A **Corpus-based Language Pedagogy** for Pre- and In-service Teachers: Theory, Research and Practice

Qing MA (Angel)

The Education University of Hong Kong

About me

01 — Researcher in CALL

02 — Experimenter of Corpus Technology in Teaching

03 — English Teacher Trainer Encouraging Independent Learning

- Associate Professor at Department of Linguistics and Modern Language Studies, EdUHK

- Programme leader of MATESOL
- University Teaching Award, 2021

- Research interests: vocabulary acquisition, corpus linguistics, corpus-based language pedagogy, CALL and MALL

- Associate Editor for *The International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT) & The Journal of China Association for Computer-Assisted Language Learning (ChinaCALL)*

Outline

Background of corpora in language education

CBLP: theory and construction

A two-step corpus training framework for teachers/student teachers

Two case studies on CBLP

CAP: CBLP resources for supporting teachers/student teachers

Future directions for CBLP research and practice

Approaches to corpora in language education

Direct vs. indirect approaches to corpus applications in language teaching (Leech, 1997)

Indirect approach:

Developing corpus-based dictionaries (e.g. Collins, Macmillan), references or teaching materials (e.g. Biber et al., 1999), or testing materials (e.g., Callies & Götz, 2015; Leńko-Szymańska, 2020)

Direct approach:

Using corpora during classroom teaching (Johns, 1991; Sinclair, 1991)

To date, there is evidence of indirect use (i.e., reference publishing) (Leńko-Szymańska, 2015; McEnery & Xiao, 2011) but little evidence of direct use in classroom teaching (Römer, 2010; Naismith, 2017; Callies, 2019) of corpora in language teaching.)

Direct corpus applications in classroom teaching

In DDL, language learners are capable of becoming "researchers" or "detectors" (Johns, 1991)

Learning through corpora encourages learner discovery skills and inductive learning, thus enhancing their autonomy (Bernardini, 2002; Boulton & Cobb, 2017; Boulton, 2017)

Nature of DDL:

- "cut out the <u>middleman</u> … give the learner direct access to the data" (Johns, 1990, p. 18)

- "the hands-on use of authentic corpus data (concordances) by <u>advanced, sophisticated</u> foreign or second language <u>learners</u> in <u>higher education</u> for <u>inductive, self-directed language learning of advanced usage</u>" (Boulton, 2011, p. 572).

Corpora use in learning: positive outcomes

- "The overall body of empirical research in DDL provides overwhelmingly favourable reactions to DDL on the part of learners". (Boulton, 2010, p. 140)

- The effect size of corpora use on learners' language improvement was large, with Cohen's d ranging from 0.95 to 1.50 (Boulton & Cobb, 2017)

- Corpora use helps students improve various language skills, including **vocabulary** (Ackerley, 2017; Lee & Liou, 2003), **collocations** (Chan & Liou, 2005; Vyatkina, 2016; Wu, 2021), **grammar** (Lin & Lee, 2015; Smart, 2014), and **writing** (Yoon & Jo, 2014; Poole, 2016; Crosthwaite, 2020).

- Most of the above empirical research is conducted by corpus linguists or teacher researchers with corpus expertise.
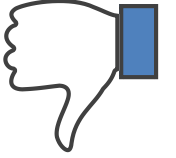
BUT,

The role of corpora in ordinary language teachers' classroom teaching is **minimal** (Boulton, 2017; Callies, 2019; Chambers, 2019).

Factors discouraging teachers' use of corpora in classroom teaching

Predominant corpus applications in higher education settings, e.g., EAP or ESP (e.g., Charles, 2014; Chen et al., 2018; Thurston & Candlin, 1998; Lee & Swales, 2006), by corpus researchers or teachers with strong research interests

Lack of suitable corpora and creative corpus-based activities for young learners (Meunier, 2019); exceptions (Kim, 2019; Crosthwaite & Stell, 2019)

Various difficulties: technical issues, suitable corpora, lack of time and confidence in operating complicated corpora (Leńko-Szymańska, 2017; Naismith, 2017; Poole, 2020; Zareva, 2017) and creation of corpus-based teaching materials

Corpus-based teacher training is largely absent from pre- and in-service teachers' education programmes or professional development (Boulton, 2017; Breyer, 2009; Callies, 2019; Chambers, 2019; Leńko-Szymańska, 2017)

Empirical studies: corpus training for teachers/student teachers

Focusing on teachers/student teachers' perceptions or attitudes toward their use of corpora after training (Breyer, 2009; Ebrahimi & Faghih, 2016; Farr, 2008; Abdel Latif, 2021; Leńko-Szymańska, 2014; Naismith, 2017; Zareva, 2017).

6 student teachers were required to design corpus-based teaching materials (Heather & Helt, 2012), but the data analysis focused on their corpus literacy with limited attention to their pedagogical skills

Leńko-Szymańska (2017) analysed 53 sets of lesson materials designed by student teachers and the results indicated that student teachers lacked pedagogical skills for successful exploitation of corpora in language teaching

After corpus training, teachers/student teachers may perceive corpus as a useful learning tool (e.g., Breyer, 2009; Heather & Helt, 2012) but less likely as a teaching tool.

Formalising corpora use in teaching: Corpus-based language pedagogy (CBLP)

- **What is CBLP?**

Corpus linguistics **CBLP** **Language pedagogy**

> "The ability to use the technology of corpus linguistics to facilitate language teaching in a classroom context" (Ma et al., 2021, p. 4).

