OPEN ACCESS

Conceptual Foundations in Language Research

Conceptual Foundations in Language Research, 1, 260104 (2026)
https://doi.org/10.65553/CFLR.260104

Vibe Coding and Specification-Driven Corpus Linguistics

Wei Lun Wong*

Faculty of Education, Universiti Kebangsaan Malaysia, Malaysia

Abstract

Corpus linguistics has long been shaped by the tools available to its users. Vibe coding, the use of natural-language instructions to develop software with generative AI, loosens that constraint by letting teachers and researchers build the instrument their question requires. This position paper argues that the shift is methodological and proposes specification-driven corpus linguistics as a third stance alongside corpus-based and corpus-driven inquiry. In this stance, an analyst's intuition is written as a natural-language specification, the specification builds the instrument, and the instrument is treated as a revisable hypothesis. The paper develops the Intuition-to-Annotation (I2A) Warrant Chain, a model in which intuition, specification, comparable evidence and contestable annotation lead to a formative decision, each move with its own warrant and failure mode. Synthesising argument-based validation, Leech's maxims of corpus annotation, learner corpus research and formative assessment, it states five testable propositions and extends Leech's maxims to specification transparency. A constructed case of pronoun reference in a Malaysian Year 6 narrative illustrates the model. A four-strand agenda addresses specification fidelity, annotation validity, pedagogical consequence and reporting standards.

Keywords: vibe coding; corpus linguistics; learner corpus; argument-based validation; formative assessment

Submission date: October 28, 2026; Revision date: November 10, 2026; Online: November 16, 2026

Copyright ©2026
The authors

(corresponding author: colinw@ukm.edu.my)

GLODAL

Introduction

An ESL teacher reads *Aina met Mei at the library after school. She gave the librarian a notebook*. The passage sounds natural. However, the teacher hesitates over she. The writer may know exactly who gave the notebook. A reader must infer it. The teacher could circle the pronoun, ask the writer to clarify who acted or explore whether the earlier paragraph makes the reference clear. Each response rests on a different judgement about the same passage. In formative assessment, that difference matters because it changes what the learner is asked to do next.

Corpus linguistics offers a long-standing answer to how such judgements might be grounded. It is to compare the passage with attested language. However, the comparisons a teacher can make have always depended on the software at hand. Anthony (2013) observes that the functions of corpus tools largely determine which research methods are available, and successive generations of tools. McEnery and Hardie (2012) have extended those functions while leaving their design to a small number of developers. A concordancer retrieves *she* efficiently. It does not show which characters were active in the preceding sentence or record why a teacher withdrew a flag. The question is often reshaped to fit the tool.

Formative assessment theory already explains why the next decision matters. Black and Wiliam (2009) locate the formative function in the elicitation, interpretation and use of evidence to improve decisions about learning. Sadler (1989) adds that learners need access to quality criteria, opportunities to compare work with those criteria and strategies for improvement. Studies of data-driven learning (DDL) indicate that corpus consultation can support language learning (Boulton & Cobb, 2017). Extending these principles to discourse-level judgement requires an account of how examples bear on a particular passage, and of how the instrument that retrieves them was built.

Recent research makes this question consequential for corpus methodology. Wang et al. (2026) find that published studies of generative AI (GenAI) in L2 writing instruction concentrate on feedback, revision and polishing. Vibe coding changes who builds the instrument. Crosthwaite and Özer (2026) describe an eight-stage approach to instrument development, illustrated by VetLingua and its corpus-supported writing feedback module. Wong (2026) distinguishes the evidence needed for computational correctness, linguistic validity and pedagogical validity. Pan et al. (2026) illustrate how learners can create tools for their own argumentative writing. These contributions establish how a tool can be built, what claims its use may support, and who can shape its design. What remains untheorised is the methodological status of the specification itself. It is the natural-language account of what a tool should notice, retrieve and display. Treated as an engineering by-product, the specification disappears from the method section. Treated as method, it becomes part of the evidence for every annotation the tool produces.

