C U E

Reports from the Online CUE Conference, September 11, 2023

OCJSI

Volume 4
2022 Special Conference Issue

Editor

Glen Hill *Obihiro University of Agriculture & Veterinary Medicine*

Guest Editors

Message from the Editor 2

Practice-Oriented Paper

Message to the Readers

Our 2022 ESP Symposium was held online for the first time in its history, and it was due to the COVID-19 pandemic still posing risks to in-person contact. As a result, we had to change from a paper poster event to something else. We decided on Powerpoint presentations, and the speaker turnout was unfortunately quite small, although the Edzilla host site recorded 57 participants. In fact, there were only five presentations, but those were bolstered by two keynote speakers. **Professor An Cheng** made his keynote talk ("But my students lack the technical knowledge in their areas of study": Key tenets and issues in the ESP genre-based approach") from Oklahoma State University, USA despite the time difference. **Professor Helen Basturkmen** had less of a time difference from the University of Auckland, NZ and spoke on "ESP research in Asia Pacific: A topical analysis of themes in recent research".

CUE continued to stay at the forefront of diversity by being awarded the red Equal Voices in ELT award for its choice of keynote speakers.

Contributors to the *OnCUE Journal Special Issue* were three of the presenters. **Jonathan Ferries** expanded his talk on corpus work and offers us his article here,

"Using Word Frequencies to Introduce Corpora in the Classroom". **Tetsuya Fukuda** and **Kimberly Klassen** teamed up for an article "Exploring IELTS Scores in an English for Academic Purposes Program Using Cluster Analysis". And, **Hiroshi Kohsaka** has an article for us entitled "Building essential and state-of-the-art vocabulary in information science: observations from classes at a Japanese university".

We at CUE feel honored by everyone's contribution to the conference, and we hope that the one in 2024 will attract more speakers as before. It is not known at this time whether it will be an online-only, hybrid, or in-person event.

I wish to offer my very special thanks to the co-editors **R. Paul Lege** and **Ryota Nishino**, both of Nagoya University.

Following the talks, CUE held a special "Open floor discussion on inclusivity in CUE". It consisted of about an hour of free talk in two Zoom breakout rooms where participants of the symposium discussed questions posed by the moderator **Emily Choong**. Thanks, Emily! The questions posed and the replies generated are shown below.

Room 1

1) What are some problems about inclusivity in the ESL/EFL teaching community?

- Native speakerism (this note was deleted by someone...I don't know why)
- Some cultures (Myanmar) don't have surnames. Causes problems in filling out forms.
- Having foreign names spelled properly, or not at all in katakana. Having foreign names shown in the proper order.
- Junior/Senior members: junior members have to wait for promotions when ability doesn't matter.
- Showing Caucasian photos for advertising, instead of non-Caucasians.
- Higher level students may ask more, or may BE asked more Qs in class.

2) What are some solutions to the problems discussed in #1?

- Either address admin about your photo, or get the other pictured people (and Japanese coworkers) to learn your situation and get solidarity to approach admin to change the photo.
- adding global issues in lessons, including photos of people from various countries, because admin may have children who will see these lessons and take the thought home and open minds
- Don't use middle names when filling out forms. Use full caps for last name.
- Support or join a labor union.
- Tell admin to update their websites to accommodate properly spelled / ordered names.
- Change the image in society to allow promotions on merit over age. Also change the image of women who should be allowed promotions or heads of committees, too.
- DON'T call on the higher level students first. Or at all. Wait and tell the class "who else has an opinion/answer?"

Room 2

1) What are some problems about inclusivity in the ESL/EFL teaching community?

language on the new JALT awards still cut out some people...we need to revise.	too many assumptions are being made	themes: too narrow? not promoting inclusiveness
language is always English	An aside: attendees are noticing who is there and not there.	blind vetting results in more 'majority' speakers
junior/senior members	lack of Asian women representationesp. Japanese professors(even in the audience(language preference (there was one Japanese participant who left a JALT event because th	veteran/new teachers is also a problem

2) What are some solutions to the problems discussed in #1?

make a thing to approach people for presentations	make diversity the goal over ALL of your events throughout the year	have an option for 'language of presentation' when you submit an abstract proposal...including Japanese etc would help more people join in
be aware	veteran people can help others!	
be aware of who is not represented either as far as chosen speakers or attendees.	award for teacher-practioner collaborations	possible prizes/recognition for collaboration, mentorship, institutional reform?

Thanks again to everyone, and see you in 2024!

Glen Hill CUE SIG Publications Chair and *OnCUE Journal*

Chief Editor

Using Word Frequencies to Introduce Corpora in the Classroom

Jonathan Ferries

Bukkyo University

Classroom use of corpora has long been advocated but remains rare in practice. In this paper, I summarize the obstacles to corpus use, describe a simple frequency-based method of analysis that avoids these problems, and offer a gateway to a more sophisticated use of corpora to promote more active, independent learning.

In this paper, I describe a two-stage method of using word frequency counts from online corpora to provide a simple introduction to using online corpora in the classroom. Researchers have long advocated using corpora to enable learners to explore and make their own inductive language discoveries through data-driven learning (see Johns, 1986). In practice, however, corpora remain rarely used in the classroom (Boulton, 2017; Breyer, 2009; Ma et al., 2022; O'Keefe & Farr, 2003). This is attributed on the teacher side to lack of training in corpus skills (Boulton, 2017; Breyer, 2009; Leńko-Szymańska, 2014), and on the student side to insufficient knowledge of the metalanguage required to perform corpus searches (Chang, 2014; Yeh et al., 2007). Studies have also drawn attention to the problems faced by both teachers and students in analyzing and interpreting concordance lines (Breyer, 2009; Ma et al., 2022).

Concordance lines are lines of text centered around a nodal word or phrase, and producing them is one of the basic functions of corpus analysis software. However, they can be challenging to read for the inexperienced (O'Keefe et al., 2007) and are considered more suitable for use with more advanced learners. This poses a challenge for practitioners in locations such as Japan, where many learners do not advance beyond the lower intermediate level (Hadley, 2001).

