori & Oxford, 2003). By using a battery of language tests and questionnaires, these studies have revealed that it is possible to identify different affective, cognitive, and achievement profiles of L2 learners.

The primary goal of the present study is to unearth hidden groups of learners in a corpus-based L2 vocabulary learning experiment using a data-mining technique that is frequently used in the field of computer science. To this end, we employed a model-based clustering technique, which uses statistical criteria to determine an optimum number of groups. This feature distinguishes itself from traditional clustering methods used thus far in the field of L2 research (e.g. hierarchical clustering or partitioning clustering; Csizér & Dörnyei, 2005; Skehan, 1986). Although these traditional clustering methods have the potential to exert strong power in identifying similar groups of learners "based on strength of and relationships among several [outcome] variables" (Papi & Teimouri, 2014: 495), they are considered heuristic because they are not based on formal models, thus requiring researchers to make a subjective decision on the optimal number of clusters (see Meilă & Heckerman, 2001; Witten, Frank, Hall & Pal, 2016). Adopting more advanced techniques from a cutting-edge data-mining approach might help us unearth hidden groups or patterns from a data set in a more reliable and precise way.

Moreover, scant attempts have been made in L2 and computer-assisted language learning (CALL) research to understand possible hidden groups or patterns from data obtained from experimental designs aimed to compare the effects of different types of treatment. To address this paucity, we used data obtained from a previous experiment in an instructed L2 context, in which concordance lines excerpted from corpora were provided as glossary information in CALL reading environments (Lee, Warschauer & Lee, 2017). On the assumption that learners would make meaning inferences of target vocabulary by exploring the concordance lines provided (so-called "data-driven learning," or DDL; Johns, 1991), their lexical inferences via the multiple contextual examples excerpted from corpora can be more accurate when additional confirmation opportunities via dictionary definitions of target vocabulary are provided. A previous variable-centred analysis revealed that, on average, DDL was effective but that providing an additional definition glossary led to even more L2 vocabulary gains (Lee *et al.*, 2017). Given that clustering groups of similar individuals in terms of the aforementioned variables (i.e. L2 aptitude, motivation, and strategy use) has provided a new angle on the problem, adopting data-mining techniques for experimental data could reveal information that is equally important, if not more so. For example, when there are significant differences between treatment and control groups in terms of outcome variables, it is possible that "there is considerable variation within these groups" (Staples & Biber, 2015: 243). A close examination of this variation could thus reveal some interesting findings about the effect of a target treatment on certain groups of learners.

Overall, we believe that adopting a data-mining approach makes a significant contribution in that we can glean information about how individuals behave in each learning condition, subsequently allowing us to provide personalized instruction. As the present study is about CALL reading environments equipped with glossary information with one type of this information being corpus-based input (i.e. concordance lines), in the following sections we review the literature on different learner types in L2 vocabulary learning with different glossary types and corpus-based L2 vocabulary learning. We then present a description of methods, our findings, and implications.

https://doi.org/10.1017/S0958344018000162 Published online by Cambridge University Press

1.1 Different learner types in L2 vocabulary learning with different glossary types

There have been continuous empirical efforts in the L2 vocabulary learning and CALL literatures to examine the impact of different glossary types, such as L1 and L2 glosses (e.g. Yoshii, 2006), multimedia glosses (e.g. Lomicka, 1998; Yanguas, 2009), glosses in different positions on screen (e.g. AbuSeileek, 2011; Chen & Yen, 2013; Lee & Lee, 2013, 2015), and concordance-based glosses (e.g. Lee *et al.*, 2017; Poole, 2012). However, only limited attempts have been made to understand different learner types in L2 vocabulary learning with different glossary types (e.g. Chun, 2001; Plass, Chun, Mayer & Leutner, 1998).

In one example, Plass *et al.* (1998) developed a computer program called CyberBush to improve L2 reading comprehension by offering different types of glossary. Participants in their study were German learners who read a story that consisted of 762 German words. Participants could click on verbal information (i.e. text translated into their first language) and/or multimedia cues (i.e. picture or video clips in their first language) on 24 target lexical items. Based on log-file data of learners' choices of glossary types, the results revealed that on average students achieved better vocabulary gains when they selected both glossary types rather than selecting one or none. Further, Plass *et al.* found different learner types who showed diverse preferences towards glossary types to maximize their reading comprehension.

In another example, Chun (2001) developed a web program called netLearn to investigate L2 learners' preferences between instructor-created glosses (internal glossary) and electronic dictionaries (external glossary). In addition to log-file data of learners' clicks on websites, a small number of participants were asked to conduct think-aloud protocols as they read. Post-intervention interviews were then conducted with a few selected students. The results from both the quantitative and qualitative data revealed that on average participants showed better reading comprehension when they were provided access to both internal and external glossaries than when provided access only to an external glossary. Further, Chun identified different learner types who had varied beliefs in which glossary type was more helpful and showed contrasting strategies for L2 vocabulary learning during their activities.

Overall, only a few studies in the field of L2 vocabulary learning have examined different learner types in CALL experiments, and researchers have emphasized that more research is needed to tackle this issue. With similar purpose, a data-mining technique that can provide a statistical basis for similarity-based aggregating (Witten *et al.*, 2016) could be used to identify different learner types based on data obtained from an experiment.

1.2 Positive impact of corpus-based glossary information and different learner types

Inferring meanings of unfamiliar L2 vocabulary, also known as L2 lexical inferencing, is considered a successful vocabulary learning strategy (e.g. Fraser, 1999; Nassaji, 2003; Schmitt, 2000). For this reason, it is believed that corpus use can promote self-driven L2 vocabulary learning by allowing learners to explore authentic language data on their own or with some level of guided induction (e.g. Cobb, 1999; Johns, 1991; Lee *et al.*, 2017). Learners can discover linguistic features of target vocabulary such as its contextual meanings and collocation patterns as they are induced by multiple contextual examples excerpted from corpora (DDL; Johns, 1991). Recent meta-analysis studies (e.g. Boulton & Cobb, 2017; Lee, Warschauer & Lee, 2018) have confirmed the overall positive effect of corpus use on L2 vocabulary learning, although they did not distinguish the studies focusing on the effects of independent lexical inferencing and those on the effects of guided induction in DDL.