This paper takes the second view. Its central claim is that vibe coding opens a specification-driven stance in corpus linguistics, in which the analyst's intuition, written as a specification, becomes an inspectable and revisable component of corpus inquiry. Within this stance, an annotation warrants pedagogical action when its interpretation is grounded in an explicit criterion and relevant evidence, when the specification that generated it is recoverable, and when the teacher or learner can revisit the judgement in light of a response. The argument draws on a purposive synthesis of corpus methodology, formative assessment, argument-based validation, learner corpus research and feedback literacy. It focuses on classroom assessment of English as a second language (ESL) writing, particularly discourse features that admit more than one plausible interpretation. The unit of analysis is the annotation episode, which starts with an observation about an identifiable passage and follows it through interpretation, a

response and a decision about the next step.

Conceptual Framework and Synthesis

Specification-driven corpus linguistics

Tognini-Bonelli (2001) distinguished corpus-based work, in which the corpus tests categories the analyst brings to it, from corpus-driven work, in which categories emerge from the data. Both stances assume a stable instrument between analyst and data. Vibe coding removes that assumption. An analyst can now state in ordinary language which contexts to display, which comparisons count and which decisions to record, and a generative model turns that statement into a working tool. The instrument becomes a third object of inquiry, alongside theory and data.

We call this stance specification-driven corpus linguistics. It rests on three commitments. First, intuition is not replaced by corpus evidence but written as a specification that can be tested against it. Second, the specification is versioned and reported as part of the method, because a change in wording can change what the tool retrieves. Third, the validity of a finding depends on a chain of warrants running from the specification to the decision it informs. The stance does not oppose corpus-based or corpus-driven work. It describes what either becomes when analysts build their own instruments (Table 1).

Table 1. *Three stances in corpus linguistics*

Dimension	Corpus-based	Corpus-driven	Specification-driven
Starting point	A prior category or theory	Patterns in the data	An intuition written as a specification
Role of the instrument	Given; selected by the analyst	Given; selected by the analyst	Built from the specification; a revisable hypothesis
What is reported	Corpus and categories	Corpus and procedures	Corpus, specification and version history
Locus of validity	Fit between category and data	Robustness of emergent patterns	A chain of warrants from specification to decision

The Intuition-to-Annotation Warrant Chain

The Intuition-to-Annotation (I2A) Warrant Chain models how a specification-driven annotation comes to justify action. Intuition, specification, comparable evidence and contestable annotation converge on a formative decision. Each move licenses the next only when its warrant holds, and each has a characteristic failure mode (Table 2). The chain adapts Kane's (2013) argument-based approach to unscored classroom decisions and maps onto the computational, linguistic and pedagogical warrants distinguished by Wong (2026). Five propositions state the chain in testable form.

Proposition 1: Teacher intuition as a hypothesis

Experienced teachers often recognise a pattern of difficulty before they can explain it. A tool

can also direct attention to a passage by flagging a pattern specified by the teacher. Either route raises a question about where a reader might hesitate. Diagnosing why requires a further inference. In the opening example, two people in the preceding sentence could plausibly be the referent. The first-mentioned subject may be easier to retrieve, while events described earlier could favour the other person. Research on local discourse coherence helps explain why pronouns must be interpreted in relation to the prominence of their possible referents (Grosz et al., 1995).

The first move turns that recognition into a hypothesis. The teacher identifies a passage, interprets a possible difficulty in relation to a criterion and asks what action the evidence warrants. The task criterion might be that a reader can follow who performs each action. The teacher can then check the surrounding paragraph or ask a peer reader which person they understood. A flag can be withdrawn if the wider text settles the question or if the writer has delayed naming the actor for an intended effect. The proposition predicts that requiring a criterion and a check of the surrounding text will reduce unwarranted flags compared with flagging surface patterns alone.

Proposition 2: Specification as method

The second move is distinctive to the specification-driven stance. Teachers using vibe coding can specify an assessment decision in ordinary language and ask a generative model to build the interface or retrieval function that supports it. For a lesson on reference, a teacher could specify that the tool should display the preceding sentences, preserve names as written, show the context and source of a retrieved comparison, and let a teacher decline a flag. Each clause is a methodological decision. It fixes the window of context, the comparison set and the record kept. Such changes give local educators a means to align the tool with the writing task, a priority in recent work on corpus pedagogy (Curry & McEnery, 2025).