Ma, Q., Tang, J., & Lin, S. (2021). The development of corpus-based language pedagogy for Language teachers: A two-step training approach facilitated by online collaboration. *Computer Assisted Language Learning,* 1-30

CBLP: from PCK to TPACK

- **PCK:** "special amalgam of content and pedagogy that is uniquely the province of teachers, their own special form of professional understanding" (Shulman, 1987, p. 8)

- **TPACK:** an extension of PCK, referring to "synthesized form of knowledge for the purpose of integrating ICT/educational technology into classroom teaching and learning" (Chai et al., 2013)

Technology

TPACK

PCK

Corpus Technology

CBLP

Language Pedagogy
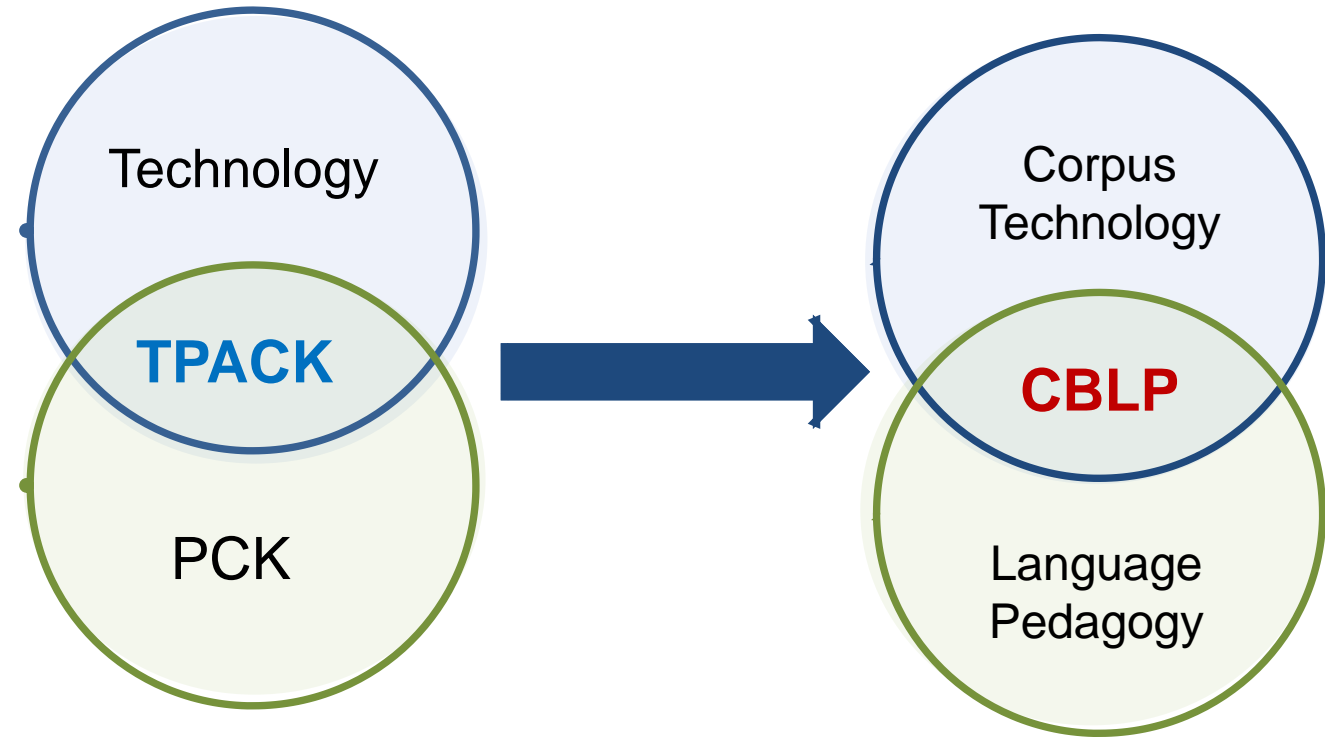
Essential knowledge/skills for conducting corpus-based classroom teaching

- **Corpus literacy**

1. Understanding of corpus

2. Corpus search skills

3. Corpus analysis skills

4. Advantages of corpus use

5. Limitations of corpus use

Mukherjee, 2006, Callies, 2016, Ma et al., 2021

- **CBLP**

- Understanding the purpose of corpus use, corpus linguistics and target students

