



DATA-DRIVEN LEARNING AND LEARNER AUTONOMY IN THE AGE OF AI

Amin Dehghan, Ph.D.

Samarkand State Institute of Foreign Languages

Email: aminnima@gmail.com

Abstract: *Data-driven learning (DDL) represents a powerful pedagogical approach in language education, enabling learners to engage with authentic corpus data through exploratory, student-centered activities. By analyzing naturally occurring language with technological tools, learners move from passive reception to active discovery, fostering autonomy and deeper linguistic awareness. This review article synthesizes current research on the intersection of DDL and learner autonomy, highlighting how corpus-based methods and emerging AI technologies reshape the learning process in higher education. It examines opportunities for empowering learners to independently analyze, experiment with, and generate language, while also addressing challenges such as over-reliance on technology and ethical considerations in AI-driven environments. By mapping recent developments and future directions, the article underscores the role of DDL in cultivating autonomous, confident learners equipped to navigate complex real-world contexts in the age of AI.*

INTRODUCTION.

Higher education has undergone significant pedagogical shifts in recent decades, moving away from traditional lecture-based instruction toward more innovative, student-centered approaches. Among these, data-driven learning (DDL) has emerged as a transformative methodology that empowers learners to engage directly with authentic corpus data. By analyzing naturally occurring language through technological tools, students can transit from passive reception to active discovery, fostering autonomy and deeper linguistic awareness (Johns, 1991; Bruner, 1961). This review synthesizes current research on the intersection of DDL and learner autonomy, situating it within the rapidly evolving landscape of artificial intelligence (AI). This review builds on insights first discussed in (Dehghan, 2025) where the foundational framework was introduced. It highlights how corpus-based methods empower learners to independently analyze and generate language, while also considering challenges posed by AI-driven environments.

DDL and Learner Autonomy

Data-driven learning (DDL) positions learners as active investigators of language, echoing discovery learning principles (Bruner, 1961) and the noticing hypothesis. Johns (1991) described students as “language detectives,” uncovering linguistic patterns through corpus consultation. This approach aligns with Nunan’s (1997) model of autonomy, where learners progress from awareness to transcendence, gradually assuming greater responsibility for their learning. Autonomy, defined by Thanasoulas (2000) as the capacity to manage one’s own learning, is operationalized in DDL through direct engagement with authentic data.

Research consistently demonstrates DDL’s effectiveness: it enhances collocation use (Daskalovska, 2015), strengthens academic writing (Chen et al., 2015; Crosthwaite, 2020), supports vocabulary growth (Lee et al., 2020), and improves grammar proficiency (Lin, 2021). Meta-analyses confirm that corpus consultation not only develops linguistic skills but also fosters independence and critical thinking (Boulton & Cobb, 2017; Dong et al., 2022). In this way, DDL shifts learners from passive recipients to autonomous explorers, equipping them with strategies for lifelong learning.

Nunan’s Levels of Autonomy in Practice

DDL activities map neatly onto Nunan’s five levels of autonomy. At the awareness stage, learners observe authentic language patterns using tools such as AntConc (Anthony, 2024), gaining insight into



linguistic structures. Involvement emerges as learners select goals and tasks, for example focusing on collocations in academic writing. Intervention occurs when learners experiment with modifying tasks, filtering corpora, or testing hypotheses. Creation is exemplified by DIY corpus-building, where learners compile texts tailored to their needs, fostering ownership of resources (Charles, 2012). Finally, transcendence is achieved when learners apply findings beyond the classroom, assuming roles as teachers or researchers and contributing to knowledge creation (Boulton, 2010). This progression illustrates how DDL nurtures autonomy through structured yet flexible engagement with authentic data.

DDL and Generative AI: Tensions and Synergies

The rise of generative AI has prompted debate about the future of DDL. While AI models such as ChatGPT offer intuitive interfaces and rapid outputs, they differ fundamentally from corpus consultation. DDL requires learners to engage critically with transparent, authentic data, whereas AI outputs are generated through opaque statistical procedures that may not always be contextually appropriate (Crosthwaite & Baisa, 2023). Moreover, DDL fosters autonomy and higher-order skills, while AI risks encouraging dependency and academic dishonesty through copy-paste practices (Day, 2024a). Corpus tools are designed for pedagogical purposes, whereas AI outputs may expose learners to unreliable or sensitive content.

Nevertheless, synergies exist. AI can lower barriers to corpus use by simplifying complex queries and providing accessible interfaces. It can generate teaching materials, lesson plans, and assessments, complementing DDL's inquiry-based rigor. A hybrid model that integrates AI's efficiency with DDL's emphasis on discovery and critical analysis could create a balanced ecosystem for language learning. Such integration would allow learners to benefit from AI's adaptability while ensuring that autonomy, transparency, and higher-order thinking remain central to the learning process.

Challenges and Ethical Considerations

Despite its promise, integrating DDL and AI raises challenges. Over-reliance on technology risks undermining learner autonomy, while inequities in access to corpus tools and AI platforms may exacerbate digital divides (Day, 2024b). Ethical concerns include plagiarism, bias, and exposure to inappropriate content. Teacher training is also essential to ensure that educators can responsibly integrate both DDL and AI into curricula. Addressing these challenges requires ongoing reflection on how technological innovations align with pedagogical goals and values.