However, to the best of our knowledge, little has been studied about how learners may differentially benefit from corpus-based glossary information (Boulton, 2009; Flowerdew, 2008; Lee *et al.*, 2017). For example, in our previous study (Lee *et al.*, 2017), we found four emerging patterns of vocabulary learning from the participants' interactions with e-glosses. However, the results were not about learner types but about vocabulary types, such that vocabulary items may

https://doi.org/10.1017/S0958344018000162 Published online by Cambridge University Press

need different glossary types for them to be learned most effectively. Furthermore, although recent meta-analyses (Boulton & Cobb, 2017; Lee *et al.*, 2018) included a moderator analysis to identify factors related to learner types in corpus-based L2 learning, these studies failed to identify such factors because of limited data and incomplete reporting of the included studies. Nevertheless, we believe that corpus-based glossary information may not be appropriate for all types of learners. For example, there could be different learner types, considering that language data excerpted from corpora could be beyond some learners' comprehension levels (Lee *et al.*, 2018) or could impose different amounts of cognitive load (Lee & Lee, 2015). Thus, the present study aimed to unearth different learner types in a corpus-based L2 vocabulary learning experiment via a data-mining approach.

2. Present study

Data mining is generally categorized into two types: (1) supervised and (2) unsupervised (Witten *et al.*, 2016). A distinguishing factor of these two types is whether data mining is being implemented with or without a response variable (i.e. predefined labels or classifications of observations). For example, supervised data mining is mainly used to predict or classify observations from a new data set based on the algorithm computed from pre-classified observations in an original data set. Conversely, unsupervised data mining is mainly used to analyse a given data set without predefined labels or classifications and identify hidden structures of the data set. In the present study, we focused on unsupervised data mining because our goal was to unearth possible different learner types from experimental data. As part of this approach, we also hypothesized that using a model-based clustering technique based on statistical distributions and probabilities in identifying clusters (see Fraley & Raftery, 2002) would expand our perspective and advance CALL research.

Based on the suggestions of researchers (e.g. Chun, 2001; Lee *et al.*, 2017; Plass *et al.*, 1998) and limited empirical evidence on learner types (e.g. Boulton & Cobb, 2017; Lee *et al.*, 2017, 2018), we conjecture that there may be different learner types who show different learning patterns to maximize their L2 vocabulary learning. Such differential learning patterns would deviate from previous variable-centred findings on the correlation between the amount of glossary information provided and L2 vocabulary knowledge gains at the group level (Lee *et al.*, 2017). We believe that this study provides a clearer picture on how to explore different learner types when a data-mining approach (as part of person-centred analysis) is accompanied by variable-centred analysis. We address the following three guiding research questions:

1. Are there hidden groups of learners in a corpus-based L2 vocabulary learning experiment?
2. If so, how similar or different are they from each other in terms of maximizing their L2 vocabulary learning?
3. What is the role of L2 proficiency in relation to different learner types?

3. Methods

In this study, we used experimental data published by Lee *et al.* (2017) that addressed effects of different glossary types on L2 vocabulary learning: (1) concordance lines and definitions, (2) concordance lines only, and (3) no glossary–all of which were equipped with English reading tasks in CALL reading environments. In this section, we describe the methodological aspects involved in comparing these three conditions as well as the details of our data analysis plan.

https://doi.org/10.1017/S0958344018000162 Published online by Cambridge University Press

3.1 Participants

A total of 132 L2 undergraduate students with a wide range of academic majors in Republic of Korea participated in this study. They were convenience samples from six intact EFL classes. We adopted a repeated-measures design to have each participant experience all three reading tasks in a random order. To determine students' L2 (English) proficiency, we collected their scores on the Test of English for International Communication (TOEIC) developed by the Educational Testing Service (2016). The average TOEIC score of these participants was 732, indicating that they were at an intermediate level of English proficiency (CEFR B1; TOEIC scores between 550 and 785). According to a post hoc power analysis, the sample size was large enough to yield sufficient statistical power for the study design.[1]

3.2 Materials

We extracted three reading texts, each consisting of approximately 500 words, from newspaper articles on social issues from an English textbook (Cunningham, Moor & Carr, 2003). Through a pilot study using students with similar characteristics to our participants, we determined key features of glossary information (e.g. a preferred and manageable number of sample sentences for students' L2 vocabulary learning) and 30 unfamiliar lexical items from the reading texts, including nouns (e.g. *endowments*, *fib*), verbs (e.g. *traipse*, *tuck*), adjectives (e.g. *mucky*, *dodgy*), and multi-lexical items (e.g. *be beset with*, *in the vicinity of*). Each reading text included 10 of the target lexical items evenly distributed throughout the text.[2]

The reading texts were made compatible with three conditions of glossary information.[3] For the concordance lines and dictionary definitions condition (CODI), each target vocabulary item was hyperlinked to a pop-up window showing three preselected concordance lines via a customized tool. After consulting these concordance lines, learners could confirm the inferred meaning through dictionary definition of the target vocabulary item.[4] For the concordance lines only condition (CONC), target vocabulary items were hyperlinked to a pop-up window for concordance lines. For the no glossary or control condition (CTRL), target vocabulary items were underlined without providing lexical information.

3.3 Study design and previous findings

To prevent possible order effects, we computed six order types to distribute the three tasks: (1) task 1–task 2–task 3, (2) task 1–task 3–task 2, (3) task 2–task 1–task 3, (4) task 2–task 3–task 1, (5) task 3–task 1–task 2, and (6) task 3–task 2–task 1. Participants within condition were then randomly assigned to these order types. After a pre-test of target lexical items, the participants completed the three reading tasks in an ordinary classroom with their own laptops, followed immediately by post-tests of target lexical items (e.g. task 1–post-test 1–task 2–post-test 2–task 3–post-test 3). The pre-test and each post-test took five minutes, and each reading task took 15 minutes. Scheduling considerations did not allow us to include a definition-only condition.

In the post-tests, the participants were required to write the meaning of each target lexical item in either the L1 (Korean) or the L2 (English). Their answers were given a zero point for

[1]Faul, Erdfelder, Lang and Buchner's (2007) G*Power software is one of the most popular tools for power analysis. Results indicated that the required sample sizes were 12, 28, or 163 for large, medium, and small effect sizes, respectively, for three conditions under a repeated-measures design with a power of 0.80.