To address these issues, the method presented below makes use of another basic function of corpus software: calculation of word frequencies. These calculations are relatively easy to perform, and they produce results that are visually easy to grasp. I illustrate how this method can be applied with an example drawn from the domain of English for academic purposes (EAP) that I have used with Japanese L1 learners of English majoring in sciences at the undergraduate level. With appropriate modifications, this method may be used with learners of lower intermediate level and above in other domains of English for specific purposes (ESP).

However, word frequencies cannot tell us how and in what context a word or phrase is used. For this, concordance lines showing a certain degree of co-text are required. The method described below serves only as a relatively accessible gateway to more sophisticated use of corpora, including concordances, in the classroom.

Use of "I" in academic writing

My example concerns use of the first-person pronoun in academic writing. This is a topic of clear relevance to EAP students, and one regarding which students may have received advice in the past. For example, both lower-intermediate textbooks (e.g., Ackert et al., 2014) and writing guides for advanced students (e.g., Bailey, 2011) from major publishing houses recommend avoidance of the first person in academic writing. Several corpus-based studies have demonstrated that this advice is not always followed in actual academic writing. Hyland (2001) and Dobakhti and Hassan (2017), for example, assembled their own large corpora to reveal extensive use of the first person in research articles and considerable variation between disciplines.

However, approaches like these that use purpose-built corpora are unfeasible for ESP classroom use. This is because they use corpora that take time to prepare and specialized software that takes more time to master. With time in the language curriculum often scarce, students and teachers need ready-made corpora that are cheap, accessible, and relevant. Several large corpora and corpus analysis tools are now accessible online that meet these requirements. These include the

British National Corpus (Davies, 2004), the Corpus of Contemporary American English (COCA) (Davies, 2008), and Lextutor (Cobb, 2017). I chose COCA because of its size and breadth in content, the simplicity of its search interface, the availability of instructions on its use in Japanese, and its cost (students can perform up to 50 searches per day for free).

Stage 1: Searching the corpus

The first stage consists of searching the chosen corpus for the features under investigation. In my example, the aim was to investigate how academic writers refer to themselves and their actions in order to determine whether they avoid using "I" and instead use structures such as the passive voice. This was operationalized by searching for occurrences of the pronoun *I* (as a proxy for first-person authorial references) and *be* + past participle (as a proxy for passive-voice authorial references) in each of the academic disciplines contained in COCA. This is potentially the most challenging step for teachers and students unfamiliar with corpora. Corpora are normally annotated by "tagging" to indicate the part of speech of each word. The purpose of this is to allow users to search for instances of, for example, *will* used as a modal verb but not as a noun. Users require some knowledge of these tags in order to perform effective searches. For simplicity, I used the query syntax `i_p*` to search for instances of the pronoun *I*, and the query syntax `are|were [vvn*]` to search for passive *be* + past participle constructions.[1]

A search of this kind produces results indicating the frequency of occurrence in each discipline of the features searched for (Figure 1).

These may be converted by students or the teacher to chart form (Figure 2). This shows, in a visually easy-to-grasp manner, that the advice not to use "I" in academic writing is not always followed by academics in practice; the first person appears to be used to some degree in all disciplines, and it is used more frequently than the passive voice in the Humanities and Philosophy/Religion.[2]

Owing to the crudity of the proxies and query syntax used, these results provide only a rough indication of the relative frequencies of use of the first person and passive voice in different disciplines. Use of pronoun *I* as a proxy

SECTION	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT
ACAD:History	17213	12.2	1,405.69	
ACAD:Education	17441	9.4	1,846.92	
ACAD:Geog/SocSci	32327	16.2	1,997.95	
ACAD:Law/PolSci	17172	8.6	1,996.65	
ACAD:Humanities	35944	11.9	3,013.80	
ACAD:Phil/Rel	20428	6.7	3,030.73	
ACAD:Sci/Tech	16173	14.1	1,149.03	
ACAD:Medicine	4786	6.7	714.28	
ACAD:Misc	34550	4.3	8,116.75	

Figure 1. Selected raw outputs (Source: COCA).
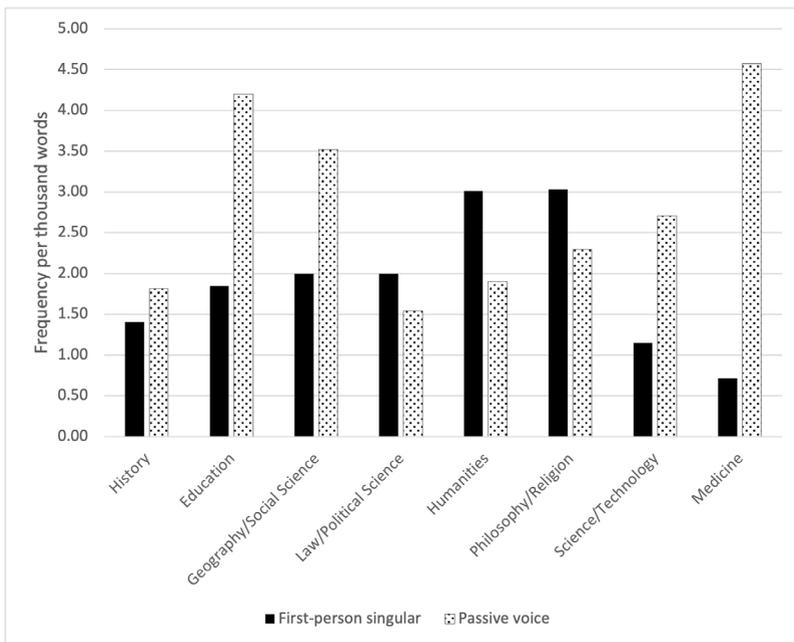
Figure 2. First person and passive voice frequencies by discipline

excluded first-person pronouns in other cases (*me*, *my*, *us*, *our*), which means that the results do not accurately reflect the frequency of use of all forms of first-person expression. The search for passive *be* + past participle using the syntax `are|were [vvn*]` is likewise limited in scope. It fails to capture, for example, constructions using the more colloquial passive auxiliary *get* and instances where

an adverb occurs between the auxiliary and past participle (e.g., *were frequently referred to*). A second, more in-depth stage of analysis is therefore required.

