Spoken Data-Driven Learning with Learner Corpora: Integrating VR-Based L2 Use and Reflection

Yoshiho Satake

Aoyama Gakuin University

Tokyo, Japan

t31330@aoyamagakuin.jp

Abstract

Data-driven learning (DDL) enables learners to investigate authentic language use through corpus consultation and has demonstrated positive effects across a range of second language learning contexts. However, spoken DDL remains comparatively underexplored because learner speech has rarely been used as the immediate object of corpus investigation. This study proposes that learner-generated spoken corpora created through virtual reality (VR) speaking activities can provide a pedagogical bridge between authentic spoken production and corpus consultation. Twelve Japanese university EFL learners participated in four instructional cycles. Because attendance varied across sessions, analyses were restricted to 13 complete instructional cycles contributed by six learners, in which learners completed a Pre-task VR speaking task, a DDL worksheet based on corpus consultation, and a Post-task speaking task. Speech measures, DDL inquiry types, application of corpus learning, Search Quality, and Noticing were analysed using the same dataset. Results showed that speech rate increased following the complete VR–DDL instructional cycle, while most DDL inquiries focused on collocations and phraseology rather than grammatical correction. Adaptive application of corpus findings occurred more frequently than direct repetition, suggesting that learners flexibly integrated corpus-supported knowledge into subsequent speaking. The findings demonstrate the pedagogical value of using learner-generated spoken corpora to connect immersive speaking tasks with corpus consultation and subsequent language refinement.

Keywords: data-driven learning, virtual reality, spoken corpora, corpus consultation, speaking, EFL

1. Introduction

Data-driven learning (DDL) encourages learners to investigate authentic language through corpus consultation rather than relying solely on teacher correction. Learners formulate questions about their own language, examine concordance lines, compare alternative expressions, and revise their hypotheses on the basis of corpus evidence (Johns, 1991). Meta-analytic research has demonstrated positive effects of DDL across a wide range of second language learning contexts (Boulton & Cobb, 2017; Ueno & Takeuchi, 2023). More recent reviews further argue that the effectiveness of corpus consultation depends not simply on access to a corpus but on meaningful pedagogical integration, carefully designed tasks, and appropriate learner scaffolding (O'Keeffe, 2021; Lusta et al., 2023; Crosthwaite & Steeples, 2024). Corpus consultation has also been shown to support the acquisition of formulaic language and phraseological competence (Geluso & Yamaguchi, 2014).

Despite these advances, spoken DDL remains comparatively underexplored. DDL research has traditionally focused on written learner production, including essays and translation tasks, because written language is relatively easy to collect, revise, and compare (Granger, 2002). In contrast, learners' spoken production has rarely been recycled as the immediate object of corpus investigation. Yet spoken communication naturally generates questions concerning collocations, phraseology, lexical choice, grammatical patterning, and pragmatic appropriateness, making it a promising source of learner-generated corpus data. When learners attempt to describe a landmark, explain a historical site, or recommend a tourist attraction, for example, they may become uncertain about natural lexical combinations, appropriate prepositions, or efficient ways of packaging meaning for oral communication.

A central challenge for spoken DDL is therefore to create meaningful spoken production that can subsequently become the object of corpus inquiry. Virtual reality (VR) offers one possible way of addressing this challenge. VR enables learners to communicate within visually rich environments while observing and responding to places, objects, and events, thereby reducing dependence on written scripts and encouraging situated language use.

Previous research has reported that VR can promote learner engagement, willingness to communicate, and speaking confidence while reducing foreign language anxiety (Dhimolea et al., 2022; Kaplan-Rakowski & Gruber, 2023; Satake et al., 2024; Yudintseva, 2023). The pedagogical potential of VR speaking activities has also been demonstrated in previous studies using Google Street View–based virtual tour tasks, which reported improvements in learner motivation and spoken English performance in Japanese university EFL contexts (Satake, 2024, 2025). Beyond these benefits, VR speaking tasks can generate learner-produced spoken data grounded in immediate communicative needs. Such data can then be

revisited, analysed, and refined through corpus consultation. In this sense, VR is not treated here as an alternative to DDL but as a task environment through which personally relevant spoken language is generated for subsequent DDL.

The present study proposes that learner-generated spoken corpora can function as a pedagogical bridge between VR speaking and corpus consultation. Rather than treating learners' recordings merely as assessment data, the recordings become small spoken corpora from which learners identify personally meaningful language problems. After completing an initial VR speaking task, learners review their recordings, select expressions they wish to improve, consult the Corpus of Contemporary American English (COCA), and then perform the speaking task again. In this instructional sequence, VR generates situated learner language, while corpus consultation transforms selected parts of that language into objects of inquiry and subsequent refinement.