[2]The list of 30 target vocabulary items and their definitions can be found in Lee *et al.* (2017), Appendix B.

[3]To ensure learners' comprehension, concordance lines from the BNC, the OANC, and the Brown Corpus were carefully selected by the authors. Detailed information about the process of our selection can be found in Lee *et al.* (2017), Appendix A. Here are hyperlinks to these material samples: CODI (http://hansol6461.dothome.co.kr/corpus(2014)/article-1-d.htm), CONC (http://hansol6461.dothome.co.kr/corpus(2014)/article-1-c.htm), and CTRL (http://hansol6461.dothome.co.kr/corpus(2014)/article-1-p.htm).

[4]We used the Merriam-Webster Online Dictionary (http://www.merriam-webster.com) for dictionary definitions of target vocabulary items.

https://doi.org/10.1017/S0958344018000162 Published online by Cambridge University Press

inaccurate meaning, one point for partially correct meaning, or two points for completely correct meaning.

In our previous study (Lee *et al.*, 2017), we found that on average the participants achieved higher post-test scores than their pre-test scores and that their gains were statistically significant ($p < .001$) within each of the different conditions. When compared between the conditions, we found that the participants under the CODI condition achieved the highest post-test scores on average, followed by those under the CONC condition and then those under the CTRL condition. The achievement differences between the conditions were statistically significant ($p < .05$). It was also found that there was no order effect and that having different sets of target lexical items in each reading task did not affect participants' overall vocabulary gains (see Lee *et al.*, 2017, for more details).

3.4 Data analysis plan

To answer the first research question, as part of our data-mining approach, we used a model-based clustering technique with *R* 3.4.2 for Windows along with the mclust package version 5.3 (entitled "Gaussian mixture modelling for model-based clustering"; Fraley, Raftery, Scrucca, Murphy & Fop, 2017; Scrucca, Fop, Murphy & Raftery, 2016). The mclust package identifies clusters in given data based on pre-defined Gaussian mixture models (GMMs; see Fraley & Raftery, 2002; Jung, Kang & Heo, 2014). Provided that any given data can be multi-dimensional, models can be distinguished from each other in terms of their distributions, volumes, shapes, and orientations (see this article's supplementary materials for detailed information about predefined GMMs included in the mclust package).[5]

In general, any statistical technique works best with normally distributed data. Although model-based clustering has been reported to be robust for non-normal data sets (Mun, von Eye, Bates & Vaschillo, 2008; Yeung, Fraley, Murua, Raftery & Ruzzo, 2001), the method assumes that identified clusters are "concentrated locally about linear subspaces" by using pre-defined normally distributed models to statistically unearth hidden groups of cases (Fraley & Raftery, 1998: 586). This indicates that it would also work best with data that are normally distributed, or Gaussian.

For this reason, we checked the normality of students' performance data, such as pre- and post-test scores, by using univariate and multivariate normality tests after applying each of four commonly used data transformations: logarithm, square root, reciprocal, and reverse score (see Field, 2009; Pires & Branco, 2010, for data transformations). First, we used a skewness and kurtosis test for each variable (i.e. the Improved D'Agostino Test suggested by Royston, 1991). This test compares the symmetry and flatness of the distribution of a variable to the normal distribution. Second, we used two widely used tests for multivariate skewness and kurtosis (i.e. the Henze–Zirkler [Henze & Zirkler, 1990] and the Doornik–Hansen [Doornik & Hansen, 2008] tests). These tests examine "whether the marginal distributions and linear combinations of *x*-variables are normal and whether observations of pairs of *x*-variables show the elliptical appearance of the equal-density contours" (Tacq, 2010: 338; see this reference for more information about the multivariate normal distribution).

The results indicated that the square root transformation satisfied the Gaussian mixture assumption for the post-test scores according to the normality tests at the 1% significance level. Although the pre-test scores were important indicators of students' achievement, we found that the data values, which measured participants' unfamiliarity with target vocabulary items, could not be normally distributed, as both the mean and standard deviation were very close to zero.

[5]Scrucca *et al.* (2016) provide more details on the model-based clustering technique, including examples with short scripts of *R* code.

https://doi.org/10.1017/S0958344018000162 Published online by Cambridge University Press

It was further found that gain scores, which can be calculated by subtracting pre-test scores from post-test scores, failed the normality tests under all the data transformations.[6] As a result, we used the post-test scores for the model-based clustering, and included the pre-test scores as covariates in the following data analysis to take account of the baseline differences.

To answer the remaining research questions, we conducted logistic regression and multiple regression analyses. With pre-test scores and gender as covariates, L2 proficiency scores were analysed to (1) check how participants of different cluster memberships were similar to or different from each other, and (2) understand the role of L2 proficiency in relation to different learner types in different learning conditions. To ensure consistency, all data analyses followed by the model-based clustering used the square roots of the vocabulary test scores.

4. Results

4.1 RQ #1. Hidden groups in a corpus-based L2 vocabulary learning experiment

The results of the data-mining technique indicated that there were two clusters across different glossary types for maximizing L2 vocabulary learning potentials. Specifically, we found the best fit at the two-cluster solution according to the lowest Bayesian information criterion (BIC) value (BIC = −511.92; see supplementary materials for detailed information about how the optimal model can be determined by BIC via the mclust package). This result indicated that there were different groups of learners in terms of the square roots of their post-test scores in three conditions.

4.2 RQ #2. Similarities and differences in L2 vocabulary learning between hidden groups

Table 1 presents the descriptive statistics of vocabulary test scores for the two different learner types, as well as the total sample. Corresponding to the overall finding that all participants, on average, gained a statistically significant amount of L2 vocabulary knowledge within each of the three conditions, both the two learner types achieved higher vocabulary post-test scores than pre-test scores in all conditions and the differences were statistically significant ($p < .001$).

To further analyse how the two groups of learners were similar to or different from each other across the different glossary types, we conducted an ANOVA by using the square roots of the post-test scores, which were used for the model-based clustering technique. Roughly speaking, the two learner types appeared to exhibit L2 vocabulary learning patterns largely corresponding to our previous finding at the group level (i.e. CTRL < CONC < CODI; see Total sample row in Table 1) but with different magnitudes of gains (i.e. higher or lower gains according to the size of *t*-values from *t*-tests).