Stage 2: Manual search

For the second stage, a manual search is made of a smaller sample of texts. In the case of my example, a manual search was made for all instances of pronoun *I* referring to the writer (excluding, for example, instances in quotations and appended questionnaires) and all instances of the passive voice used to describe actions (including judgments and other mental processes) performed by the author.

Given time and other classroom constraints, the manual analysis was limited to the two disciplines found to exhibit the greatest relative differences in frequency of use of the first person and passive voice, namely Education and the Humanities (Figure 2). Three research articles (RAs) were randomly selected from the most recent issue of a journal from each of these disciplines—*ELT Journal* and *The Art Bulletin*—chosen based on online availability to students and the advice of expert informants in each discipline. However, alternative methods of selection, such as a journal rating metric or students' own knowledge of the field of interest, could be used.

This analysis, performed manually by the teacher and students, revealed that passive-voice use exceeded first-person use in all three Education RAs (Figure 3). Conversely, first-person use exceeded passive-voice use in all three Humanities RAs (Figure 4). These results corroborate the results of the cruder computerized analysis at the first stage.

Implications for the classroom

The first stage of the above analysis introduces newcomers to corpora to the tags and metalanguage needed to perform more sophisticated corpus analyses. The second stage introduces learners to complete texts. This stage exposes them to the organization and constitution of texts in their fields of interest. Thus, in the case of the example described above, students noted, unprompted, the different writing strategies employed in the acknowledgement sections compared to the other parts of academic articles. Importantly, from the point of view of data-
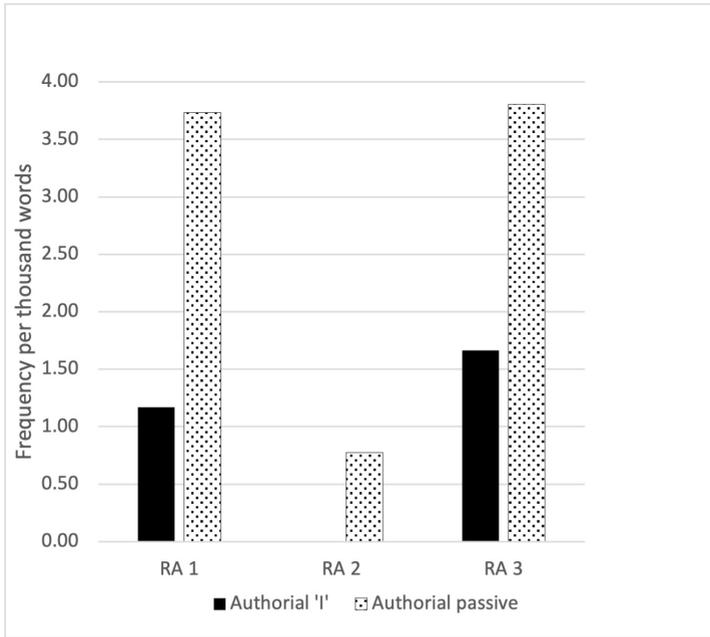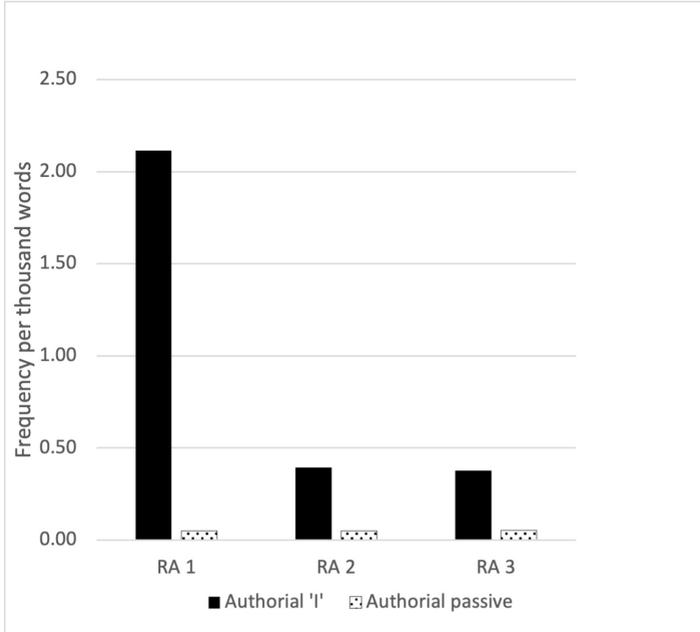
Figure 3. Education sample.

Figure 4. Humanities sample.

driven learning, the second stage raises students' awareness of how the words that they searched for at the first stage are used in context. This serves as a gateway for introducing concordances once students have mastered the necessary basic corpus skills. Concordances, and the tools used to obtain them, are considered a mainstay of data-driven learning, and the simple frequency-based corpus analysis proposed here offers one way of overcoming the obstacles to corpus use outlined at the outset to equip students to become more active, independent learners (Chen, 2011).

Notes

1. A full list and descriptions of these and other tags used by many online corpora, including COCA, can be found at the CLAWS part-of-speech tagger for English website. https://ucrel.lancs.ac.uk/claws

2. The disciplinary categories cited here are those used in COCA.

References

Ackert, P., Lee, L., Haynes, H., & Beck, J. (2014). *New reading and vocabulary development 2: Thoughts and notions (teacher's edition)*. Cengage Learning.

Bailey, S. (2011). *Academic writing: A handbook for international students*. Routledge.