- Transforming corpus literacy into designing corpus-based teaching materials

- Conducting the CBLP lesson with appropriate instructional strategies

- Evaluate and self-reflect on the CBLP teaching

Based on Shulman' (1987) PCK framework

Two-step CBLP training framework
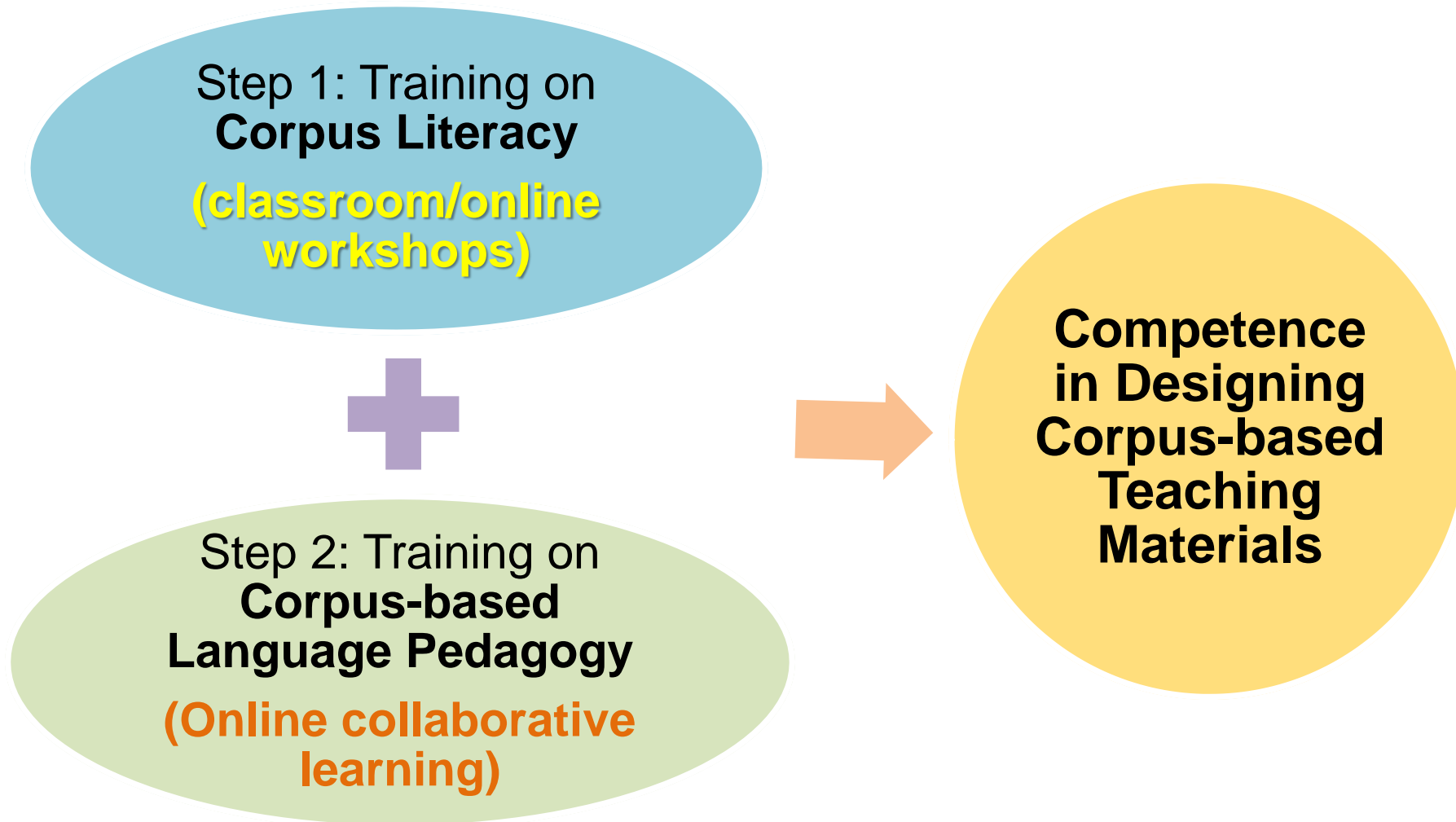
Figure 1. A two-step framework for providing corpus-based teacher training.

Using Interactive Learning to Energise CBLP training

1. Individual learning

2. Intra-group discussion

3. Inter-group feedback

4. Community sharing
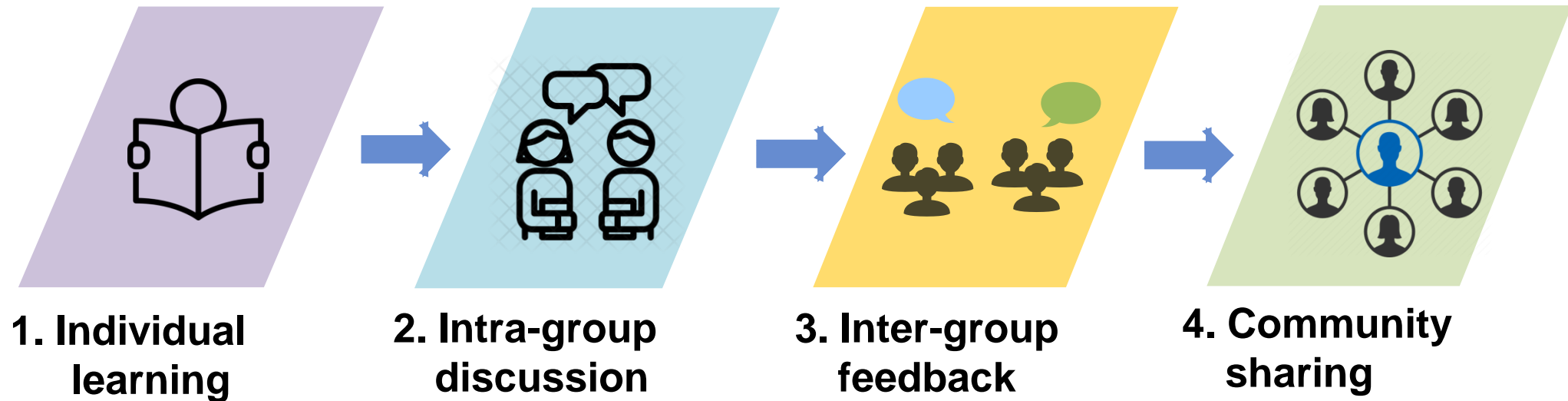
✓ **Effective for both classroom and online CBLP training**

