Asia Pacific Journal of Language, Culture, and Education
Vol. 1, No. 1

http://www.apalce.org
ISSN 3092-362X
https://doi.org/10.23403/apjlce.2026.1.1.1

Corpus Tools in a World of Generative AI: What Role Should They Play?

Laurence Anthony
(Waseda University)

Anthony, L. (2026). Corpus tools in a world of generative AI: What role should they play? *Asia Pacific Journal of Language, Culture, and Education, 1*(1), 1-11.

Generative AI has dramatically transformed the conditions under which language teaching and learning take place, enabling teachers to rapidly develop materials and evaluate learner outputs, and helping learners to produce fluent and well-structured texts with minimal effort. This development raises important questions about the continuing role of corpus tools and methods, which have traditionally supported the analysis and learning of language through engagement with authentic data, and have also served as an in-class learning approach in the form of data-driven learning (DDL). In this paper, I offer a conceptual framework and design-oriented perspective on how the role of corpus tools needs to be reinterpreted and redesigned in this new world of generative AI. I argue that while AI has democratized the production of language, there is a growing need to validate these outputs, especially in highly specialized domains and under-resourced language contexts. At the same time, learners still need to develop the ability to express ideas, structure arguments, and make appropriate linguistic choices, and teachers still require ways to ground evaluation and feedback in evidence. Corpus tools should therefore evolve as complementary systems to AI, supporting evidence-based learning, pattern detection, and the verification of AI-generated output. As an illustrative example of this direction, the paper describes a recent update to AntConc (version 4.4), which integrates AI-supported exploration with traditional corpus analysis.

Keywords: AntConc, Corpus Tools, Data-Driven Learning, Generative AI, Language Teaching, Verification

1. Introduction: Users, Goals, and a Changed World

Corpus tools have long played a central role in language teaching and learning by providing systematic access to authentic language data (Biber et al., 1998). In classroom contexts, they have been used by teachers to inform learning objectives, design materials, and ground feedback in evidence, and by learners as part of data-driven learning (DDL) approaches that encourage noticing, comparison, and independent discovery (Johns, 1991; Boulton & Cobb, 2017; Gilquin & Granger, 2022). Over time, these approaches have been further supported by the development of accessible tools designed specifically for classroom use (Anthony, 2022). In recent years, however, the rapid development of generative AI has begun to reshape the environment in which corpus tools operate. Learners can now generate fluent texts, summarize sources, and reformulate ideas with minimal effort, while teachers can use AI to rapidly develop materials, generate examples, and support evaluation (Kasneci et al., 2023; Law, 2024).

At first glance, these developments may appear to reduce the need for corpus-based approaches. If AI can instantly produce examples, explanations, and extended pieces of writing, then activities such as examining concordance lines or comparing authentic usage may seem unnecessarily slow or demanding. In this sense, generative AI is able to perform many of the functions traditionally associated with corpus tools and methods, including providing examples of usage, highlighting phraseological patterns, and supporting language production. However, this shift raises important

concerns. AI-generated outputs, while often highly plausible, are not inherently grounded in verifiable evidence and may not accurately reflect the conventions of specific discourse communities, particularly in specialized or under-resourced contexts. At the same time, the ability of AI to produce high-quality texts can allow learners to bypass the processes of analysis, comparison, and decision-making that are central to language development.

In this paper, I present a conceptual framework for understanding the role that corpus tools can play in this new AI-rich environment, together with a design-oriented perspective on how they might be developed. I argue that corpus tools should be understood not as alternatives to generative AI, but as complementary systems that address two critical needs. First, they provide a means of verifying AI-generated language against authentic usage, allowing both teachers and learners to ground their decisions in evidence. Second, they support learner development by maintaining engagement with the processes through which linguistic competence is acquired. Building on this perspective, the paper considers how corpus tools might be reinterpreted and redesigned to meet these needs and presents recent developments in AntConc (Version 4.4; Anthony, 2026) as an illustrative example of this emerging direction.

2. Generative AI, Teachers, and Learners

Any reconsideration of corpus tools and methods must begin with the users they are intended to serve. In educational contexts, these users are primarily teachers and learners, whose needs, lacks, and wants shape not only the tools they adopt, but also the design of courses, materials, and evaluation practices. In recent years, generative AI has begun to meet many of the needs that corpus tools were previously used to address. For example, learners can use AI to generate examples of language use, paraphrase expressions, summarize texts, and explore alternative ways of structuring their writing. Similarly, teachers can use AI to develop teaching materials, generate model answers, and produce explanations of language features. In these ways, AI provides immediate and flexible access to functions that once required more time, effort, or technical knowledge through corpus consultation.