First, the finding that the participants performed well under the CONC and CODI conditions, when compared to CTRL, indicates that both the two learner types might have harnessed corpus-based language input as a source of their vocabulary learning, supporting researchers' advocacy for providing concordance lines as glossary information for L2 vocabulary learning (e.g. Frankenberg-Garcia, 2012, 2014; Lee *et al.*, 2018; Poole, 2012).

Second, participants' higher scores under the CODI condition compared to the CONC condition indicated that providing definitions of target lexical items in addition to concordance lines had a positive impact on L2 vocabulary learning. This condition likely enabled learners to confirm their inferred meaning of target lexical items (e.g. Cobb, Greaves & Horst, 2001; Fraser, 1999). This finding also corresponds to Godwin-Jones's (2001) idea that learners' inaccurate inferences from inductive learning would be minimized if their meaning inferences can later be confirmed through definition of target lexical items.

[6]Furthermore, according to Fitzmaurice, Laird, and Ware (2012), using pre-test scores as covariates is more appropriate than using gain scores for our study design (i.e. a randomised controlled trial). See Maris (1998) for an in-depth discussion about covariance adjustment versus gain scores.

https://doi.org/10.1017/S0958344018000162 Published online by Cambridge University Press

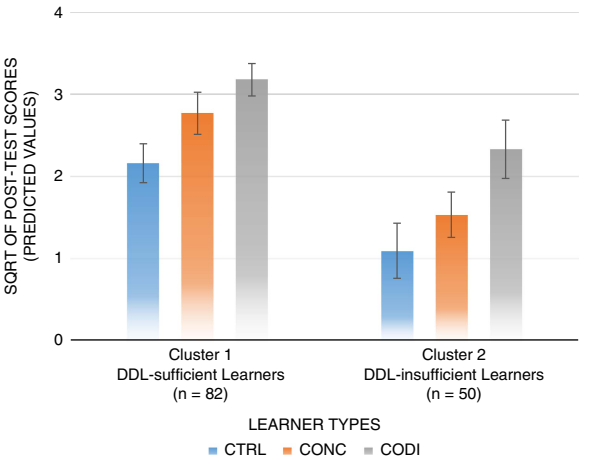


Figure 1. Two clusters (learner types) and their L2 vocabulary learning patterns
Note. Values are from Table 2. Error bars represent 95% confidence intervals.

Table 1. Descriptive statistics of vocabulary test scores

	Learner type	Condition	SQRT of pre-test	SQRT of post-test	*t*-test[a]	ANOVA[b]
Before clustering	Total sample ($N = 132$)	CTRL	.21 (.47)	1.76 (.93)	18.11***	$F(2, 393) = 49.46$***
		CONC	.19 (.40)	2.30 (.97)	25.34***	CTRL < CONC < CODI
		CODI	.13 (.36)	2.85 (.89)	34.79***	
After clustering	Cluster 1 ($n = 82$)	CTRL	.27 (.52)	2.17 (.65)	21.83***	$F(2, 243) = 44.60$***
		CONC	.22 (.43)	2.77 (.66)	33.30***	CTRL < CONC < CODI
		CODI	.17 (.41)	3.17 (.63)	38.83***	
	Cluster 2 ($n = 50$)	CTRL	.13 (.36)	1.10 (.94)	6.86***	$F(2, 147) = 22.35$***
		CONC	.13 (.36)	1.54 (.90)	10.83***	CTRL < CONC < CODI
		CODI	.06 (.24)	2.32 (1.00)	15.96***	

Note. Standard deviations are in parentheses. SQRT = square root.
***$p < .001$.
[a]*t*-tests examined whether the square roots of post-test scores were different from the square roots of pre-test scores.
[b]ANOVA results compared the square roots of post-test scores between the conditions.

However, results from the multiple regression analysis with the square roots of the vocabulary post-test scores as the dependent variable and the square roots of the vocabulary pre-test scores, gender, and L2 proficiency scores as independent variables or covariates reported a slightly different picture.

We obtained bar graphs (see Figure 1) by using the predicted values of the dependent variable (i.e. the square roots of the post-test scores across three different conditions) from the multiple regression analysis (see Table 2). Although the bars largely correspond to the overall finding (i.e. CTRL < CONC < CODI), examination of error bars suggested that some of the differences between the conditions could be statistically insignificant (i.e. overlapping error bars). In brief, the vocabulary gains of Cluster 1 in CONC and CODI conditions and the gains of Cluster 2 in CTRL and CONC were not statistically different, respectively (see Table 2). We named these two clusters as follows: (1) "DDL-sufficient learners," and (2) "DDL-insufficient learners."

First, we named Cluster 1 DDL-sufficient learners ($n = 82$) to refer to those who achieved the highest post-test scores when they performed DDL under CONC and CODI conditions (no statistically significant difference between these two conditions; $t = 2.04$, $p = .08$). This learning

https://doi.org/10.1017/S0958344018000162 Published online by Cambridge University Press

Table 2. Predicted values of vocabulary post-test scores from multiple regression analysis

		Predicted values			
	Learner type	CTRL	CONC	CODI	Contrasts at 5% significance level
Before clustering	Total sample ($N = 132$)	1.76 (.14)	2.30 (.11)	2.86 (.10)	CTRL < CONC < CODI
After clustering	DDL-sufficient learners ($n = 82$)	2.16 (.12)	2.77 (.13)	3.18 (.10)	CTRL < CONC ≤ CODI
	DDL-insufficient learners ($n = 50$)	1.09 (.17)	1.53 (.14)	2.33 (.18)	CTRL ≤ CONC < CODI

Note. Values came from multiple regression analysis with the square roots of the vocabulary post-test scores as the dependent variable and vocabulary pre-test scores, gender, and L2 proficiency scores as independent variables or covariates. See supplementary materials for the complete results of the analysis. Standard errors are in parentheses.

pattern indicated that receiving concordance lines as glossary information was sufficient for these learners to be successfully induced to discover the meaning of target lexical items by independently exploring concordance lines. Further, given that this learner type has higher English proficiency (TOEIC scores: $M = 759.04$, $SD = 81.07$) than the average L2 proficiency level of the DDL-insufficient learners ($M = 688.50$, $SD = 86.39$), we can speculate that DDL can be more effective for learners with higher L2 proficiency (Boulton, 2009; Flowerdew, 2015; Leńko-Szymańska & Boulton, 2015).