Boulton, A. (2017). Corpora in language teaching and learning. *Language Teaching*, *50*(4), 483–506. https://doi.org/10.1017/S0261444817000167

Breyer, Y. (2009). Learning and teaching with corpora: Reflections by student teachers. *Computer Assisted Language Learning*, *22*(2), 153–172. https://doi.org/10.1080/09588220902778328

Chang, J.-Y. (2014). The use of general and specialized corpora as reference sources for academic English writing: A case study. *ReCALL*, *26*(2), 243–259. https://doi.org/10.1017/S0958344014000056

Chen, H.-J. H. (2011). Developing and evaluating a web-based collocation retrieval tool for EFL students and teachers. *Computer Assisted Language Learning*, *24*(1), 59–76. https://doi.org/10.1080/09588221.2010.526945

Cobb, T. (2017). *Range for texts v.3* [computer software]. Lextutor. https://

www.lextutor.ca/cgi-bin/range/texts/index.pl

Davies, M. (2004). *British national corpus* (from Oxford University Press). https://www.english-corpora.org/bnc/

Davies, M. (2008). The corpus of contemporary American English (COCA): 560 million words, 1990-present. https://corpus.byu.edu/coca

Dobakhti, L., & Hassan, N. (2017). A corpus-based study of writer identity in qualitative and quantitative research articles. *The Southeast Asian Journal of English Language Studies, 23*(1), 1–14. https://doi.org:10.17576/3L-2017-2301-01

Hadley, G. (2001). Concordancing in Japanese TEFL: Unlocking the power of data-driven learning. In K. Gray, M. Ansell, S. Cardew, & M. Leedham (Eds.). *The Japanese learner: Context, culture and classroom practice* (pp. 138–144). Oxford University Department for Continuing Education.

Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes, 20*, 207–226. https://doi:10.1016/S0889-4906(00)00012-000012-0 "10")

Johns, T. (1986). Micro-concord: A language learner's research tool. *System*, *14*(2), 151–162. https://doi.org/10.1016/0346-251X(86)90004-790004-7 "11")

Leńko-Szymańska, A. (2014). Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. *ReCALL*, *26*(2), 260–278. https://doi.org/10.1017/S095834401400010X

Ma, Q., Chiu, M. M., Lin, S., & Mendoza, N. B. (2023). Teachers' perceived corpus literacy and their intention to integrate corpora into classroom teaching: A survey study. *ReCALL*, *35*(1), 19–39. https://doi.org/10.1017/S0958344022000180

O'Keefe, A., & Farr, F. (2003). Using language corpora in initial teacher education: Pedagogic issues and practical applications. *TESOL Quarterly*, *37*(3), 389–418. https://doi.org/10.2307/3588397

O'Keefe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.

Yeh, Y., Liou, H. C., & Li, Y. H. (2007). Online synonym materials and concordancing for EFL college writing. *Computer Assisted Language Learning*, *20*(2), 131–152. https://doi.org/10.1080/09588220701331451

Author bio

Jonathan Ferries is a contract lecturer at Bukkyo University in Kyoto, Japan. f-paul@bukkyo-u.ac.jp

Exploring IELTS Scores in an English for Academic Purposes Program Using Cluster Analysis

Tetsuya Fukuda

International Christian University

Kimberly Klassen

Tokyo Woman's Christian University

We describe an investigation into individual differences in an English for Academic Purposes program. The aim is to explore characteristics of successful students in order to better support those less likely to succeed. Students' ($N = 349$) IELTS scores and survey responses were analyzed using cluster analysis. The survey concerned test preparation, prior test experience, and extrinsic motivation. Six subgroups with distinct characteristics were identified. Recommendations for program evaluation are made based on the findings.

English for Academic Purposes (EAP) is defined as "language research and instruction that focuses on the communicative needs and practices of individuals working in academic contexts" (Hyland & Shaw, 2016, p. 1). According to Jordan (2002), EAP is considered to be a key area within English for Specific Purposes (ESP). An important aspect of EAP programs for curriculum writers and teachers is formative evaluation of the program's objectives: are students achieving the learning goals, and why are some students more successful than others? Proficiency tests such as IELTS are often used to gauge students' academic English abilities after completing an EAP program. Many factors are involved in students' success on such proficiency tests, not the least of which are individual differences. The challenge for program heads is to tease apart these influences in

order to better support those learners who are less successful. Here, we describe an approach to identify patterns among students of a large cohort in order to determine whether a certain combination of factors can account for students' performance on IELTS, the program exit exam. The approach employs cluster analysis as a complementary analytical tool, which is easily adaptable in terms of the factors of interest to the program evaluation.

Cluster analysis is a multivariate procedure used to group people, as opposed to grouping variables as in factor analysis (Skehan, 1986). Measures are taken of a population sample on different variables, and based on patterns of scores on these variables, the sample is grouped into clusters of similar people (Skehan, 1986). Cluster analysis is often used to complement other traditional statistical procedures, as it can shed light on the role of individual differences that may not be readily apparent otherwise, as in Gu and Johnson's (1996) study on vocabulary learning strategies.

To investigate students' performance on the exit exam from our EAP program, we used cluster analysis as a complementary procedure to prior analyses (Erdelyi et al., 2018; Yagi & Fukuda, 2020), which had indicated much individual variation. This example analysis from our EAP program may be of interest to researchers seeking alternative approaches to program evaluation. Our research question is: what patterns among EAP program participants can be found in terms of IELTS test scores, test preparation, prior test experience, and extrinsic motivation?

Method

The data were collected from 349 first-year students at a Japanese university ($n = 183$ from the 2020 cohort and $n = 166$ from the 2021 cohort). Students join the EAP program in April, and after completing the program in February the following year, they take IELTS in March as an exit test to determine their academic English proficiency.

The data consist of two components: IELTS scores and responses to a survey. Table 1 shows the four 5-point Likert-scale items in the survey. They were drawn from our previous exploratory study (Yagi & Fukuda, 2020). The survey

aimed to ascertain how prepared, experienced, and motivated the participants were regarding the IELTS test. The responses to Item 1 and Item 2 were conceptualized as *preparation*. The responses to Item 3 represent *experience*, and those to Item 4 were *future use* (i.e., extrinsic motivation represented by the need for the test score in the future).