Importantly, generative AI can also be used in dialogic ways to support language development. Through iterative prompting, learners can engage in scaffolded interactions in which the system provides explanations, reformulations, and feedback. Teachers can design activities that use AI for guided noticing, Socratic questioning, or formative feedback. Recent work in AI and education suggests that large language models can support learners in interactive and reflective ways, not only as generators of text (Law, 2024; Mannekote et al., 2024). AI can therefore, in principle, support many of the same learning processes that corpus methods have traditionally emphasized.

At the same time, generative AI tools do more than assist; they can also complete substantial parts of the task itself. In many language teaching contexts, learners are asked to produce essays, reports, presentations, or other extended texts in order to develop control over structure, argumentation, and rhetorical choice. AI can support planning, phrasing, and revision, but it can also generate these outputs directly, often at a level that exceeds the learner's current competence. This shift brings clear advantages: it can democratize access to high-quality language support, reduce barriers to participation, and allow learners to focus on higher-level concerns such as ideas and organization, while enabling teachers to prepare materials and feedback more efficiently.

However, these developments also introduce two important problems. The first concerns the reliability of AI-generated output. While such output is often fluent and persuasive, it is not inherently grounded in verifiable evidence. AI systems may produce hallucinated content, reflect biases in training data, or generate language that appears appropriate but does not accurately reflect the conventions of a target discourse community. These issues can be particularly pronounced in specialized domains or in under-resourced language contexts, where training data may be limited or

unbalanced. The second problem concerns learner engagement and development. As AI becomes more capable of producing complete texts, there is a growing risk that learners will bypass the processes that tasks are intended to develop.

A simple analogy helps to clarify this issue. Figure 1 shows an imaginary art class in which a teacher introduces an AI-assisted paintbrush that can support students. Initially, the students use the tool to question what they might paint, but soon they start to ask it to sketch ideas and choose colors for them. Before long, the students are asking the tool to replicate the works of experienced artists, aiming to impress the teacher but at the same time losing any ability to paint on their own. A similar pattern is starting to be observed in writing classrooms. Learners may submit polished texts that appear successful, while bypassing the reasoning, noticing, and decision-making that the task was designed to develop.

Figure 1. Storyboard of an AI-Assisted Paintbrush Dramatically Changing the Nature of an Art Classroom

This does not mean that AI should be rejected, but it does suggest that users need ways to verify AI-generated output before taking responsibility for it, and to integrate AI use with complementary approaches that promote the forms of learning that lie at the core of the language classroom. In the following section, I argue that corpus tools and methods can play an important role in addressing both issues by providing access to authentic language data and supporting evidence-based engagement with language.

3. Corpus Tools for Evidence, Verification, and Learner Development

Generative AI can significantly change how language tasks are completed, but it does not fundamentally change the responsibilities of teachers and learners in terms of ownership. In educational contexts, both teachers and learners remain accountable for the language they produce, select, and evaluate. From this perspective, the continuing value of corpus tools in a world of generative AI can be understood in terms of two closely related functions: verification and learner development. These functions are not entirely separate, but together they help explain why corpus methods remain pedagogically important even when AI can produce fluent and apparently convincing language.

3.1. Corpus Tools for AI Verification

Generative AI produces language that is often fluent, coherent, and rhetorically effective. However, this high quality should not be confused with evidence. In educational contexts, both teachers and learners increasingly need to verify whether AI-generated language is appropriate for a particular genre, audience, discourse community, or institutional setting. This need arises in classroom materials, model texts, assignments, feedback, and revision.

For teachers, the issue is prominent during planning and materials design. AI can generate sample texts, task prompts, explanations of genre conventions, and suggested classroom activities. These outputs may be useful, but they are not guaranteed to reflect actual usage in the target domain. A phrase may sound acceptable but might be rare in authentic texts; an organizational pattern may appear convincing but not be typical of the genre; a stylistic feature may be overgeneralized. Corpus tools allow teachers to check whether suggested forms have grounded truth, are frequent enough to require attention, and are appropriate to the discourse context in question.