The role of the model during development and its role during classroom use are distinct (Wong, 2026). AI can generate code for a tool that retrieves a stored passage and lets a teacher propose a label. An optional model used at run time can propose a candidate annotation whose wording and interpretation are then reviewed. Retrieval can be tested against known examples and passages with missing context. Model suggestions can be checked against adjudicated cases, including acceptable variants and competing interpretations. Versioned criteria and examples let teachers see which rule produced a result and whether a revision changed earlier behaviour. The characteristic failure is specification drift, in which a revised instruction silently alters which passages are flagged. The proposition predicts that teachers given control over criteria, examples and annotation decisions will adapt the tool more successfully across tasks, with gains contingent on their ability to inspect changes and on the effort required.

Proposition 3: Comparability of corpus evidence

Learner and reference corpora answer different questions. A corpus of comparable learner drafts can show how writers at a similar stage handle reference when two characters are active. A reference collection of successfully interpreted narratives can illustrate several workable strategies. Writers may repeat a name when the actor changes, use a pronoun when the actor is clear or supply a contextual clue. The relevant comparison is between choices made under similar discourse demands. In corpus design, the target population and the variation it contains determine which generalisations the sample can support (Biber, 1993).

A high frequency of *she* in a broad English corpus offers little guidance about this particular referential decision. Comparisons should preserve enough surrounding text to show the candidate referents, and the selected texts should suit the task and its readership. The example set should represent successful local usage, including Malaysian English where it fits the writing

goal. When the corpus offers few comparable cases, the annotation can disclose the gap and draw on a peer reading of the actual draft. Because a vibe-coded tool retrieves only what it is told counts as comparable, comparability must be written into the specification rather than assumed. The proposition predicts that comparable passages with sufficient context will support more defensible feedback decisions than generic concordance lines, even when the number of examples is held constant.

Proposition 4: Contestable annotation

Error-tagged learner corpora offer a basis for analysing recurring difficulties and designing instruction (Granger, 2003). Leech (1993) cautioned that every act of corpus annotation is also an act of interpretation, and that schemes should state how and by whom annotation was carried out. A classroom annotation extends this maxim. It must make enough of the interpretation visible for a learner or another reader to challenge it. A useful record contains the marked passage, the assessment criterion, the proposed interpretation, the evidence and its source, the specification version that produced the flag and a possible next step. It also records whether a reader confirmed the concern, asked for more context or withdrew it. If the available evidence supports two readings that call for different actions, the annotation can stay provisional and ask the writer a question.

The learner should be able to see the reason in ordinary language and recover enough context to respond. In the narrative case, *Who gave the notebook? I could read this as Aina or Mei* opens a question about meaning. Zhang's (2026) account of distributed metacognitive agency provides a way to locate this reasoning across the teacher, the learner and the tool. The proposition predicts that exposing the reason for a flag and inviting a response will increase justified revisions and justified rejections compared with showing a tag and an example alone.

Proposition 5: Formative decision

An annotation becomes formative when it informs a decision about what to do next (Black & Wiliam, 2009). In the example, the writer might replace *she* with *Mei*, keep the pronoun and strengthen an earlier cue, or explain why the passage already works. Teachers may use evidence from several comparable annotations to plan a short lesson on maintaining and shifting reference. A count of flags can reflect how many opportunities the task creates or how readily the tool flags them.

Feedback literacy explains why learners need to practise making and acting on judgements (Carless & Boud, 2018). Learners may engage differently with the same kind of comment (Han & Hyland, 2015). A learner who repairs the highlighted sentence shows successful use of help on that draft. A new narrative with an unmarked reference problem can reveal whether the learner independently notices and handles a similar difficulty. The proposition predicts that guided comparison and a learner decision will produce stronger independent monitoring on a later task than an immediate rewrite for the learner to adopt.

Table 2. *The Intuition-to-Annotation (I2A) Warrant Chain*

Move	Warrant	Failure mode	Observable consequence
1. Intuition	A flag is linked to a task criterion	Flagging surface patterns	Checking the criterion and passage reduces unwarranted flags.
2. Specification	The tool does what its versioned specification states (computational)	Specification drift	Teachers adapt tools across tasks without losing earlier behaviour.
3. Comparable evidence	Examples match the context, readership and function (linguistic)	Frequency substituted for comparability	Matched passages improve the quality of feedback judgements.
4. Contestable annotation	Reasoning is visible and open to challenge	Opaque tagging	Learners can justify both changes and decisions to retain wording.
5. Formative decision	Evidence fits the action taken (pedagogical)	Counting flags as learning	Uptake is visible now; independent monitoring is tested on a new task.