The IELTS scores ranged from 4.5 to 8.5 (M = 6.37, SD = 0.81). Since

Table 1

The Question Items and the Options in the Survey

Item 1	How many hours did you study to prepare for the IELTS exam?
	More than 10 hours
	Between 5 and 9 hours
	Between 2 and 5 hours
	Between 1 and 2 hours
	None
Item 2	Did you participate in the information sessions for IELTS held at Guidance Seminar in December and Preparation Seminar in January?
	Both days
	January only
	December only
	Part of one day
Item 3	Had you taken IELTS before you took it in March?
	4 or more times
	3 times
	2 times
	1 time
	Never
Item 4	How likely are you to use the IELTS results to study abroad in the near future, such as on an exchange program or graduate school?
	Very likely
	Quite likely
	Somewhat
	Not very likely
	Not at all

test scores and responses to survey items were used in the analysis, all data were standardized to z-scores. To answer the question of what patterns can be identified among participants, the data were analyzed employing cluster analysis with four factors explained above. The cluster analysis was conducted following the procedure in Staples and Biber (2015).

Results

As a result of the cluster analysis, the participants were categorized into six groups, which are shown in Figure 1. ANOVAs were conducted to confirm the validity of this six-group solution: the groups were significantly different in all four factors ($p < .001$), and the effect sizes were medium to high (.49 to .69 in eta squared).

The results from the analysis found that the six clusters of participants have the following characteristics. Cluster 1 ($n = 16$) were most experienced with the test. They did not prepare much, and their IELTS scores were above average. Cluster 2 ($n = 50$) reported the lowest need for the test scores in the future. They prepared less than average, and their scores were low. Cluster 3 ($n = 65$) had an average need for the test score in the future. However, they did not prepare for
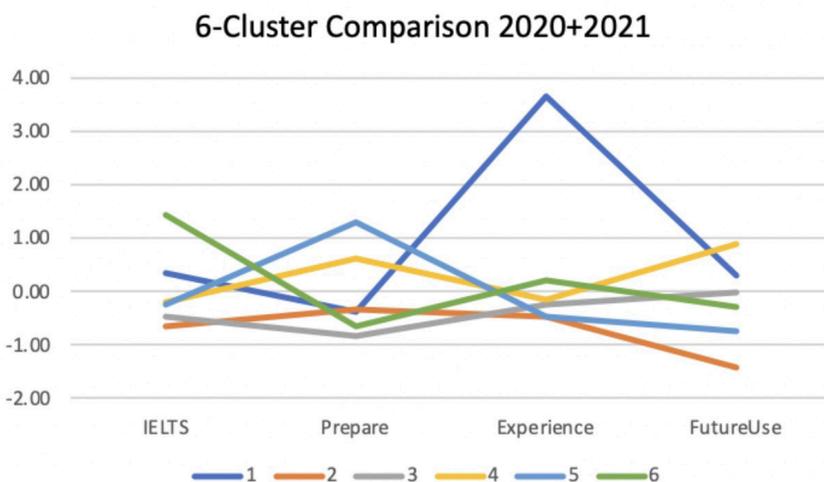

Figure 1. Six-cluster comparison 2020+2021

the test, and their scores were low. Cluster 4 (*n* = 123) had the most interest in using the score for the future. They prepared, but their scores were below average. Cluster 5 (*n* = 33) prepared the most, but the scores were below average. Their interest in using the score for the future was also below average. Cluster 6 (*n* = 62) had the highest scores but prepared less than average. Their test experience was just above average.

Discussion

As noted above, cluster analysis is an exploratory tool, and the findings can raise unexpected insights into the participant group (Staples & Biber, 2015). Here, we focus on three such observations among the clusters, which we will explore in future research.

First, we are interested in exploring further the effect of test experience on test scores. For example, Clusters 1 and 6 had the most experience, scored the highest, and prepared the least. This finding suggests that test experience may be more valuable than attending test information sessions before test day. Currently, there is no specific IELTS test preparation course available to the majority of these students before the exam, and yet having practical test experience seems to be helpful. Opening test prep courses is, therefore, an important consideration for the EAP program heads.

Secondly, we need to look further at the interaction between motivation and test preparation. For example, Clusters 2 and 5 both reported the lowest future need for the score (i.e., extrinsic motivation), yet Cluster 5 prepared the most for the test of all groups. It was unexpected that students with low extrinsic motivation would prepare a lot for the test. Thus, we need to further refine the factor of motivation for the purposes of the program evaluation analysis.

Lastly, we found that the factors of test experience, test preparation, and extrinsic motivation were not sufficient to explain the differences in test scores between all the participants. When we look at Clusters 3 and 6, these two groups seem to show similarities in terms of average test experience, below average test preparation, and an average need for the test in the future. However, in terms of test scores, Cluster 3 scored below average while Cluster 6 scored well above

average. What this suggests is that there are factors beyond those examined in this analysis that need to be explored to account for differing test performance.

Conclusion

Cluster analysis can be a useful complementary statistical tool for EAP program evaluation, as it attempts to group participants according to patterns of factors chosen by the researcher. This can bring to light individual differences that are not obvious through traditional analyses. Our findings have indicated several ways that the program heads can better support students in the exit exam process. For example, we found that some students identified a need for the test score, yet did not prepare much for the exam. Thus, we recommend making test prep courses available to these students before they sit the exam. We also found that having test experience was more valuable than attending the test information sessions. Again, this highlights the need for focused test preparation courses in which students can get hands-on experience with the IELTS format. Finally, we were not able to collect program entrance test scores due to Coronavirus restrictions. Our future program evaluations will include a comparison of entrance and exit test scores, as well as fine tuning complex factors such as motivation.

References

Erdelyi, T., Fukuda, T., & Yagi, K. (2018). Analysis of study abroad summer program. *OnCUE Journal, 11*(2), 116–122. https://jaltcue.org/files/OnCUE/OCJ11.2/OCJ11.2_pp116-122_Erdelyi_Fukuda_Yagi.pdf

Gu, Y., & Johnson, R. K. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning, 46*(4), 643–679. https://doi.org/10.1111/j.1467-1770.1996.tb01355.x

Hyland, K., & Shaw, P. (2016). Introduction. In K. Hyland, & P. Shaw (Eds.), *The Routledge handbook of English for academic purposes* (pp. 1–13). Routledge.