For learners, similar issues arise during drafting and revision of their language outputs. A student may use AI to propose alternative phrasings, improve their sentences, or even generate a part of an essay. The resulting text can often appear to be highly polished, but it may still contain unusual collocations, atypical phraseology, or patterns that do not align well with the target genre. Corpus tools provide a way for a learner to investigate these possibilities directly. By examining concordance lines, clusters, collocates, and full-text examples, learners can ask not simply whether a form sounds good, but whether it is actually used in relevant texts and how it functions there.

The pedagogical importance of this process is that it repositions AI output as provisional. Rather than treating generated language as the definitive answer, teachers and learners can treat it as a hypothesis to be checked against observable data. This shift is important because it places judgment back in the hands of the user, shifting the focus from acceptance to evaluation. Corpus tools thus function as a source of evidence, which complements rather than replaces AI. This verification role becomes even more important in specialized or under-resourced contexts. When teachers and learners are working with discipline-specific genres or less well-supported languages, AI output is likely to perform less reliably and hence requires stricter verification. In such cases, a modest but carefully selected corpus can provide a useful basis for decision-making.

3.2. Corpus Tools for Support of Learner Development

Verification is important, but perhaps a more important role of corpus tools is preserving the kinds of learning processes that risk being reduced when AI is used uncritically. If learners can generate acceptable products with minimal engagement, they may have fewer opportunities to analyze language, compare alternatives, and make reasoned decisions for themselves. Corpus methods push learners in the opposite direction. Instead of offering a single ready-made answer, they present multiple examples that need to be interpreted. Learners must notice recurrent patterns, compare possibilities, and decide which forms are appropriate in a given context. This is one reason why data-driven learning (DDL) has proven successful, as it encourages noticing, hypothesis formation, and independent learning (Chambers, 2007; Flowerdew, 2015; Vyatkina, 2016). In a world of generative AI, these processes may become even more important, as learners are increasingly asked to not only produce language, but also to evaluate and justify it.

For teachers, this creates an opportunity to redesign classroom tasks. Rather than banning AI or accepting it uncritically, teachers can ask learners to verify, challenge, and refine AI-generated suggestions through corpus analysis. A student might begin with an AI-generated paragraph, then use corpus evidence to test whether the language is typical of the target genre, whether the phrasing is natural, and whether the rhetorical choices are well aligned with authentic texts. In such a workflow,

AI does not replace learning; instead, it becomes one resource within a broader evidential process.

This suggests the emergence of a new educational competence: evidence-based language evaluation. As AI now makes it relatively trivial to produce outputs that appear acceptable at first sight, learners need to ask whether a form is typical, whether a pattern is supported by authentic examples, and whether a revision reflects actual usage rather than generated plausibility. Corpus tools provide a practical means of developing this competence by connecting language judgment directly to observable data (Mizumoto, 2023).

3.3. Limitations of Corpus Tools and the Role of AI Integration

The case for corpus tools is strong but it should not be overstated. Corpus methods bring their own limitations. For example, they depend on the availability of suitable corpora, which are not equally accessible across languages, institutions, and domains. Also, many teachers lack the time or training required to design corpus-based activities, and learners may find corpus consultation difficult, particularly when it involves unfamiliar interfaces, complex search syntax, or the interpretation of large amounts of data. These challenges have long been noted in the literature on corpus use and DDL (Boulton & Cobb, 2017; Gilquin & Granger, 2022).

In this respect, generative AI can play an important complementary role in the reverse direction. While corpus tools provide access to evidence, AI can help users interact with that evidence more effectively. For example, AI can assist in formulating search queries using part-of-speech tags or based on regular expressions. Similarly, it can help summarize concordance results, highlight key tendencies, and provide explanations that support interpretation. For teachers, this can reduce the effort required to prepare corpus-informed materials. For learners, it can lower the barrier to entry, making corpus exploration more accessible and less cognitively demanding.

More broadly, AI integration offers the possibility of addressing one of the longstanding challenges in corpus-based pedagogy: the balance between authenticity and usability. Corpus tools provide access to authentic data, but this data can be difficult to navigate without guidance. AI can provide that guidance, helping both teachers and learners to move more efficiently from raw data to meaningful insights. In this way, AI does not replace corpus methods, but rather, it increases their usability and pedagogical value.

Taken together, these considerations suggest that corpus tools and generative AI should not be viewed as competing approaches, but as complementary components of a broader ecosystem. Corpus tools provide the evidence and transparency that AI lacks, while AI can provide the accessibility and interpretive support that corpus tools have traditionally struggled to offer. Thus, the challenge is not to choose between them, but to design ways in which they can be used together effectively.