An illustrative case

The constructed scenario begins with the draft excerpt about Aina and Mei. The lesson goal is for another pupil to follow who does what in a short narrative. The teacher's specification is brief: show the two preceding sentences, list every named character, retrieve passages in which two characters are active, and record whether a reader confirmed the concern. The teacher marks *she* for review, reads the surrounding paragraph and asks a peer to identify who handed over the notebook. A comparison passage shows that a pronoun can still work with two characters in the story. *Mei brought a notebook to show Aina. Aina waited by the library door. When Mei reached the desk, she handed the notebook to the librarian.* Here, Mei is named again just before she, and the sequence of actions supports the same reading. Both passages are constructed for illustration.

The annotation shown to the learner

She could refer to Aina or Mei in this paragraph. A classmate read it as Aina. Who did you mean? If you meant Mei, try one way to make that clear and ask a classmate to read the sentence again.

Suppose the writer meant Mei. The writer could revise the second sentence to *Mei gave the librarian a notebook* and explain that the name makes clear who carried out the action. A peer rereading can test whether the revision helped. If the writer shows that earlier context already establishes Mei as the actor, the teacher can reconsider the original flag. When the teacher later adapts the tool to opinion essays, it is the specification, not the code, that is revised, versioned and reported.

Theoretical and Practical Implications

Theoretical implications

The framework contributes to corpus linguistic theory on three fronts: what counts as a methodological object, how analyst and instrument are related, and where validation takes place.

The specification as a methodological object. Methodological debate in corpus linguistics has addressed corpus design (Biber, 1993), annotation (Leech, 1993), the relation between theory and data (Tognini-Bonelli, 2001) and the balance between statistical reporting and linguistic description (Larsson et al., 2022). Each debate made a previously tacit decision reportable. Anthony (2013) extended this logic to software, arguing that tool functionality shapes the methods available. While tools were shared and stable, their limits could be documented once. When each study generates its own instrument from natural language, the specification becomes study-specific and must be reported with the same care as the corpus. Hence, Leech's (1993) maxims can be extended. A vibe-coded annotation should be recoverable to the specification version that produced it, the specification should be published in a form another analyst can rerun, and revisions should be logged with their effect on earlier cases.

Instrumental genesis made explicit. Vérillon and Rabardel (1995) treat an instrument as an artefact joined to its user's schemes of use, formed through instrumentalisation, in which users adapt the artefact, and instrumentation, in which the artefact shapes how users act. With off-the-shelf concordancers, instrumentalisation was confined to options a developer anticipated, and instrumentation went largely unexamined. Vibe coding enlarges instrumentalisation and makes it linguistic, since the teacher's adaptation exists as text. The specification is therefore an analysable trace of instrumental genesis. The same account identifies the risk. A generated tool carries defaults nobody requested, and those defaults can shape what teachers notice. Thus, specification drift is a theoretical concern about how instruments reshape judgement.

Validation located in the annotation episode. Kane's (2013) interpretation builds on Toulmin's (2003) layout of claim, data, warrant and rebuttal, and Chapelle et al. (2008) showed how such arguments can organise validation in language assessment. The I2A chain moves this logic from the test to the single annotation, which places three inferences within one episode. A reader may have trouble identifying a referent. That difficulty may matter for the task criterion. A particular intervention may help the writer respond. Evidence for one inference can leave the next undecided, so the warrant is specific to the action. Rebuttals come from learners themselves when they contest a flag, which makes consequences part of the evidence, in line with Messick's (1989) inclusion of consequences within validity. The chain also gives DDL a defined role in formative assessment: corpus examples support comparison, while the decision rests on how the comparison illuminates a local writing problem (cf. Dong & Wang, 2025).

Practical implications

Teachers can begin by specifying a criterion and collecting a small set of comparable passages that show both clear and difficult cases. They can write the tool specification in the plain language they would use with a colleague and save each version with the examples it was tested on. For pronoun reference, the context usually needs to include at least the preceding sentence and sometimes more of the paragraph. During moderation, teachers can compare why they flagged or retained a form and agree on what further evidence would resolve a disagreement.