Jordan, R. R. (2002). The growth of EAP in Britain. *Journal of English for Academic Purposes, 1*(1), 69–78. https://doi.org/10.1016/s1475-1585(02)00004-8

Skehan, P. (1986). Cluster analysis and the identification of learner types. In V. Cook (Ed.), *Experimental approaches to second language learning* (pp. 81–94). Pergamon Press.

Staples, S., & Biber, D. (2015). Cluster analysis. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 243–274). Routledge. https://doi.org/10.4324/9781315870908-14

Yagi, K., & Fukuda, T. (2020). Exploratory investigation into successful and less successful learners of English with interviews. *Language Research Bulletin, 35*, 84–94. https://doi.org/10.34577/00004769

Author bios

Tetsuya Fukuda (PhD in Education, Temple University Japan) is a senior associate professor at Juntendo University. He teaches academic English courses and is interested in psychological factors in language learning. He has published papers in national and international journals, including more than 10 book chapters. t.fukuda.wv@juntendo.ac.jp

Kimberly Klassen (PhD, Applied Linguistics, Cardiff University, Wales) is an associate professor at Tokyo Woman's Christian University. Recent publications have appeared in Language Teaching, Reading Psychology, The Reading Matrix and The Curriculum Journal. She has taught in Korea, Taiwan, the United Arab Emirates and Japan. kimik@lab.twcu.ac.jp

Building Essential and State-of-the-art Vocabulary in Information Science: Observations from Classes at a Japanese University

Hiroshi Kohsaka

The University of Electro-Communications

The terminology of information science has increased significantly, which makes it difficult to effectively cover all essential technical terms in a class in one given semester. To provide students with an opportunity to improve their English proficiency, an autonomous class activity to collect keywords in information science was incorporated into an English course at a Japanese university. Results suggest that students in their second year and above can build both basic and state-of-the-art vocabulary autonomously.

The vocabulary in information science has expanded at a rapid pace in recent years. Since the invention of the first programmable digital computer ENIAC (Electronic Numerical Integrator and Computer) in 1946, novel technical terms have been added to the vocabulary of information science (Comer, 2017). Furthermore, the number of technical words is increasing in association with the exponential development of new technology, including artificial intelligence (Russell & Norvig, 2022) and quantum computation (Nielsen & Chuang, 2010).

Vocabulary acquisition is one of the core elements of language learning (Hu & Nation, 2000; Thornbury, 2002). A number of teaching strategies to support building vocabulary have been developed, such as verbal interaction with others (Wang, 2015), consulting dictionaries (Zhang et al., 2021), note-taking (Jin & Webb, 2021), and guessing from context (Nassaji, 2006). These effective learning strategies are implemented based on reliable vocabulary lists built in existence (Nation & Beglar, 2007; Browne et al., 2013). Thus, establishing

relevant vocabulary lists is prerequisite to conduct vocabulary building activities in ESP (English for Specific Purposes) classes. However, in the field of information science, although several vocabulary lists of the technical terms have been published (Esteras & Fabre, 2007; Hirai, 2012), it is still challenging for university teachers and students to follow the rapid pace of the appearance of new words due to the fact that published lists can become obsolete quickly.

This paper reports a survey and the results of an English class in a Japanese university aiming at the development of vocabulary in information science by a student-centered activity. Within a theme selected by a teacher, students selected relevant keywords at their discretion. The comparison between the keywords collected by students and a vocabulary textbook demonstrates two-fold results. First, the vocabulary developed by a large majority of students successfully nominated basic keywords in information science. Second, the students' list covers state-of-the-art words related to novel computer technology. These observations suggest that an autonomous class activity would be a viable option to tie together two areas of vocabulary in information science: a) permanent and foundation vocabulary and b) new and potentially transient vocabulary.

Methods

The data for this study were collected in ESP classes taught at the University of Electro-Communications in Tokyo, Japan. The classes were elective-compulsory subjects; each student selected one class among other compulsory language classes. The students were in their second year or above and have passed both Academic Written English and Academic Spoken English classes (taught by other instructors) in their first year. Therefore, the students possess adequate command of English for the specific needs of information science. The information science classes were taught face-to-face in the standard 15-week semester, and each class meeting lasted 90 minutes once per week. The students' proficiency in English is intermediate level, and they possess high literacy in information science.

A typical instructor-supervised student-centered activity took the following form throughout the course. The instructor assigned one topic from information science in each class (Table 1).

Table 1

The Topic of Each Class

Week	Topic	Week	Topic
1	Discrete math	7	Artificial intelligence
2	Digital logic	8	Deep learning
3	Data representation	9	Computer vision
4	Processors	10	Robotics
5	Memories	11	Quantum computing
6	Input/Output devices		

At the beginning of the class, each student brainstormed the topic for five minutes to find a keyword about the topic. The instructor allowed the students to access the Internet during this activity. Students conducted group discussions with two to four classmates to share their keywords. At the end of the class, each student searched for one keyword in five minutes and submitted both the keyword and its explanation to Google Forms. In total, 306 keywords were collected in 11 weeks. In the last week of the course (week 15), the instructor conducted a survey with a questionnaire to ask about the rigor of the course for reference to improve next year's class, in particular, to adjust the speed and the level of the course to the students' demand. The survey included multiple-choice single-answer questions (Figure 1). Eighty out of 87 students agreed to participate in this study. The instructor received informed consent from all the students whose work was used in this report. Python 3.7 was used to perform all the statistical analyses. The questionnaire responses about the level of the course were analyzed using the one-way ANOVA (analysis of variance) and post hoc Tukey test.