Second, we named Cluster 2 DDL-insufficient learners ($n = 50$) because they were unlikely to achieve significantly higher scores when they received concordance lines only as glossary information compared to under the CTRL condition where no glossary was provided (no difference between these two conditions; $t = 1.52$, $p = .17$). In other words, DDL was not an effective way of learning L2 vocabulary for this learner type compared to DDL-sufficient learners. As suggested by previous researchers (e.g. Huang, 2011; Lee & Lee, 2015), students of this learner type may find that the concordance lines provided are incomprehensible or unsuitable in accordance with their relatively low L2 proficiency level ($M = 688.50$, $SD = 86.39$). This assumption, as well as findings of a study by Chun (2001) that participants generally prefer glossary information that requires a lower cognitive load, enabled us to suggest that DDL-insufficient learners probably learned target lexical items by leveraging dictionary definitions under the CODI condition (Cobb *et al.*, 2001; Fraser, 1999) without benefiting much from DDL activities.

4.3 RQ #3. Role of L2 proficiency in relation to different learner types

Finally, we used the English proficiency (TOEIC) variable to check if it was related to cluster membership. In addition to the simple comparison between DDL-sufficient learners and DDL-insufficient learners in terms of L2 proficiency level, we conducted a logistic regression analysis to predict probabilities of a participant falling into specific learner types. The results revealed that there was a statistically significant odds ratio of 1.01, $\chi 2(5) = 77.69$, $SE = .002$, $p < .001$, for DDL-sufficient learners, indicating that the odds for being this learner type increase about 1% for every one point increase in TOEIC score, and that students with higher L2 proficiency levels were more likely to fall in this learner type (see supplementary materials for the complete results of the analysis). On the other hand, the odds for being the other learner type (DDL-insufficient learners) increase about 1% for every one-point decrease in TOEIC score; thus, those with lower L2 proficiency levels were more likely to fall into the DDL-insufficient learner type.

To further investigate the role of L2 proficiency, we conducted a multiple regression analysis with L2 proficiency as an independent variable for each treatment condition. As shown in Table 3, the results indicated that L2 proficiency had positive associations with vocabulary post-test scores for DDL-sufficient learners in the CONC ($b = .002$, $SE = .001$, $p < .05$) and CODI ($b = .002$, $SE = .001$, $p < .05$) conditions, and for DDL-insufficient learners in the CONC

https://doi.org/10.1017/S0958344018000162 Published online by Cambridge University Press

Table 3. Role of L2 proficiency identified from multiple regression analysis

		Coefficient of TOEIC		
	Learner type	CTRL	CONC	CODI
Before clustering	Total sample ($N = 132$)	.002* (.001)	.004*** (.001)	.003*** (.001)
After clustering	DDL-sufficient learners ($n = 82$)	.001 (.001)	.002* (.001)	.002* (.001)
	DDL-insufficient learners ($n = 50$)	−.002 (.002)	.003* (.001)	.002 (.002)

Note. Values came from multiple regression analysis with the square roots of the vocabulary post-test scores as the dependent variable and vocabulary pre-test scores, gender, and L2 proficiency scores as independent variables or covariates. See supplementary materials for the complete results of the analysis. Standard errors are in parentheses.
*$p < .05$. ***$p < .001$.

($b = .003$, $SE = .001$, $p < .05$) condition. Given that L2 proficiency was significantly associated with L2 vocabulary learning across all three conditions including CTRL in the full sample, the null impact of L2 proficiency under CTRL for the two learner types indicated an exclusive role of L2 proficiency in determining the success of DDL tasks. Along these lines, the null impact of L2 proficiency under CODI for DDL-insufficient learners partly corroborated one of our speculations about possible absence of DDL efforts of this learner type in CODI. Overall, although the practical magnitude of the impact was not large enough to draw a meaningful interpretation (i.e. small coefficients of the L2 proficiency variable), we found that L2 proficiency played a significant role in influencing successful DDL.

5. Discussion

In this study, we adopted a data-mining approach to unearth hidden groups of learners in an instructed L2 vocabulary-learning context using a model-based clustering technique. In doing so, we were able to extend our previous finding that more glossary information led to better overall learning outcomes (Lee *et al.*, 2017), suggesting that identified learner types from the clustering technique may not exactly follow the overall learning pattern. We found that participants exposed to all three learning conditions through a repeated-measures design could be divided into two learner types: (1) DDL-sufficient learners, whose ability to make use of concordance lines made them suitable for DDL (cf. Johns, 1991), but who did not benefit as much from additional dictionary information as had been expected (e.g. Godwin-Jones, 2001; Lee *et al.*, 2017); and (2) DDL-insufficient learners, who did not benefit much from only receiving concordance lines as glossary information. That is, by using data mining, this study shed light on unidentified learner types overshadowed by the average obtained through data analysis at the group level. The two learner types had distinctively different learning patterns, so combining them produced a poorly defined "one-size-fits-all" conclusion. We thus offered support to what Staples and Biber (2015: 243) asserted, that clustering techniques can "provide a bottom-up way to identify new groups that are better defined with respect to target variables."

In this way, we were able to identify the complex role of learners' L2 proficiency level in this CALL intervention. The model-based clustering technique we employed here is designed to identify statistically more similar and homogeneous groups of learners, and this led us to recognize a strong predictive power of L2 proficiency not only between but also within the identified learner types. Between the two learner types, we found that students with higher L2 proficiency were more likely to be DDL-sufficient learners, and those with lower L2 proficiency were more likely to be DDL-insufficient learners. Within each learner type, data mining helped us find that learners' high levels of L2 proficiency were especially crucial for successful DDL activities, unlike the previous finding at the group level, where students' overall L2 proficiency level is expected to influence the magnitude of L2 vocabulary gains equally across

https://doi.org/10.1017/S0958344018000162 Published online by Cambridge University Press

all conditions. That is, we found more direct statistical evidence that DDL can be beneficial to vocabulary learning for L2 learners in general, but the impact can be greater for those with higher L2 proficiency (e.g. Boulton, 2009; Flowerdew, 2015; Leńko-Szymańska & Boulton, 2015). This may bring us one step closer to understanding "the types of learners who take most readily to DDL or extract most benefit from it" (Boulton, 2009: 87), corroborating the significant role of L2 proficiency in DDL highlighted by recent meta-analyses (Boulton & Cobb, 2017; Lee *et al.*, 2018).