4. AntConc for Verification and Learner Development

The discussion so far has argued that corpus tools retain value in an AI-rich environment through their roles in verification and learner development. The latest version of AntConc (Version 4.4; Anthony, 2026) provides a concrete example of how these roles can be operationalized in practice. Rather than treating corpus tools and generative AI as separate or competing systems, AntConc integrates them within a single environment, allowing teachers and learners to smoothly switch between language generation, inspection, and analysis. In this sense, AntConc offers an early illustration of how corpus tool design might respond to the pedagogical conditions outlined in the previous sections.

4.1. ChatAI as a Tool for AI Literacy and Critical Use

A central feature of the latest version of AntConc is the ChatAI tool, which allows users to interact with multiple AI systems, including both cloud-based models from vendors such as OpenAI, Google Gemini, and Claude, and locally hosted models such as DeepSeek, Gemma, Llama, and Qwen (see Figure 2).

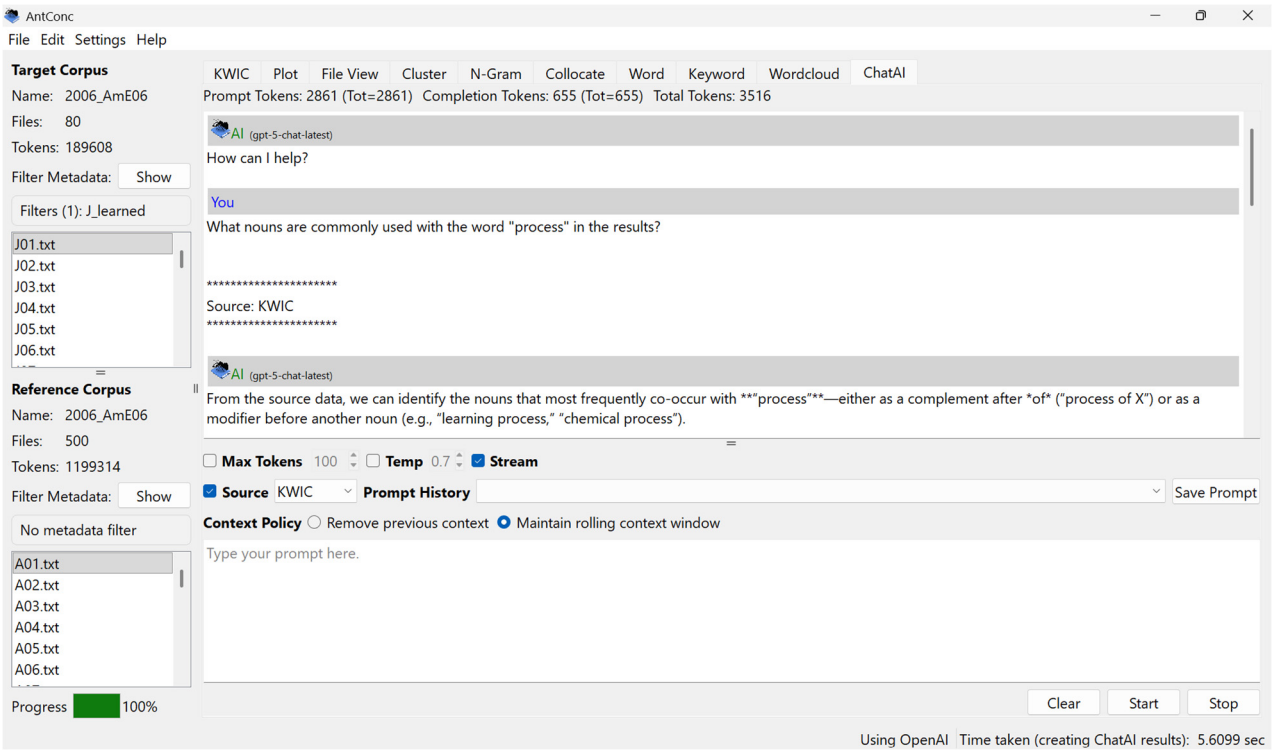

Figure 2. ChatAI Interface for Prompt-Based Output Comparison

From a teaching perspective, this provides a practical way to introduce AI literacy into the classroom. Teachers can demonstrate how different models respond to the same prompt, highlighting variation in language output quality, style, and reliability. For example, a teacher might ask multiple models to generate an introduction to a research report and then compare the results, drawing attention to differences in structure, tone, and disciplinary appropriateness. By adjusting parameters such as temperature, system prompts, and user prompts, teachers can also show learners how outputs can be shaped and how easily they can become overly general, stylistically inconsistent, or rhetorically inappropriate.