CBLP training steps & procedure

Steps	Learning activities and assessment
Step 1: **Corpus skills** ***(f2f/online teaching)*** Individual - Intragroup -Intergroup – Community	• Lecture input, corpus demonstration • Hands-on tasks on corpus searches • Group brainstorming & presentation
Step 2: **Pedagogical skills** ***(online learning)*** Individual - Intragroup -Intergroup - Community	• Individual tasks: online interactive quiz • Within-group work: collaborative corpus-based lesson design • Intergroup peer feedback: providing comments to different groups • Community sharing

Four-step model for designing CBLP lesson activities (Ma et al., 2021)

1. **Test students' knowledge –** *detect lexical errors/gaps*

2. **Hands-on corpus searches by students –** *look for language patterns*

3. **Inductive discovery by students –** *summarize language patterns*

4. **Output exercise -** *USAGE*

Input →

Understanding →

Integration →

Output

Interaction, Feedback

(Gass, 2001)
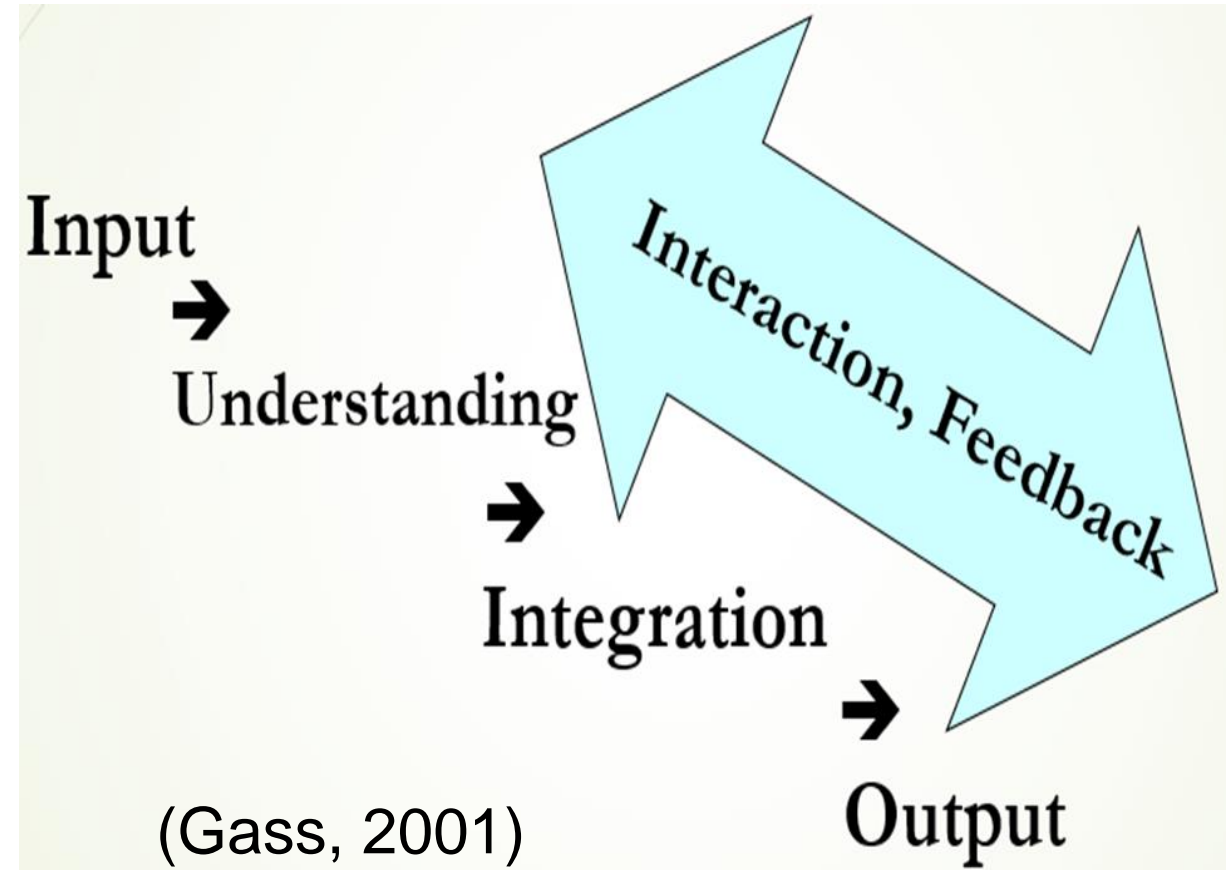
CBLP lessons designed by student teachers (EdUHK)

**Ms. Angel Zhao:
Make vs. Do
Upper Primary Level**
(https://www.youtube.com/embed/ECxqkWXSYSs **)**

**Kevin Zhu: 'Very' and
'Important'
Secondary Level**
(https://www.youtube.com/embed/yilfIBCm80g **)**

CBLP Training Workshop Information (EdUHK)

➢ Since 2017, **we** have conducted **40** workshops to disseminate CBLP in Hong Kong, mainland China and elsewhere

➢ CBLP has reached:

- Teachers (primary, secondary and tertiary): **1000**

- Principals & English panel chairs: **80**

- Schools/universities: **300**

- Students/student teachers: **750**

➢ Integrated into EdUHK courses for student teachers or as stand-alone teacher professional development programmes for in-service teachers

Case Study 1: Research on Student Teachers' CBLP Development

Computer Assisted Language Learning

The development of corpus-based language pedagogy for TESOL teachers: a two-step training approach facilitated by online collaboration

Qing Ma, Jinlan Tang & Shanru Lin

19

Selected Research Questions

To what extent can student teachers develop their **corpus literacy** by participating in the training?