Results

The 306 keywords collected in the class were compared with the vocabulary list from a published book listing the essential vocabulary in information science

Figure 1. Part of the questionnaire conducted to ask about the level of the course

($N = 512$ words in Hirai, 2012). Figure 2 shows a Venn diagram of the number of overlapping words between student-selected words and the textbook vocabulary. The students' list included 27 essential keywords in information science (Figure 3).

To test whether the students could acquire state-of-the-art terminology, the instructor included four modern technology fields in the topic list ("Artificial intelligence," "Deep learning," "Computer vision," and "Quantum computing", Table 1). Students could find some state-of-the-art words through this class activity. As a consequence, several state-of-the-art keywords were listed in the vocabulary selected by students but not in the basic textbook (Figure 4): "object recognition" and "face authentication" from artificial intelligence; "quantum interference" and "q-bit" from quantum computing.
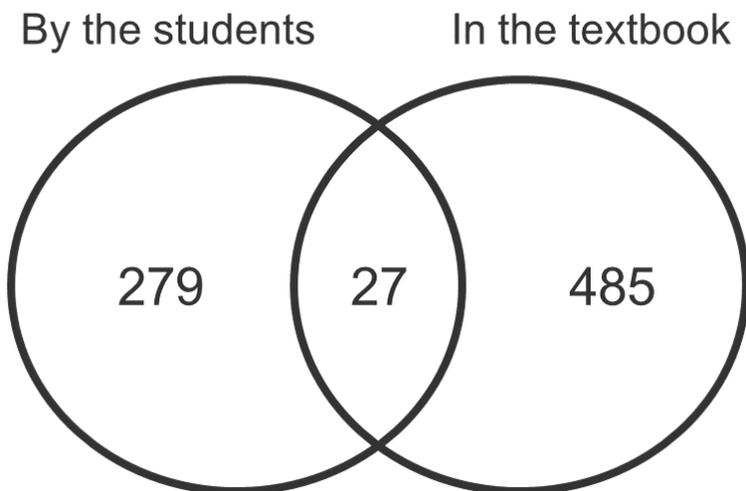
By the students   In the textbook

279   27   485

*Figure 2.*Comparison of the vocabulary lists created by the students and described in the textbook

algorithm	bit	input	byte	error	pixel	CPU
storage	bus	register	memory	cache	decode	volatile
scanner	flash memory	interface	monitor	mouse	smartphone	screen
hardware	speech recognition	layer	search	OCR	simulation	

Figure 3. A list of terms selected by the students overlapping the vocabulary in the textbook

algorithm	bit	input	byte	error	pixel	CPU
storage	bus	register	memory	cache	decode	volatile
scanner	flash memory	interface	monitor	mouse	smartphone	screen
hardware	speech recognition	layer	search	OCR	simulation	

Figure 4. A list of terms selected by the students that did not overlap the vocabulary in the textbook

This observation demonstrates that autonomous vocabulary building is capable of developing a word list including keywords from new technology in information science. One of the advantages of this strategy is that teachers do

not have to prepare the comprehensive vocabulary list beforehand. In reality, it is impossible for English teachers to cover the wide range of information science fields. The student-centered activity is not only effective for students to learn vocabulary but also practical for teachers to provide keywords from new technology.

To evaluate the course level for the students, the instructor conducted a survey with a questionnaire to ask students about the difficulty of the class activity (Figure 1). Students felt the class level was neither too easy nor too difficult for most topics (Table 2).

It should be noted, however, that the one-way ANOVA yielded significant variation in difficulty among topics, $F(9, 744) = 9.226$, $p < 10\text{-}12$. The post hoc Tukey test demonstrated that quantum computing was challenging for some students with a significance level at $p < 0.001$. In sum, whereas the topics and other class design, such as the progression and structure of content, should be

Table 2

Results of the Questionnaire about the Level of the Course

Topic	Too easy	Easy	Just right	Difficult	Too difficult
Digital logic	2 (3%)	22 (28%)	41 (51%)	13 (16%)	2 (3%)
Data representation	2 (3%)	12 (15%)	49 (61%)	16 (20%)	1 (1%)
Processor	1 (1%)	15 (19%)	46 (58%)	18 (23%)	0 (0%)
Memories	2 (3%)	21 (26%)	46 (58%)	9 (11%)	2 (3%)
Input/Output device	4 (5%)	18 (23%)	44 (55%)	11 (14%)	3 (4%)
Artificial intelligence	2 (3%)	20 (26%)	42 (54%)	13 (17%)	1 (1%)
Deep learning	2 (3%)	7 (9%)	41 (51%)	27 (34%)	3 (4%)
Computer vision	0 (0%)	11 (14%)	45 (56%)	24 (30%)	0 (0%)
Robotics	1 (3%)	9 (24%)	21 (57%)	6 (16%)	0 (0%)
Quantum computing	0 (0%)	4 (5%)	28 (35%)	37 (47%)	10 (13%)

carefully chosen, as a whole, it appears that the topics selected in the classes suited the students.

Discussion

This article has reported results from a survey in ESP classes that attempted to build a vocabulary list for information science with an autonomous class activity at a Japanese university. The list developed in classes included both essential keywords and state-of-the-art terminology. Furthermore, the results from the questionnaire show that the topics and rigor of the autonomous class activity matched the level of the students in their second year or above majoring in computer and information science. These observations indicate that building vocabulary through students' autonomous activity could constitute one learning method to keep up with the ever-growing vocabulary in information technology. This strategy will be applicable to vocabulary building in other fields evolving at a fast pace where English teachers cannot keep up with the state-of-the-art terminology.

A couple of issues remain unexplored. First, the basic vocabulary used in this article was from a book published over 10 years ago (Hirai, 2012). That age may cause an underestimation of the number of keywords between the students' choices and the basic vocabulary. Thus, an updated essential vocabulary list will be required for adequate evaluation of the word list. Second, although several state-of-the-art words were found in the students' list, the quality of the words needs to be verified. To implement autonomous vocabulary building in classes, it would be necessary to consider a way to verify the adequacy of the keywords in the list. One possible strategy to guarantee the quality of words would be teachers' helpful intervention so that the students can recognize what vocabulary is more important. Lastly, building vocabulary lists is not the goal and should be followed by the next step—to design effective and practical methods to use the vocabulary list aiming at the improvement of students' proficiency. The utility of the vocabulary list should be amplified when it is used with proper strategies. Adopting efficient learning strategies can enable students majoring in information science to master the requisite vocabulary in their ever-growing

field. In particular, the students' ability to update their vocabulary is undoubtedly essential after they graduate from university.