Given that corpus use for L2 vocabulary learning is "an active, creative, and socially interactive process" (Rüschoff & Ritter, 2001: 223), more research is required to understand which learner factors could affect successful DDL activities, such as motivation (e.g. Gass, Behney & Plonsky, 2013), strategy use (e.g. Tseng & Schmitt, 2008), use of knowledge sources (e.g. Nassaji, 2003), learning styles (e.g. Flowerdew, 2008), and working memory capacities (e.g. Martin & Ellis, 2012). Again, findings of the present study highlight that results from aggregating individuals (as in the case of data analysis at the group level) should be interpreted cautiously as providing only a partial perspective on L2 vocabulary learning.

6. Implications and limitations

Our study has several implications for future research on L2 learning. It is expected that the use of data-mining techniques in analysing experimental data sets will expand research paradigms in several possible ways. First, we recommend the use of data-mining techniques in addition to variable-centred statistical analysis. Such an implementation could either produce similar results across different analyses (and thus enhance the validity of the overall result), or present a conflicting finding that could provide useful information about hidden groups of learners with different profiles. Second, researchers may administer a series of pre-tests and questionnaires on individual differences and use data-mining techniques to examine if their participants can be clustered in meaningful ways prior to an intervention (see Staples & Biber, 2015, for similar suggestions). Researchers could then further examine possible interaction effects between each cluster and target intervention(s), which could provide valuable findings regarding personalized instruction.

The limitation of the present study is that we could not fully harness our data set by excluding the vocabulary pre-test scores in the model-based clustering technique. We used these scores to interpret the results of clustering in a statistically robust way, and yet we wonder how results might have differed if the pre-test score variable had been normally distributed, in which case it would have been possible to include it in the data mining. According to our clustering simulation with the six variables (i.e. square roots of pre-test and post-test scores for CTRL, CONC, and CODI), we found six clusters with uneven sizes (large differences between the size of the clusters; i.e. 9, 12, 12, 13, 35, and 51, respectively), which could be a convincing sign of bias (see Firooz, 2015, for a discussion). Moreover, the small sample size of the data set could be another limitation in this case, because the number of variables (i.e. degree of data dimensionality) generally requires a corresponding sample size. Although there is no rule of thumb about the sample size necessary for clustering techniques, one suggests using $5 \times 2k$ (k = number of variables) as the minimum sample size (see Dolnicar, 2002, for a review), which would be a minimum of 320 samples for this case. Overall, it is complex and difficult to evaluate findings of unsupervised data mining due to an absence of true labels or classifications. For this reason, future studies on this topic with larger samples are warranted to understand how individual learners gain their L2 vocabulary knowledge in diverse ways.

7. Conclusion

Using a data-mining technique at the individual level, our results indicated that there were different learner types who exhibited learning patterns that differed slightly from our previous

https://doi.org/10.1017/S0958344018000162 Published online by Cambridge University Press

finding at the group level–a positive association between the amount of glossary information provided and post-test scores of L2 vocabulary knowledge. As a result, we identified that individual learners might require different accommodations to maximize their L2 vocabulary learning potentials. For example, for the DDL-sufficient learner type, receiving concordance lines was enough for successful L2 vocabulary learning, while the process of confirming inferred meaning did not substantially contribute to L2 vocabulary learning. The vocabulary post-test scores of DDL-insufficient learners also corresponded to a concern that DDL may not be as effective for some learners (Schmitt, 2008). Therefore, closer attention to individual types of learners is required (e.g. Boulton, 2009; Flowerdew, 2015; Lee & Lee, 2015; Leńko-Szymańska & Boulton, 2015). If L2 researchers can implement similar approaches in their future studies, this could contribute to a better understanding of CALL environments equipped with different learning accommodations and the development of more personalized instruction.

Supplementary materials. For supplementary materials referred to in this article, please visit https://doi.org/10.1017/S0958344018000162

Acknowledgements. The authors would like to thank Jamal Abedi, Dorothy M. Chun, Joshua Lawrence, Robin C. Scarcella, and Julio R. Torres for their helpful comments on an earlier version of the manuscript; Kameryn Denaro, Jacquelynne S. Eccles, Alexander Ihler, and Steve Peck for statistical guidance and support; and the editor and three anonymous reviewers for their thoughtful feedback and suggestions.

Ethical statement. The present study uses data from a previously published study (Lee *et al.*, 2017). In Lee et al. (2017) and the present study, effort was made to ensure the participants' anonymity. We also declare that there is no conflict of interests.

References

AbuSeileek, A. F. (2011) Hypermedia annotation presentation: The effect of location and type on the EFL learners' achievement in reading comprehension and vocabulary acquisition. *Computers & Education*, 57(1): 1281–1291. https://doi.org/10.1016/j.compedu.2011.01.011

Bergman, L. R. & Magnusson, D. (1997) A person-oriented approach in research on developmental psychopathology. *Development and Psychopathology*, 9(2): 291–319. https://doi.org/10.1017/S095457949700206X

Boulton, A. (2009) Data-driven learning: Reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics*, 35(1): 81–106.

Boulton, A. & Cobb, T. (2017) Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2): 348–393. https://doi.org/10.1111/lang.12224

Chen, I.-J. & Yen, J.-C. (2013) Hypertext annotation: Effects of presentation formats and learner proficiency on reading comprehension and vocabulary learning in foreign languages. *Computers & Education*, 63: 416–423. https://doi.org/10.1016/j.compedu.2013.01.005

Chun, D. M. (2001) L2 reading on the Web: Strategies for accessing information in hypermedia. *Computer Assisted Language Learning*, 14(5): 367–403. https://doi.org/10.1076/call.14.5.367.5775

Cobb, T. (1999) Applying constructivism: A test for the learner-as-scientist. *Educational Technology Research and Development*, 47(3): 15–31. https://doi.org/10.1007/BF02299631

Cobb, T., Greaves, C. & Horst, M. (2001) Can the rate of lexical acquisition from reading be increased? An experiment in reading French with a suite of on-line resources. In Raymond, P. & Cornaire, C. (eds.), *Regards sur la didactique des langues seconds*. Montréal: Éditions logique, 133–153.