To what extent can student teachers develop their **CBLP**, i.e., designing satisfactory corpus-based lessons?

Methodology: Mixed Methods

- **Participants:** 31 MATESOL students, aged 24-30

- **Context:** CBLP training embedded in a MA course on vocabulary teaching

- **Procedure**

Table 1. Procedure of the two-step corpus-based teacher training.

Time	Training steps	Training activities
Week 1	Step 1 (classroom teaching)	Lecture and workshop (3 hours)
Week 2	Step 2 (virtual classroom hosted on online Moodle platform)	Study pedagogical resources available on the CAP website
Week 3	Step 2(virtual classroom hosted on online Moodle platform)	Complete three individual learning tasks on corpus searches and design principles for corpus lessons
Week 4	Step 2(virtual classroom hosted on online Moodle platform)	Complete one group corpus-based vocabulary lesson design Provide online peer feedback on other group lesson designs

Instruments and Data Collection

Self-designed corpus literacy survey (16 6-point Likert scale questions)

Self-designed evaluation criteria for evaluating CBLP lessons.

8 group CBLP lessons

8 group interviews

Results (Corpus Literacy)

Table 1. Survey results for corpus literacy.

N = 33	No. of items	Sample item	Reliability[a] (Cronbach's α)	Mean (max. = 6)	SD
Understanding of corpora	4	I understand what a concordance line is	0.87	5.39	0.78
Advantages of using corpora	3	I can draw conclusions about language use after searching corpus data	0.89	5.00	0.81
Limitations of using corpora	3	I am aware of the limitations of using corpus data for language learning	0.84	4.76	1.21
Search skills of corpora	3	I know how to search words in corpus data	0.96	5.24	0.78
Analysis of corpus data	3	I will examine the words before or after the keyword in concordance lines	0.77	4.92	0.91

[a] the reliability for each dimension was calculated based on a larger sample size to be reported in a separate study (N = 101).

Results (CBLP)

Table 2. The score and ranking of lesson design (2 raters).

Group	Scores max. = 60 (100%)	Ranking
C	58 (97%)	1
F	56 (93%)	2
D	55 (91%)	3
G	53 (88%)	4
B	51 (85%)	5
E	49 (82%)	6
H	46 (77%)	7
A	43 (72%)	8

Case Study 2: In-service Teachers' CBLP Development

Computer Assisted Language Learning

Routledge
Taylor & Francis Group

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ncal20

Teacher paths for developing corpus-based language pedagogy: a case study

Qing Ma, Rui (Eric) Yuan, Lok Ming Eric Cheung & Jing Yang

To cite this article: Qing Ma, Rui (Eric) Yuan, Lok Ming Eric Cheung & Jing Yang (2022): Teacher paths for developing corpus-based language pedagogy: a case study, Computer Assisted Language Learning, DOI: 10.1080/09588221.2022.2040537

To link to this article: https://doi.org/10.1080/09588221.2022.2040537

Published online: 10 Mar 2022.

25

Selected Research Question

How did two university English teachers develop their CBLP?

Five stages of CBLP

Comprehension ⬅- - - ➡ Understanding of subject knowledge (e.g., CL and English language) and teaching purposes

Transformation ⬅- - - ➡ Transferring comprehended ideas to design of corpus-based teaching activities

Instruction ⬅- - - ➡ Conducting the teaching and sequencing/managing student learning

Evaluation ⬅- - - ➡ Self-evaluating the teaching practice/outcome

Reflection ⬅- - - ➡ Reflecting on the teaching practice (self-criticism)

Adapted from **Model of Pedagogical Reasoning and Action** for PCK (Shulman, 1987, p. 15)

Research context

- **Workshop training** for language teachers

- Teachers designed a **CBLP lesson** (plan, activities, worksheets) addressing their students' needs: sharing and revision

- Invitation to **implement** the CBLP lesson in real classroom teaching

- **Two university English teachers** prepared and conducted CBLP teaching

Participants

Name (pseudonym)	Level of teaching	Teaching experience	Target students	Teaching focus	Familiarity with corpora
Tim *A functional linguist and lover of corpus*	University	10 years	Final year Science students	Academic writing; grammar	Proficient
May *A curriculum reformer and experimenter of pedagogy*	University	15 years	Second year English majors	Critical reading; discourse features of speech text	Nil knowledge of corpora

Method: Case Study

Data sources: (1) CBLP lesson materials; (2) pre-interview (before classroom teaching); (3) lesson observations (during the teaching); (4) post-interview (after classroom teaching)

Analytical framework: comprehension, transformation, instruction, evaluation and reflection (Shulman, 1987)