The inclusion of local AI models is particularly important in this context. Teachers can use these models to demonstrate how smaller or less capable systems may produce weaker or less reliable outputs, thereby reinforcing the need for verification. At the same time, local models allow for privacy-sensitive use in classroom settings, enabling students to experiment with AI tools without sending data to external services. This is especially relevant in the Asia Pacific region, where educational conditions are highly varied. In some contexts, teachers and learners have ready access to commercial AI systems and stable digital infrastructure. In others, access may be uneven due to cost, privacy concerns, institutional policies, or limited support for local languages and varieties. The region also includes many multilingual and under-resourced language contexts, where the mismatch between AI fluency and local language practice may be particularly pronounced. In such settings, the ability to work with local AI models alongside domain-specific or locally compiled corpora can provide a more appropriate and reliable basis for both language learning and verification.

For learners, ChatAI provides a space to explore AI-generated language critically rather than

passively. Students can generate drafts, alternative phrasings, or explanations, but the tool design encourages them to question the outputs they receive. For example, a student might ask an AI system to generate a paragraph for a presentation, then compare outputs across models and identify features that seem overly generic, stylistically unusual, or inconsistent with the target genre. In this way, ChatAI becomes not only a tool for generation, but also a tool for developing critical awareness of AI behavior.

4.2. Verifying AI Output Through Corpus Evidence

While ChatAI enables the generation and exploration of AI outputs, the core strength of AntConc lies in its ability to verify these outputs against authentic language data. A key design feature is the way ChatAI can be linked with traditional corpus tools, allowing generated text to be examined systematically using corpus methods (see Figure 3).