An interface can show the learner's text, a short comparison and the question the teacher wants

the writer to consider. It can keep proposed labels editable and retain the reason for a changed decision. Source records can preserve the task and year group while omitting details that identify pupils.

A research agenda for specification-driven corpus linguistics

The agenda is ordered as a trajectory from instrument to discipline. Strands 1 and 2 ask whether vibe-coded tools and their annotations can be trusted; Strand 3 asks whether they change learning; Strand 4 asks how findings can accumulate. Each strand states a guiding question, a design and the direction it opens.

Strand 1. Specification fidelity (computational warrant). *Does the tool do what its specification says, across revisions?* Crosthwaite and Özer (2026) outline a staged workflow for building instruments, and Wong (2026) separates computational correctness from linguistic and pedagogical validity. Fidelity now needs direct testing. Regression sets of passages with known answers, including missing context and local varieties, can be rerun after every revision to detect drift. Ask teachers to adapt an annotation design across two genres, comparing a conventional editor with a vibe-coding environment that allows equivalent changes. Record what teachers adjust, whether the changes still implement agreed criteria, and how much time they spend checking earlier cases. The longer-term direction is a shared benchmark of specification - behaviour pairs against which new tools and models can be compared.

Strand 2. Annotation validity (linguistic warrant). *When are vibe-coded annotations defensible readings of learner text?* Build a held-out set of full paragraphs that includes readily interpretable references, consequential ambiguities and cases that remain uncertain. Ask independent teachers to record the criterion, their reading and which classroom response, if any, they would use. Adjudication should preserve plausible alternative readings rather than force a single gold label, consistent with Leech's (1993) view of annotation as interpretation. Comparison sets should be sampled with the attention to representativeness expected of research corpora (Biber, 1993), and analyses should examine whether local varieties such as Malaysian English are flagged when they suit the task. Report missed concerns, unwarranted flags and deferred decisions alongside teacher checking time.

Strand 3. Pedagogical consequence (pedagogical warrant). *Do annotation episodes build learners' independent judgement?* Meta-analytic evidence shows that DDL can support learning (Boulton & Cobb, 2017), and a thirty-year review of DDL research traces which recommended directions the field has and has not taken up (Boulton & Vyatkina, 2021). Delayed and transfer effects are among those this strand would address. Compare teacher-approved comments that give a label and corpus example with comments that also explain why the example is relevant and invite learners to justify a change or retention decision, followed by a delayed task with a new prompt and no annotation assistance. Learners' accounts of their decisions can show whether feedback literacy develops (Carless & Boud, 2018). The evidence can distinguish improved revision of the marked passage from later monitoring of an unmarked passage.

Strand 4. Reporting standards (disciplinary warrant). *How can vibe-coded studies accumulate?* Replication is rare in second language research, and Marsden et al. (2018) recommend changes to design, infrastructure and incentives to support it. Vibe coding adds an obstacle. Two studies naming the same tool may run different code. Thus, shared conventions should require the specification text, version log, regression set, the run-time role of any model and the human checks applied. Because reporting priorities in corpus linguistics shift over time (Larsson et al., 2022), this strand would also track whether specification reporting becomes

routine and whether it improves replication.

Together, the strands move the field from isolated proof-of-concept tools towards cumulative evidence: shared repositories of specifications, benchmarks and adjudicated cases that allow vibe-coded corpus research to be compared, replicated and taught.

Conclusion

This paper has argued that vibe coding marks a methodological turn in corpus linguistics. When analysts build instruments by describing them, the specification joins theory and data as an object of inquiry, and specification-driven corpus linguistics names the stance that follows. The I2A Warrant Chain shows how an intuition becomes a classroom decision through specification, comparable evidence and contestable annotation, and where that chain can break. The framework's immediate claims concern low-stakes writing decisions about particular features. Extension to classroom speaking or summative scoring would require a different account of the evidence and its use. Future work should pursue the four strands of the agenda, extend the chain beyond referential cohesion to other discourse features and learner populations, and establish reporting conventions for vibe-coded corpus instruments. The main empirical question is whether learners can draw on these annotation episodes to make sound judgements in new writing when the highlighted passage and suggested response are absent.