Csizér, K. & Dörnyei, Z. (2005) Language learners' motivational profiles and their motivated learning behavior. *Language Learning*, 55(4): 613–659. https://doi.org/10.1111/j.0023-8333.2005.00319.x

Cunningham, S., Moor, P. & Carr, J. C. (2003) *Cutting edge: Advanced with phrase builder*. Harlow: Pearson Education.

Dolnicar, S. (2002) A review of unquestioned standards in using cluster analysis for data-driven market segmentation. In Shaw, R. N., Adam, S. & McDonald, H. (eds.), *ANZMAC 2002: Proceedings of the Australian and New Zealand Marketing Academy Conference 2002*. Deakin University, 2–4 December, 31–37.

Doornik, J. A. & Hansen, H. (2008) An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, 70(s1): 927–939. https://doi.org/10.1111/j.1468-0084.2008.00537.x

Educational Testing Service (2016) *TOEIC® listening and reading test scored and the CEFR levels*. https://www.etsglobal.org/Tests-Preparation/The-TOEIC-Tests/TOEIC-Listening-Reading-Test/Scores-Overview

https://doi.org/10.1017/S0958344018000162 Published online by Cambridge University Press

Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007) G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39: 175–191. https://doi.org/10.3758/BF03193146

Field, A. P. (2009) *Discovering statistics using SPSS* (3rd ed.). London: Sage.

Firooz, H. (2015, March 4) *When not to use Gaussian mixture model (EM clustering)*. https://hameddaily.blogspot.com/2015/03/when-not-to-use-gaussian-mixtures-model.html

Fitzmaurice, G. M., Laird, N. M. & Ware, J. H. (2012) *Applied longitudinal analysis* (2nd ed.). Hoboken: John Wiley & Sons.

Flowerdew, L. (2008) Pedagogic value of corpora: A critical evaluation. In Frankenberg-Garcia, A. (ed.), *Proceedings of the 8th Teaching and Language Corpora conference*. Associação de Estudos e de Investigação Científica do ISLA-Lisboa, 115–119.

Flowerdew, L. (2015) Data-driven learning and language learning theories: Whither the twain shall meet. In Leńko-Szymańska, A. & Boulton, A. (eds.), *Multiple affordances of language corpora for data-driven learning*. Amsterdam: John Benjamins, 15–36.

Fraley, C. & Raftery, A. E. (1998) How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal*, 41(8): 578–588. https://doi.org/10.1093/comjnl/41.8.578

Fraley, C. & Raftery, A. E. (2002) Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458): 611–631. https://doi.org/10.1198/016214502760047131

Fraley, C., Raftery, A. E., Scrucca, L., Murphy, T. B. & Fop, M. (2017) *mclust: Gaussian mixture modelling for model-based clustering, classification, and density estimation* (*R* package version 5.3) https://CRAN.R-project.org/package=mclust

Frankenberg-Garcia, A. (2012) Learners' use of corpus examples. *International Journal of Lexicography*, 25(3): 273–296. https://doi.org/10.1093/ijl/ecs011

Frankenberg-Garcia, A. (2014) The use of corpus examples for language comprehension and production. *ReCALL*, 26(2): 128–146. https://doi.org/10.1017/S0958344014000093

Fraser, C. A. (1999) Lexical processing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition*, 21(2): 225–241. https://doi.org/10.1017/S0272263199002041

Gass, S. M., Behney, J. & Plonsky, L. (2013) *Second language acquisition: An introductory course* (4th ed.). New York: Routledge.

Godwin-Jones, R. (2001) Tools and trends in corpora use for teaching and learning. *Language Learning & Technology*, 5(3): 7–12. https://doi.org/10125/44559

Henze, N. & Zirkler, B. (1990) A class of invariant consistent tests for multivariate normality. *Communications in Statistics – Theory and Methods*, 19(10): 3595–3617. https://doi.org/10.1080/03610929008830400

Huang, L.-S. (2011) Language learners as language researchers: The acquisition of English grammar through a corpus-aided discovery learning approach mediated by intra- and interpersonal dialogues. In Newman, J., Baayen, H. & Rice, S. (eds.), *Corpus-based studies in language use, language learning, and language documentation*. Amsterdam: Rodopi, 91–122.

Hummel, K. M. & French, L. M. (2016) Phonological memory and aptitude components: Contributions to second language proficiency. *Learning and Individual Differences*, 51: 249–255. https://doi.org/10.1016/j.lindif.2016.08.016

Johns, T. (1991) Should you be persuaded: Two examples of data-driven learning. In Johns, T. & King, P. (eds.), *Classroom concordancing. English Language Research Journal*, 4: 1–16.

Jung, Y. G., Kang, M. S. & Heo, J. (2014) Clustering performance comparison using *K*-means and expectation maximization algorithms. *Biotechnology & Biotechnological Equipment*, 28(Supp. 1): S44–S48. https://doi.org/10.1080/13102818.2014.949045

Lee, H. & Lee, J. H. (2013) Implementing glossing in mobile-assisted language learning environments: Directions and outlook. *Language Learning & Technology*, 17(3): 6–22. https://doi.org/10125/44334

Lee, H. & Lee, J. H. (2015) The effects of electronic glossing types on foreign language vocabulary learning: Different types of format and glossary information. *The Asia-Pacific Education Researcher*, 24(4): 591–601. https://doi.org/10.1007/s40299-014-0204-3

Lee, H., Warschauer, M. & Lee, J. H. (2017) The effects of concordance-based electronic glosses on L2 vocabulary learning. *Language Learning & Technology*, 21(2): 32–51. https://doi.org/10125/44610

Lee, H., Warschauer, M. & Lee, J. H. (2018) The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*. Advance online publication. https://doi.org/10.1093/applin/amy012

Leńko-Szymańska, A. & Boulton, A. (2015) Introduction: Data-driven learning in language pedagogy. In Leńko-Szymańska, A. & Boulton, A. (eds.), *Multiple affordances of language corpora for data-driven learning*. Amsterdam: John Benjamins, 1–14.