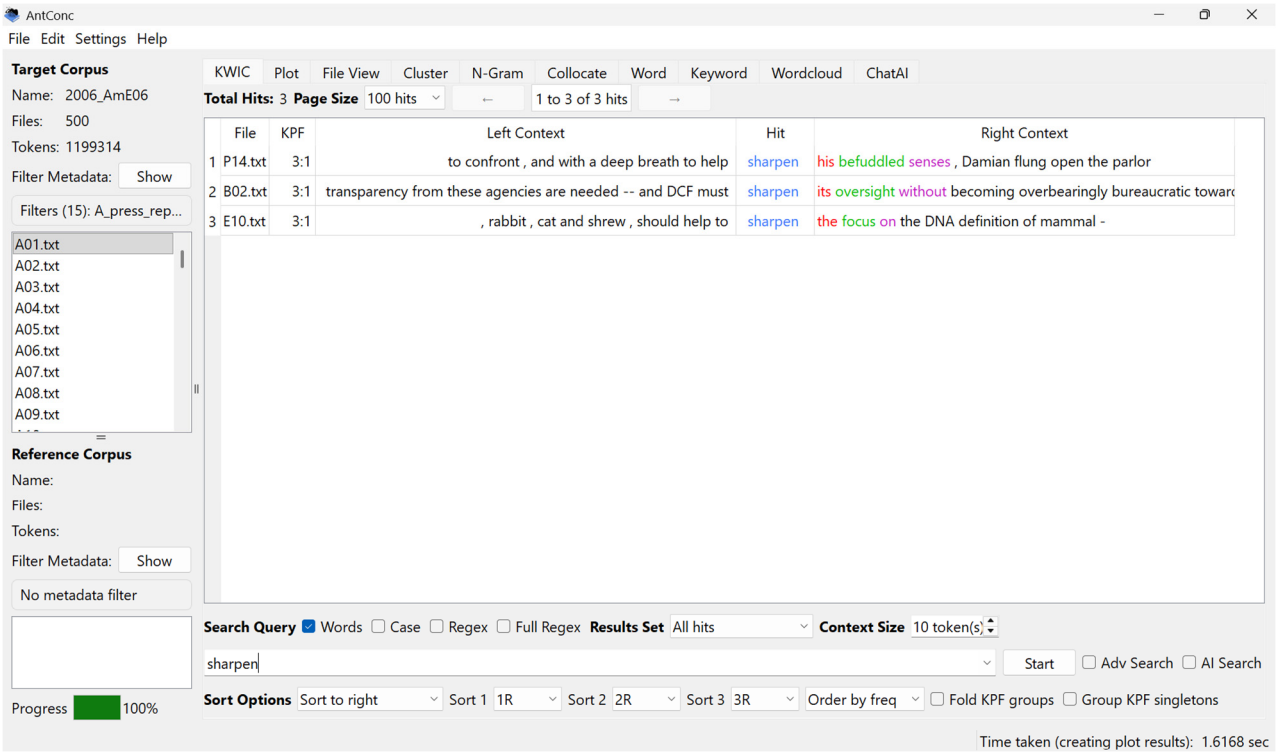

Figure 3. ChatAI and KWIC Workflow for Verifying the AI-Generated Word “Sharpen” in a 1 Million Word Corpus of General English (AmE06)

From a teacher perspective, this supports the validation of teaching materials. For example, a teacher might use AI to generate example sentences for a lesson on academic writing, but then use corpus tools to check whether key phrases actually occur in a corpus of research articles. If the AI suggests expressions that are grammatically acceptable but rarely used, or introduces stylistic patterns such as excessive use of em dashes or vague evaluative verbs such as “sharpen” or “enhance”, these can be identified and revised through corpus analysis. Similarly, teachers can verify whether the structure of an AI-generated text aligns with authentic examples of the target genre.

For learners, the same workflow supports the evaluation of their own work. A student who uses AI to draft an essay or presentation script can check whether specific expressions are typical, whether collocations are appropriate, and whether the overall structure matches that of authentic texts. For example, if an AI-generated report uses an excess of general transitions or repetitive phrasing, the student can examine corpus data to identify more typical alternatives and revise accordingly.

These verification processes are supported by a range of corpus tools within AntConc. The Key-Word-In-Context (KWIC) tool allows users to investigate how words and phrases are used in the local

context. The File View tool provides access to broader discourse context. The Plot tool makes it possible to observe how patterns are distributed across texts. Cluster, N-Gram, and Collocate tools reveal recurring phraseology and typical combinations, and Keyword analysis highlights features that are distinctive of a particular genre or domain. Importantly, these analyses can be grounded either in curated corpora from the built-in corpus repository, including spoken, written, learner, professional, and academic texts, or in user-created corpora that reflect local needs and interests (Anthony, 2022).

This integration of AI and corpus tools encourages users to treat AI-generated output as provisional. Rather than accepting generated text at face value, both teachers and learners are able to test it against authentic evidence, identifying where it aligns with real usage and where it deviates.

4.3. AI-Assisted Data-Driven Learning

In addition to supporting verification, the latest version of AntConc addresses longstanding challenges in data-driven learning (DDL). While DDL has been widely recognized as a powerful approach to language learning, it has also been associated with practical difficulties, including access to suitable data, the formulation of effective search queries, and the interpretation of results. The latest version of AntConc introduces several features designed to lower these barriers.

First, as mentioned above, AntConc includes access to curated corpora, allowing both teachers and learners to work immediately with authentic data without the need to build their own datasets. This is particularly valuable in classroom settings, where time and technical constraints often limit the use of corpus methods.

Second, AntConc includes AI-assisted search functions, which allow users to formulate queries more easily. Rather than relying on prior knowledge of search syntax or regular expressions, users can describe what they are looking for in natural language and allow the system to generate appropriate search patterns. This makes it easier for learners to explore language patterns without being hindered by technical complexity (see Figure 4).