Data analysis:

Coding & themes

Analytical memo

Cross-case comparison (Creswell & Guetterman, 2019)

Validation through member checking

Results: Tim's path to CBLP
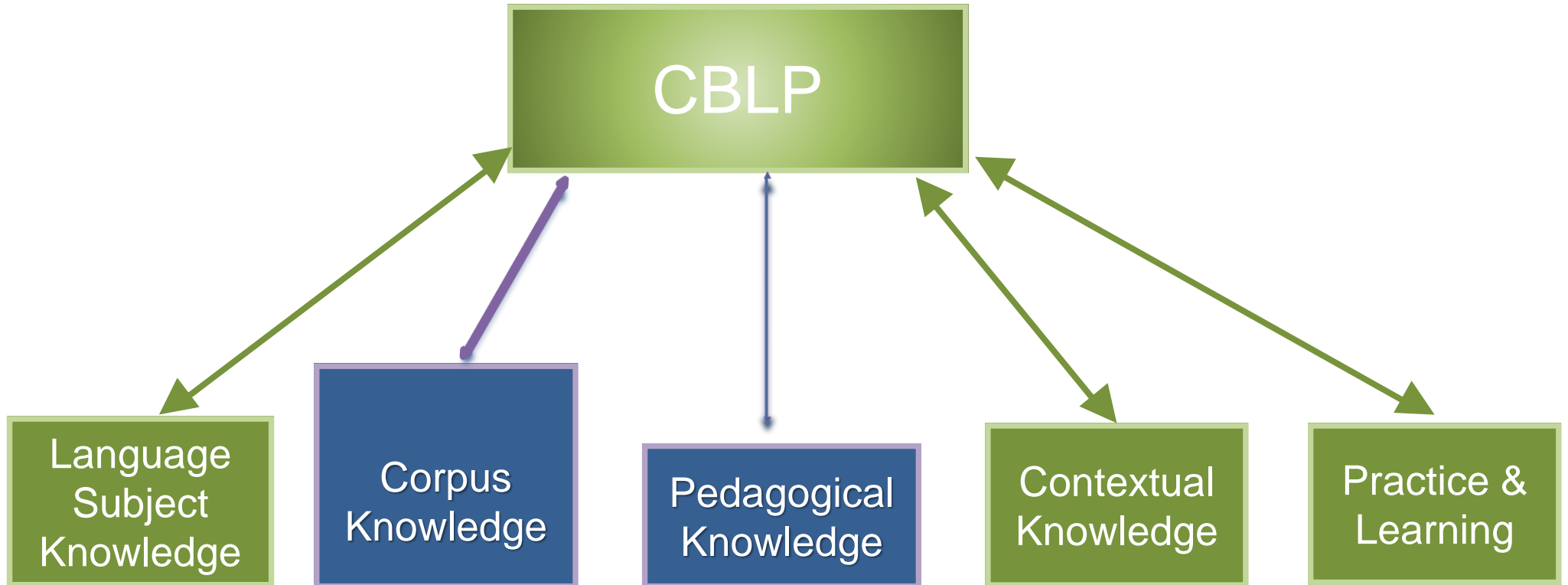
Figure 1: Tim's path to CBLP

Results: May's path to CBLP
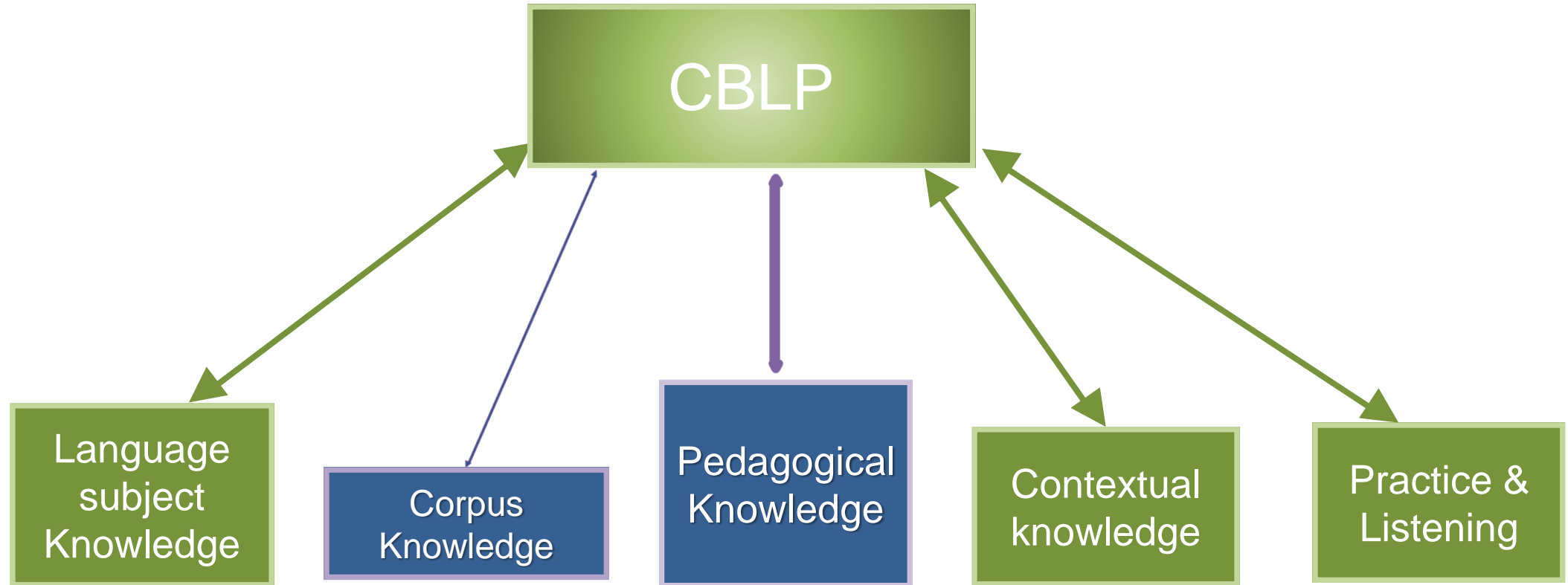
Figure 2: May' path to CBLP