References

Anthony, L. (2013). A critical look at software tools in corpus linguistics. *Linguistic Research, 30*(2), 141–161. https://doi.org/10.17250/khisli.30.2.201308.001

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing, 8*(4), 243–257. https://doi.org/10.1093/llc/8.4.243

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology, 25*(3), 66–89. https://doi.org/10.64152/10125/73450

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

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (Eds.). (2008). *Building a validity argument for the Test of English as a Foreign Language*. Routledge.

Crosthwaite, P., & Özer, M. (2026). Vibecoding as method: A structured workflow for AI-assisted research instrument development in applied linguistics. *Research Methods in Applied Linguistics, 5*(3), Article 100356. https://doi.org/10.1016/j.rmal.2026.100356

Curry, N., & McEnery, T. (2025). Corpus linguistics for language teaching and learning: A research agenda. *Language Teaching, 58*(2), 232–251. https://doi.org/10.1017/S0261444824000430

Dong, J., & Wang, H. (2025). Exploring the effectiveness of a DDL-GenAI teaching mode in students' EFL writing classroom. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2025.2577369

Granger, S. (2003). Error-tagged learner corpora and CALL: A promising synergy. *CALICO Journal, 20*(3), 465–480. https://doi.org/10.1558/cj.v20i3.465-480

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics, 21*(2), 203–225. https://aclanthology.org/J95-2003/

Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing, 30*, 31–44. https://doi.org/10.1016/j.jslw.2015.08.002

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1–73. https://doi.org/10.1111/jedm.12000

Larsson, T., Egbert, J., & Biber, D. (2022). On the status of statistical reporting versus linguistic description in corpus linguistics: A ten-year perspective. *Corpora, 17*(1), 137–157. https://doi.org/10.3366/cor.2022.0238

Leech, G. (1993). Corpus annotation schemes. *Literary and Linguistic Computing, 8*(4), 275–281. https://doi.org/10.1093/llc/8.4.275

Marsden, E., Morgan-Short, K., Thompson, S., & Abugaber, D. (2018). Replication in second language research: Narrative and systematic reviews and recommendations for the field. *Language Learning, 68*(2), 321–391. https://doi.org/10.1111/lang.12286

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). American Council on Education & Macmillan.

Pan, M., Li, K., & Lai, C. (2026). AI-assisted vibe coding with Base44: Creating personalized tools for argumentative writing. *Assessing Writing, 69*, Article 101076. https://doi.org/10.1016/j.asw.2026.101076

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144. https://doi.org/10.1007/BF00117714

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. John Benjamins. https://doi.org/10.1075/scl.6

Toulmin, S. E. (2003). *The uses of argument* (Updated ed.). Cambridge University Press.

Vérillon, P., & Rabardel, P. (1995). Cognition and artifacts: A contribution to the study of thought in relation to instrumented activity. *European Journal of Psychology of Education, 10*(1), 77–101. https://doi.org/10.1007/BF03172796

Wang, Z., Lei, H., & Chen, K. (2026). The influence of GenAI on second language writing instruction: A systematic review. *System, 140*, Article 104081. https://doi.org/10.1016/j.system.2026.104081

Wong, W. L. (2026). Vibe coding for corpus-based language research and learning: From working tools to warranted claims. *Emerging Educational Technologies*. https://doi.org/10.65553/EET.260105

Zhang, L. J. (2026). Metacognitive agency in language learning: A dynamic, distributed, and language-anchored framework. *Conceptual Foundations in Language Research, 1*, Article 260102. https://doi.org/10.65553/CFLR.260102

Author bio

Wei Lun Wong is a senior lecturer at Faculty of Education, Universiti Kebangsaan Malaysia. His research interests include GenAI learner corpora, teacher education and applied linguistics. His works appear at *International Journal of Applied Linguistics, The Asia-Pacific Education Research, Research in Science and Technological Education, SAGE Open, Applied Linguistics, Innovation in Language Teaching and Learning* and *Scientific Reports* among others.
Email: colinw@ukm.edu.my
ORCID: https://orcid.org/0000-0002-4367-0157