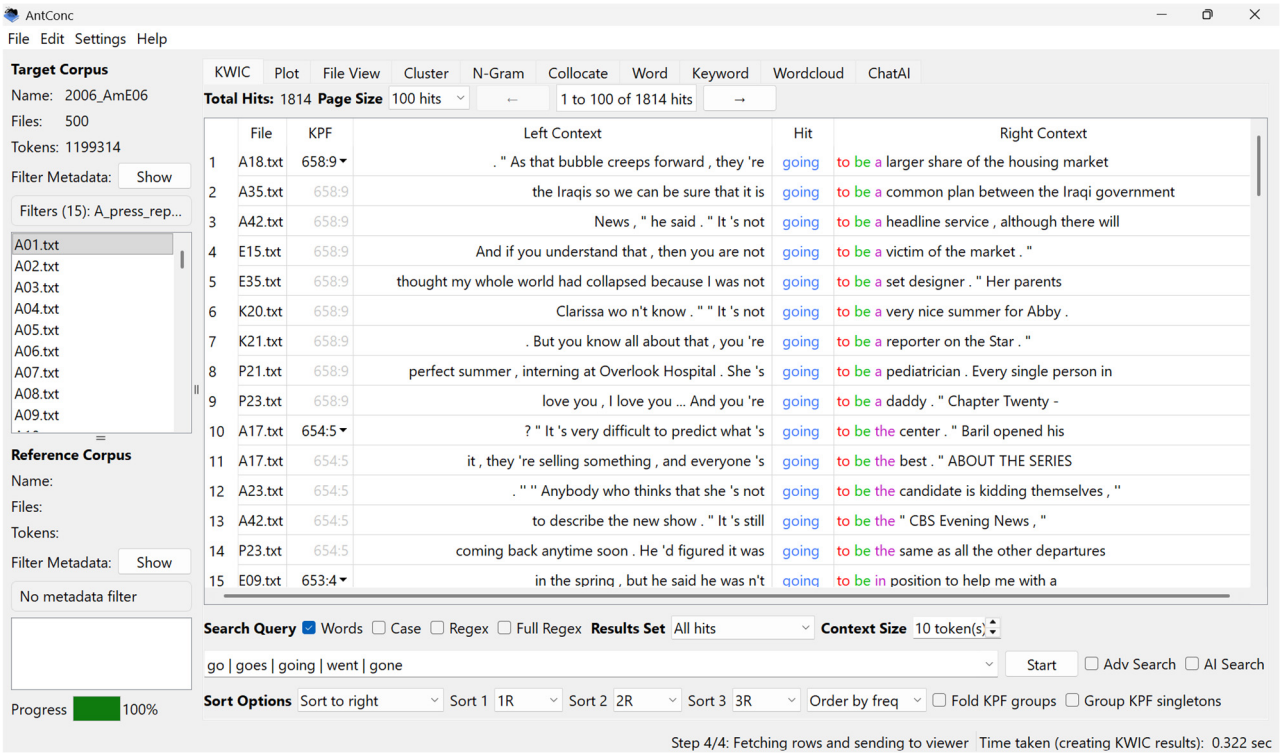

Figure 4. AI-Assisted Search in the KWIC Tool Using the Prompt "Show Me All the Inflections of the Word 'Go'"

Third, AntConc allows results from corpus tools to be passed directly into ChatAI for interpretation. For teachers, this supports classroom preparation by enabling the rapid identification of key patterns, the summarization of findings, and the generation of explanations grounded in real data. For learners, it provides a way to make sense of corpus results, helping them articulate patterns and draw conclusions from evidence.

Crucially, this integration allows for a comparison between general AI outputs and corpus-grounded insights. For example, a learner might first ask an AI system to explain how a particular phrase is used and then compare this explanation with patterns observed in corpus data. Differences between the two can then be discussed and resolved, reinforcing the importance of evidence-based understanding.

4.4. Integrated AI–Corpus Workflows

Taken together, the new features of AntConc suggest a potential future direction for corpus tool design in general. Rather than functioning as standalone systems, corpus tools may evolve to support integrated workflows in which generative AI and corpus methods are used in combination. On the one hand, they can encourage the verification of AI-generated language through access to authentic data. On the other hand, they can support learner development by facilitating the processes of analysis, comparison, and decision-making.

However, it is important to note that some caution is needed. The value of these features is currently easier to argue in design terms than to demonstrate empirically. This paper does not report classroom trials or user studies of AntConc 4.4, although an investigation of how well its design encourages a critical approach to AI use is currently underway. Outcomes of particular interest include whether the tool encourages more evidence-based revision practices, and whether learners develop stronger critical awareness of AI-generated language.

5. Conclusion

The rapid development of generative AI has transformed how language tasks are approached in educational contexts, making it possible to produce fluent and well-structured texts with minimal effort. However, as this paper has argued, these changes do not alter the underlying goals of language teaching and learning. Learners are still expected to develop the ability to express ideas, structure arguments, and make appropriate linguistic choices, and teachers remain responsible for designing learning environments that support this development. What has changed is the relationship between these goals and the processes through which they are achieved.

In this context, the role of corpus tools can be understood in terms of two complementary functions: verification and learner development. As generative AI produces increasingly plausible outputs, both teachers and learners must take greater responsibility for evaluating the language they use. Corpus tools provide a means of grounding this evaluation in authentic data, allowing users to move beyond surface plausibility and verify whether linguistic choices reflect actual usage. At the same time, corpus methods support the processes of noticing, comparison, and evidence-based decision-making that remain central to language learning, helping to ensure that the availability of AI-generated output does not displace the development of underlying competence.