Lomicka, L. L. (1998) "To gloss or not to gloss": An investigation of reading comprehension online. *Language Learning & Technology*, 1(2): 41–50. https://doi.org/10125/25020

Maris, E. (1998) Covariance adjustment versus gain scores—revisited. *Psychological Methods*, 3(3): 309–327. http://dx.doi.org/10.1037/1082-989X.3.3.309

Martin, K. I. & Ellis, N. C. (2012) The role of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, 34(3): 379–413. https://doi.org/10.1017/S0272263112000125

https://doi.org/10.1017/S0958344018000162 Published online by Cambridge University Press

Meilă, M. & Heckerman, D. (2001) An experimental comparison of model-based clustering methods. *Machine Learning*, 42 (1/2): 9–29. https://doi.org/10.1023/A:1007648401407

Mun, E. Y., von Eye, A., Bates, M. E. & Vaschillo, E. G. (2008) Finding groups using model-based cluster analysis: Heterogeneous emotional self-regulatory processes and heavy alcohol use risk. *Developmental Psychology*, 44(2): 481–495. https://doi.org/10.1037/0012-1649.44.2.481

Nassaji, H. (2003) L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly*, 37(4): 645–670. https://doi.org/10.2307/3588216

Papi, M. & Teimouri, Y. (2014) Language learner motivational types: A cluster analysis study. *Language Learning*, 64(3): 493–525. https://doi.org/10.1111/lang.12065

Pires, A. M. & Branco, J. A. (2010) Projection-pursuit approach to robust linear discriminant analysis. *Journal of Multivariate Analysis*, 101(10): 2464–2485. https://doi.org/10.1016/j.jmva.2010.06.017

Plass, J. L., Chun, D. M., Mayer, R. E. & Leutner, D. (1998) Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology*, 90(1): 25–36. https://doi.org/10.1037-/0022-0663.90.1.25

Poole, R. (2012) Concordance-based glosses for academic vocabulary acquisition. *CALICO Journal*, 29(4): 679–693. https://www.jstor.org/stable/pdf/calicojournal.29.4.679.pdf

Royston, P. (1991) sg3.5: Comment on sg3.4 and an improved D'Agostino test. *Stata Technical Bulletin*, 3: 23–24. https://stata-press.com/journals/stbcontents/stb3.pdf

Rüschoff, B. & Ritter, M. (2001) Technology-enhanced language learning: Construction of knowledge and template-based learning in the foreign language classroom. *Computer Assisted Language Learning*, 14(3-4): 219–232. https://doi.org/10.1076/call.14.3.219.5789

Schmitt, N. (2000) *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Schmitt, N. (2008) Review article: Instructed second language vocabulary learning. *Language Teaching Research*, 12(3): 329–363. https://doi.org/10.1177/1362168808089921

Scrucca, L., Fop, M., Murphy, T. B. & Raftery, A. E. (2016) mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, 8(1): 289–317. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5096736

Skehan, P. (1986) Cluster analysis and the identification of learner types. In Cook, V. (ed.), *Experimental approaches to second language acquisition*. Oxford: Pergamon, 81–94.

Staples, S. & Biber, D. (2015) Cluster analysis. In Plonsky, L (ed.), *Advancing quantitative methods in second language research*. New York: Routledge, 243–274.

Tacq, J. (2010) Multivariate normal distribution. In Peterson, P., Baker, E. & McGaw, B. (eds.), *International encyclopedia of education* (3rd ed.). Oxford: Elsevier, 332–338. https://doi.org/10.1016/B978-0-08-044894-7.01351-8

Tseng, W.-T. & Schmitt, N. (2008) Toward a model of motivated vocabulary learning: A structural equation modeling approach. *Language Learning*, 58(2): 357–400. https://doi.org/10.1111/j.1467-9922.2008.00444.x

Witten, I. H., Frank, E., Hall, M. A. & Pal, C. J. (2016) *Data mining: Practical machine learning tools and techniques* (4th ed.). Cambridge, MA: Morgan Kaufmann.

Yamamori, K., Isoda, T., Hiromori, T. & Oxford, R. L. (2003) Using cluster analysis to uncover L2 learner differences in strategy use, will to learn, and achievement over time. *International Review of Applied Linguistics in Language Teaching*, 41(4): 381–409. https://doi.org/10.1515/iral.2003.017

Yanguas, I. (2009) Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. *Language Learning & Technology*, 13(2): 48–67. https://doi.org/10125/44180

Yeung, K. Y., Fraley, C., Murua, A., Raftery, A. E. & Ruzzo, W. L. (2001) Model-based clustering and data transformations for gene expression data. *Bioinformatics*, 17(10): 977–987. https://doi.org/10.1093/bioinformatics/17.10.977

Yoshii, M. (2006) L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology*, 10(3): 85–101. https://doi.org/10125/44076

About the authors

Hansol Lee is a PhD candidate in the School of Education at the University of California, Irvine. As an applied linguist with a wide range of research interests and methodologies, his recent work has been published in *Applied Linguistics*, *Language Learning & Technology*, and *ReCALL*, among others.

Mark Warschauer is a professor of Education and Informatics at the University of California, Irvine, where he directs the Digital Learning Lab and the Teaching and Learning Research Center, and the editor of *AERA Open* journal. His research focuses on digital media and learning.

https://doi.org/10.1017/S0958344018000162 Published online by Cambridge University Press

Jang Ho Lee received his DPhil in education from the University of Oxford. He is presently an associate professor in the Department of English Education at Chung-Ang University. His work has been published in *Applied Linguistics*, *Language Teaching Research*, *The Modern Language Journal*, *TESOL Quarterly*, *System*, *Language Awareness*, and so on. All correspondence regarding this publication should be addressed to him.

Author ORCiD. Hansol Lee, http://orcid.org/0000-0002-6912-7128
Author ORCiD. Mark Warschauer, http://orcid.org/0000-0002-6817-4416
Author ORCiD. Jang Ho Lee, http://orcid.org/0000-0003-2767-3881

https://doi.org/10.1017/S0958344018000162 Published online by Cambridge University Press