CBLP resources

- **The Corpus-Aided Platform for Language Teachers (CAP)** (https://corpus.eduhk.hk/cap/)

The Corpus-Aided Platform for Language Teachers (CAP)

Home | News ▾ | Teacher Training ▾ | Workshop Series ▾ | Teaching Activities ▾ | Resources ▾ | Contact Us

✓ **Esperanto "Access to Language Education" Award, CALICO, 2020**

✓ **Silver Medal of the 47th International Exhibition of Inventions of Geneva, 2019**

The Award-winning website
The Corpus-Aided Platform

Ease of access

Free and open to all

580 visits per month from more than **20** countries/regions:
China, Hong Kong, Japan, South Korea, Thailand, US, UK, Germany, France, Netherlands, Italy, Turkey, etc.

Teacher feedback:
"I like the Teacher-Training section guiding me at every step of corpus-aided approach with valuable information, and the Teaching Activities section makes me much clearer about how to create corpus-based materials for my language class."

Versatility of the resources

- Short & practical tutorials (texts and videos) on how to use online corpora (e.g., BNC, COCA, Lextutor, etc.)

- **11** self-designed training videos to show how to design corpus-based lessons

- **8** videoed corpus-based lessons conducted in real classrooms (teachers/student teachers)

Breath of the resources

- **64** corpus-based lesson plans (inclu. work sheets)

- Various language skills: *vocabulary, grammar, pronunciation, reading, writing, translation*

- Target student levels: primary, secondary and tertiary

Teacher feedback:
➤ "It encourages students to search and discover vocabulary and grammatical points inductively."

➤ "I personally believe that the main advantage of corpora is to help students with independent learning."

Conclusion & future directions for CBLP research and practice

- ✓ CL and CBLP are two theoretically distinct concepts

- ✓ A two-step framework involving online collaboration is suitable for developing both pre- and in-service teachers' CBLP

- ✓ PCK (e.g., Ma et al., 2021) and TPACK (e.g., Meunier, 2019; Crosthwaite et al., 2021; Ma et al., 2022) approaches are useful for investigating teachers' CBLP development

- ✓ Corpus educators should help teachers and students solve technical issues regarding corpus use (e.g., developing user-friendly corpus technology)

- ✓ A need for investigating how online CBLP teaching could be designed and conducted effectively

- ✓ How to make CBLP lessons more interesting (or attractive to learners)

- ✓ How to motivate busy school teachers to try out CBLP lessons (teacher is the key to pass corpus knowledge to students)

References

Abdel Latif, M. (2020). Corpus literacy instruction in language teacher education: Investigating Arab EFL student teachers' immediate beliefs and long-term practices. *ReCALL*, *31*(1), 34–48

Ackerley, K. (2017). Effects of corpus-based instruction on phraseology in learner English. *Language Learning & Technology, 21*(3), 195–216.

Biber, D., Reppen, R., & Conrad, S. M. (2002). Developing linguistic literacy: Perspectives from corpus linguistics and multi-dimensional analysis. *Journal of Child Language*.

Boulton, A. (2011). Data-driven learning: The perpetual enigma. In S. Goźdź-Roszkowski (Ed.), Explorations across languages and corpora (pp. 563-580). Frankfurt: Peter Lang.

Boulton, A. (2017) Corpora in language teaching and learning. *Language Teaching*, *50*(4), 483–506.

Boulton, A., & Cobb, T. (2017) Corpus use in language learning: A meta-analysis. *Language Learning*, *67*(2), 348–393.

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

Callies, M. (2019) Integrating corpus literacy into language teacher education. In S. Götz & J. Mukherjee (Eds.), *Learner Corpora and Language Teaching* (pp. 245–263). Amsterdam, Philadephia: John Benjamins.

Callies, M., & Götz, S. (Eds.), (2015). *Learner corpora in language testing and assessment* (Vol. 70). John Benjamins Publishing Company.

Chai, S. C., Koh, J. H., & Tsai, C. C. (2013). A review of Technological Pedagogical Content Knowledge. *Educational Technology & Society*, *16*(2), 31–51. Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching*, *52*(4), 460–475.

Chan, T. P., & Liou, H. C. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations. *Computer Assisted Language Learning*, *18*(3), 231–251.

Charles, M. (2014). Getting the corpus habit: EAP students' long-term use of personal corpora. *English for Specific Purposes*, *35*, 30–40.