Conclusion

An autonomous class activity was implemented in ESP classes in a Japanese university to develop a functional vocabulary list in the field of information science. The list built through the classes included both essential and state-of-the-art vocabulary in information science. This pedagogical intervention will help teachers and students in design and implement effective procedures to acquire the necessary vocabulary with greater facility in the actual application outside of the classroom environment. Having an ability to learn novel vocabulary by themselves is a critical skill to survive in an ever-growing and competitive world.

References

Browne, C., Culligan, B., & Phillips, J. (2013). The new academic word list. http://www.newgeneralservicelist.org.

Comer, D. (2017). *Essentials of computer architecture* (2nd ed.). CRC Press. https://doi.org/10.1201/9781315226262

Esteras, R., & Fabre, E. M. (2007). *Professional English in use ICT: For computers and the Internet.* Cambridge University Press.

Hirai, M. (2012). *Kikutansaiensu joho kagaku-hen* [Kikutan science information science edition]. Alc.

Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*(1), 403–430. https://nflrc.hawaii.edu/rfl/item/43

Jin, Z., & Webb, S. (2021). Does writing words in notes contribute to vocabulary learning? *Language Teaching Research*. https://doi.org/10.1177/13621688211062184

Nassaji, H. (2006). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Modern Language Journal*, *90*, 387–401. https://doi.org/10.1111/j.1540-4781.2006.00431.x

Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9–13. https://doi.org/10.26686/wgtn.12552197.v1

Nielsen, M. A., & Chuang, I. L. (2010). *Quantum computation and quantum information* (10th ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511976667

Russell, S., & Norvig, P. (2022). *Artificial intelligence: A modern approach* (4th ed.). Pearson. http://aima.cs.berkeley.edu/

Thornbury, S. (2002). *How to teach vocabulary*. Pearson.

Wang, K.-H. (Isobel). (2015). The use of dialogic strategy clusters for vocabulary learning by Chinese students in the UK. *System*, *51*, 51–64. https://doi.org/10.1016/j.system.2015.04.004

Zhang, S., Xu, H., & Zhang, X. (2021). The effects of dictionary use on second language vocabulary acquisition: A meta-analysis. *International Journal of Lexicography*, *34*, 1–38. https://doi.org/10.1093/ijl/ecaa010

Author bio

Hiroshi Kohsaka is an associate professor at the University of Electro-Communications, Tokyo, Japan. He has an interest in developing materials for English for science and technology. kohsaka@uec.ac.jp

C U E OnCUE Journal Call for Papers

Aims: To provide a forum for the presentation and discussion of a broad range of topics relevant to college and university educators.

Articles Sought

Submissions

Send your proposed manuscripts and questions about proposals to the following section editors.

Features

APA-referenced articles relating to language education at the tertiary level (2,000-4,000 words). Criteria for Feature Articles

- Consideration of issues likely to be peceived by college and university eductors as relevant to language teaching in Japan.
- Well-designed and well-reported quantitative and qualitative research.
- Writing situated within the context of relevant previous work, adding to or commenting on the current body of research.
- Thought-provoking theoretical papers with clear, practical implications.

Inquiries: Glen Hill
  <glenahill@gmail.com>

Research Digest

Research Digest articles are perspectives on applying theory to practice in SLA or a related field. The format includes reviews of research studies, analyses or personal reflections of research methodologies, and reviews of resources for teachers and researchers. Up to 2,500 words.

Inquiries: Joseph Falout
  <researchdigest@gmail.com>

Professional Development

The term Professional Development (PD) covers issues including vocational education, training, employment, conducting research, publishing, presenting, resume-building and professionalism, among many others. Articles of approximately 1,000-2,500 words may be reflective, opinion-based or empirical (qualitative and/or quantitative) studies pertaining to the above.

Inquiries: Paul Moritoshi
  <oncuejournalpd@gmail.com>

Opinion and Perspective (including Readers Respond and Right to Reply)

Short opinion pieces, interviews, perspectives on teaching and learning, point-counterpoint, responses to any articles in previous issues. Also, brief articles about teaching/learning experiences that offer insights into the process of professional development. Articles up to 3,000 words.

Inquiries: Keith Ford
  <fordkeith@hotmail.com>

@CUE

Papers on CALL-related research, technology and technology-related classroom practices (1,000-2,500 words).

Inquiries: Branden Kirchmeyer
  <brandenk@m.sojo-u.ac.jp>

Book Reviews

Reviews of books, textbooks, videos, presentations, etc. 600 words; 1,500 words for scholarly review essays.

Inquiries: Steve Fukuda
  <steve@koshigaya.bunkyo.ac.jp>

Conference Reviews

Reviews of academic conferences or workshops inside or outside Japan (500-1,500 words).

Inquiries: Jonathan Brown
  <cueconferencereview@gmail.com>

*All submissions should be prepared in accordance with APA style and CUE documentation guidelines (see homepage and author template).

How can you learn more about the CUE SIG?

We've got the social network for you!

Add some tweets to Twitter on the CUE site:
https://twitter.com/jaltcue

Show your face on CUE's Facebook site:
http://www.facebook.com/groups/48199667152/

Shout yahoo (or anything else) on the Yahoo Groups site:
http://groups.yahoo.com/group/JALTCUESIGmembers/

And, if that isn't enough, check out the article below by Beck and Hillis:
 "...we hope to introduce social networking sites such as FB and Twitter to the uninitiated and to convince readers of *OnCUE Journal* to consider using these tools as a way to become more involved in CUE."

CUE is the biggest JALT SIG.

Let's get together and talk about it!