The case for corpus tools should also be kept in proportion. Corpus tools depend on the availability of suitable data, require a degree of corpus literacy, and do not by themselves resolve the broader challenges that generative AI creates for education. Even so, they offer something that AI alone does not, i.e., a clear route back to transparent language evidence. This function becomes particularly important in specialized domains and in contexts where AI output may be less reliable or less well

aligned with actual language use.

The integration of generative AI and corpus methods represents a promising direction for the future design of language tools. As illustrated through recent developments in AntConc, it is possible to bring together AI-supported generation and corpus-based verification within a single environment, supporting workflows that link production with evidence-based analysis. While AntConc provides an early example of this approach, such integration remains relatively uncommon across corpus tools more generally. This suggests an opportunity to align future tool development more closely with pedagogical goals.

More broadly, this discussion points to the importance of developing a new form of competence in a world of generative AI: the ability to evaluate and justify language use through evidence. In contexts where fluent output is readily available, the ability to question, verify, and refine that output becomes increasingly central. Corpus tools, and the methods they embody, provide a practical foundation for developing this competence, keeping teachers and learners close to the evidence needed to verify, interpret, and learn from language in use.

AI Use Disclosure

AI tools were used in the creation of Figure 1. AI tools were also used to check the final manuscript for typographical and formatting issues.

References

Anthony, L. (2022). What can corpus software do? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics*. Routledge Press.

Anthony, L. (2026). *AntConc* (Version 4.4.0) [Computer software]. Waseda University. https://www.laurenceanthony.net/software.html

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press. https://doi.org/10.1017/CBO9780511804489

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

Chambers, A. (2007). Popularising corpus consultation by language learners and teachers. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the Foreign Language Classroom* (pp. 3-16). Rodopi.

Flowerdew, L. (2015). Data-driven learning and language learning theories. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple Affordances of Language Corpora for Data-Driven Learning* (pp. 15-36). John Benjamins. https://doi.org/10.1075/scl.69.02flo

Gilquin, G., & Granger, S. (2022). Using data-driven learning in language teaching. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (2nd ed., pp. 430-442). Routledge.

Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. In T. Johns & P. King (Eds.), *Classroom Concordancing* (pp. 1-16). ELR Journal.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open, 6*, 100174. https://doi.org/10.1016/j.caeo.2024.100174

Mannekote, A., Davies, A., Pinto, J. D., Zhang, S., Olds, D., Schroeder, N. L., Lehman, B., Zapata-Rivera, D., & Zhai, C. (2024). Large language models for whole-learner support: Opportunities and challenges. *Frontiers in Artificial Intelligence, 7*, 1460364. https://doi.org/10.3389/frai.2024.1460364

Mizumoto, A. (2023). Data-driven learning meets generative AI: Introducing the framework of metacognitive resource use. *Applied Corpus Linguistics, 3*(3), 100074. https://doi.org/10.1016/j.acorp.2023.100074

Vyatkina, N. (2016). Data-driven learning of collocations: Learner performance, proficiency, and perceptions. *Language Learning & Technology, 20*(3), 159-179. https://doi.org/10.64152/10125/44487

THE AUTHOR

Laurence Anthony is Professor of Applied Linguistics at the Faculty of Science and Engineering, Waseda University, Japan. He has a BSc degree (Mathematical Physics) from the University of Manchester, UK, and MA (TESL/TEFL) and PhD (Applied Linguistics) degrees from the University of Birmingham, UK. He is a founding member of the Center for English Language Education in Science and Engineering (CELESE), which runs discipline-specific language courses for the 10,000 students of the faculty. His research focuses on language data science, AI, corpus linguistics, educational technology, and English for Specific Purposes (ESP). He received the National Prize of the Japan Association for English Corpus Studies (JAECS) in 2012 for his work in corpus software tools design, including the creation of AntConc.

THE AUTHOR'S ADDRESS

First and Corresponding Author
Laurence Anthony
Professor
Waseda University
3-4-1 Okubo, Shinjuku, Tokyo 169-0072, JAPAN
E-mail: anthony@waseda.jp

Received: 2026.04.18
Received in Revised Form: 2026.06.04
Accepted: 2026.06.23