References

Chen, M., & Flowerdew, J. (2018). A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom. *International Journal of Corpus Linguistics*, *23*(3), 335–369.

Crosthwaite, P., & Stell, A. (2019). It helps me get ideas on how to use my words": Primary school students' initial reactions to corpus use in a private tutoring setting. *Data driven learning for the next generation: Corpora and DDL for pre-tertiary learners*, 150–170.

Crosthwaite, P. (2020). Taking DDL online: Designing, implementing and evaluating a SPOC on data-driven learning for tertiary L2 writing. *Australian Review of Applied Linguistics*, *43*(2), 169–195.

Ebrahimi, A., & Faghih, E. (2016). Integrating corpus linguistics into online language teacher education programs. *ReCALL: the Journal of EUROCALL*, *29*(1), 120.

Farr, F. (2008). Evaluating the Use of Corpus-based Instruction in a Language Teacher Education Context: Perspectives from the Users. *Language Awareness*, *17*(1), 25–43.

Heather, J., & Helt, M. (2012). Evaluating corpus literacy training for pre-service language teachers: Six case studies. *Journal of Technology and Teacher Education*, *20*(4), 415–440.

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

Kim, H. (2019). The Perception of Teachers and Learners towards an Exploratory Corpusbased Grammar Instruction in a Korean EFL Primary School Context. Primary English Education, 25(1), 123–152

Lee, C. Y., & Liou, H. C. (2003). A study of using web concordancing or English vocabulary learning in a Taiwanese higher school context. *English Teaching and Learning*, *27*(3), 35–56.

Lin, M. H., & Lee, J. Y. (2015) Data-driven learning: Changing the teaching of grammar in EFL classes. *ELT Journal*, *69*(3), 264–274.

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for specific purposes*, *25*(1), 56–75.

References

Leech, G. (2014, June). Teaching and language corpora: A convergence. In *Teaching and language corpora* (pp. 1–24). Routledge.

Leńko-Szymańska, A. (2015). A teacher-training course on the use of corpora in language education: Perspectives of the students. *Insights into Technology Enhanced Language Pedagogy. Peter Lang*.

Leńko-Szymańska, A. (2020). *Defining and assessing lexical proficiency*. Routledge.

Ma, Q., Tang, J., & Lin, S. (2021). The development of corpus-based language pedagogy for Language teachers: A two-step training approach facilitated by online collaboration, *Computer Assisted Language Learning*, 1–30.

Ma, Q., Yuan, R., Cheung, L. M. E., & Yang, J. (2022). Teacher paths for developing corpus-based language pedagogy: a case study. *Computer Assisted Language Learning*, 1–32.

Meunier, F. (2019). A case for constructive alignment in DDL: Rethinking outcomes, practices, and assessment in (data-driven) language learning. In P. Crosthwaite (ed.) *Data-Driven Learning for the Next Generation* (pp. 13-30). Baton Rouge, FL: Routledge / Taylor & Francis.

McEnery, T., & Xiao, R. (2011) What corpora can offer in language teaching and learning. *Handbook of research in second language teaching and learning*, *2*, 364–380.

Mukherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the art–and beyond. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpora and language pedagogy: New resources, new tools, new methods*. Frankfurt am Main, Germany: Peter Lang.

Naismith, B. (2017) Integrating corpus tools on intensive CELTA courses. *ELT Journal*, *71*(3), 273–283.

Poole, R. (2016). A corpus-aided approach for the teaching and learning of rhetoric in an undergraduate composition course for L2 writers. *Journal of English for Academic Purposes*, *21*, 99–109.

Poole, R. (2020). "Corpus can be tricky": Revisiting teacher attitudes towards corpus-aided language learning and teaching. *Computer Assisted Language Learning*, 1–22.

References

Römer, U. (2010) Using general and specialized corpora in English language teaching: Past, present and future. In M. C. Campoy-Cubillo, B. Bellés-Fortuño & L. Gea-Valor. (Eds.), *Corpus-based approaches to English language teaching* (pp. 18–35). London: Continuum.

Roszkowski (Ed.), Explorations across languages and corpora (pp. 563-580). Frankfurt: Peter Lang.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–23.

Sinclair, J. (1991) *Corpus concordance collocation.* Oxford: Oxford University Press.

Smart, J. (2014). The role of guided induction in paper-based data-driven learning. *ReCALL*, *26*(2), 184–201.

Thurston, J., & Candlin, C. N. (1998). Concordancing and the teaching of the vocabulary of academic English. *English for specific purposes*, *17*(3), 267–280.

Vyatkina, N. (2016). Data-driven learning of collocations: learner performance, proficiency, and perceptions. *Language Learning & Technology, 20*(3), 159–179.

Wu, P., & Yu, S. (2017). Developing Pedagogical Content Knowledge (PCK) Through Module Team Collaboration: A Case Study of Business English Teachers in China. *The Asia-Pacific Education Researcher*, *26*(1), 97–105.

Yoon, H., & Jo, J. W. (2014). Direct and indirect access to corpora: and exploratory case study comparing students' error correction and learning strategy use in L2 writing. *Language Learning & Technology*, *18*(1), 96–117.

Zareva, A. (2017). Incorporating corpus literacy skills into TESOL teacher training. *ELT Journal*, *71*(1), 70–79.

Thank you!

Q & A

maqing@eduhk.hk