Implications and Next Steps

Future research should explore hybrid pedagogies that combine corpus-based inquiry with AI-driven support, balancing autonomy and accessibility. Advances in corpus visualization, such as collocation maps and interactive dashboards, promise to enhance learner engagement (Sinclair & Rockwell, 2016). Cross-disciplinary applications of DDL, including in medical and technical education, highlight its broader relevance (Dehghan et al., 2024). Ethical frameworks must also be developed to guide responsible AI integration in higher education. By embracing innovation while safeguarding educational ideals, educators can ensure that learners remain active agents in their own learning journeys.

Conclusion

DDL remains a cornerstone of learner autonomy in higher education, offering inquiry-based, transparent, and empowering approaches to language learning. While AI introduces both opportunities and risks, its integration with DDL can create a dynamic, synergistic model that fosters autonomy, critical thinking, and lifelong learning. By continuously monitoring technological impacts and aligning strategies with pedagogical goals, educators can cultivate autonomous, confident learners equipped to navigate complex real-world contexts in the age of AI.

REFERENCES:

1. Anthony, L. (2024). AntConc (Version 4.3.1) [Computer Software]. Waseda University. <https://www.laurenceanthony.net/software/AntConc>
2. Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*,



- 60(1), 534–572. <https://doi.org/10.1111/j.1467-9922.2010.00566.x>
3. 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>
 4. Bruner, J. (1961). The act of discovery. *Harvard Educational Review*, 31(1), 21–32.
 5. Charles, M. (2012). ‘Proper vocabulary and juicy collocations’: EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes*, 31(2), 93–102. <https://doi.org/10.1016/j.esp.2011.12.003>
 6. Chen, M. H., Huang, S. T., Chang, J. S., & Liou, H. C. (2015). Developing a corpus-based paraphrase tool to improve EFL learners’ writing skills. *Computer Assisted Language Learning*, 28(1), 22–40. <https://doi.org/10.1080/09588221.2013.783873>
 7. 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. <https://eric.ed.gov/?id=EJ1261152>
 8. Crosthwaite, P., & Baisa, V. (2023). Generative AI and the end of corpus-assisted data-driven learning? Not so fast! *Applied Corpus Linguistics*, 3(3), 100066. <https://doi.org/10.1016/j.acorp.2023.100066>
 9. Daskalovska, N. (2015). Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning*, 28(2), 130–144. <https://doi.org/10.1080/09588221.2013.803982>
 10. Day, M. J. (2024a). Digital divides in Chinese HE: Leveraging AI as student’s partner to reduce piracy. *Quantum Journal of Social Sciences and Humanities*, 5(1), 165–183. <https://doi.org/10.55197/qjssh.v5i1.343>
 11. Day, M. J. (2024b). Reimagining library learning spaces, or risking digital piracy in universities. *Compass: Journal of Learning and Teaching in HE*, 17(1), 65–81. <https://doi.org/10.21100/compass.v17i1.1492>
 12. Dehghan, A. (2025). Empowering Autonomy in Language Learning: The Transformative Potential of Data-Driven Learning (DDL) in Higher Education. In: Day, M.J. (eds) *Emancipatory Education Without Boundaries in the Age of Neoliberalism, Artificial Intelligence and Digital Learning Platforms*. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-031-97999-6_14
 13. Dehghan, A., Fazilatfar, M. A., & Jabbari, A. K. (2024). The impact of data-driven learning on medical students’ English academic writing. *Journal of Medical Education and Development*, 18(4), 668–676. <https://doi.org/10.18502/jmed.v18i4.15180>
 14. Dong, J., Zhao, Y., Buckingham, L., 2022. Charting the landscape of data-driven learning using a bibliometric analysis. *ReCALL* 1–17. <https://doi.org/10.1017/S0958344022000222>
 15. Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. *ELR Journal*, 4(1), 1-16. <https://doi.org/10.1017/S0267190500001501>
 16. Lee, H., Warschauer, M., & Lee, J. H. (2020). Toward the establishment of a data-driven learning model: Role of learner factors in corpus-based second language vocabulary learning. *The Modern Language Journal*, 104(2), 345-362. <https://doi.org/10.1111/modl.12634>
 17. Lin, M. H. (2021). Effects of Data-Driven Learning on College Students of Different Grammar Proficiencies: A Preliminary Empirical Assessment in EFL Classes. *SAGE Open*, 11(3). <https://doi.org/10.1177/21582440211029936>
 18. Nunan, D. (1997). Designing and adapting materials to encourage learner autonomy. In P. Benson & P. Voller (Eds.), *Autonomy and Independence in Language Learning* (1st ed., pp. 12). Routledge. <https://doi.org/10.4324/9781315842172>
 19. Sinclair, J., & Rockwell, S. (2016). Voyant Tools: A web-based text reading and analysis environment. *Literary and Linguistic Computing*, 31(2), 1–15.
 20. Thanasoulas, D. (2000). Autonomy and learning: An epistemological approach. *Applied Semiotics*, 10(1), 115-131.