This approach differs from many existing DDL studies in two important respects. First, corpus consultation is initiated by learners' own communicative needs rather than by teacher-selected target items. The linguistic focus therefore emerges directly from problems encountered during meaningful speaking. Second, learning is evaluated not simply according to whether learners reproduce corpus examples verbatim, but by examining how corpus-supported knowledge is subsequently integrated into oral performance. Because spoken communication requires online planning and adaptation, successful learning may be reflected through direct application, linguistic adaptation, functional transfer, or confirmation that an original expression was already acceptable.

At the same time, improvements between repeated speaking performances cannot be attributed automatically to corpus consultation. Task repetition, increased familiarity with the virtual location, and rehearsal may also influence subsequent performance (Bygate, 2001). The present study therefore interprets changes in speech measures as outcomes of the complete VR–DDL instructional cycle rather than as isolated effects of DDL. Item-level analyses of learners' worksheets and subsequent speech are used to examine more directly how corpus-supported knowledge entered the Post-task performances.

Accordingly, this study investigates learner-generated spoken corpora created through repeated VR speaking activities. Particular attention is paid to the linguistic issues learners chose to investigate, the quality of their corpus searches, the nature of the noticing documented in their worksheets, and the ways in which corpus-informed learning appeared in subsequent speaking performances. By integrating speech measures, worksheet evidence, and Post-task performance within the same complete instructional cycles, the study aims to provide a more comprehensive account of how DDL can support spoken language development in an immersive task environment.

The study addresses the following research questions:

(1) How did learners' speech performance change following the complete VR–DDL instructional cycle?

(2) What types of linguistic issues did learners investigate through corpus consultation?

(3) How were corpus-supported findings applied in subsequent speaking performance?

(4) How were Search Quality, Noticing, and subsequent application related?

2. Method

2.1 Participants and Instructional Design

The study was conducted in an undergraduate English course at a Japanese university. Twelve Japanese EFL learners participated in a series of VR speaking activities integrated with corpus consultation during one semester. Four instructional cycles (Sets 1–4) were implemented, each following the same instructional sequence.

Attendance varied across the four instructional cycles, and not every learner completed every stage of every cycle. To ensure that each analysis reflected the complete instructional sequence, analyses were restricted to complete DDL instructional cycles, defined as cycles in which a learner completed (a) the Pre-task VR speaking performance, (b) the DDL worksheet based on corpus consultation, and (c) the Post-task speaking performance. Applying this criterion yielded 13 complete instructional cycles contributed by six learners. All quantitative and qualitative analyses reported below were based exclusively on these 13 complete instructional cycles, ensuring consistency across speech measures, worksheet analyses, and application analyses.

2.2 Instructional Procedure

Each instructional cycle consisted of four stages. Learners wore Meta Quest 3S head-mounted displays and completed the speaking task using the Wander application, which enables users to explore locations through Google Street View in an immersive VR environment.

First, learners entered a VR environment representing a tourist destination or culturally significant location and completed an individual speaking task. They were instructed to describe the location naturally for approximately five minutes while navigating freely within the virtual environment. Learners recorded their performances on their own smartphones. Although the target speaking time was approximately five minutes, the actual recording length varied across learners depending on the pace and extent of their exploration of the virtual environment. These recordings constituted the Pre-task performance.

Second, learners reviewed their own recordings and selected one to three linguistic issues that they wished to improve. Rather than correcting teacher-selected errors, learners determined their own learning targets according to difficulties encountered during speaking.

Third, learners consulted the Corpus of Contemporary American English (COCA) to investigate authentic language use. They searched for lexical items,

collocations, grammatical patterns, or phraseological expressions related to their selected problems and completed a structured DDL worksheet. The worksheet required learners to record (a) the original expression, (b) the reason for selecting the item, (c) corpus search terms and concordance evidence, (d) observations derived from the corpus, and (e) the revised or confirmed expression.

Finally, learners completed the same VR speaking task again while attempting to incorporate insights gained through corpus consultation. This recording constituted the Post-task performance.

The instructional sequence therefore followed the cycle:

VR speaking → self-identification of language problems → corpus consultation → revised VR speaking.

As shown in Figure 1, learners navigated virtual locations while describing landmarks and tourist attractions without using a prepared script.

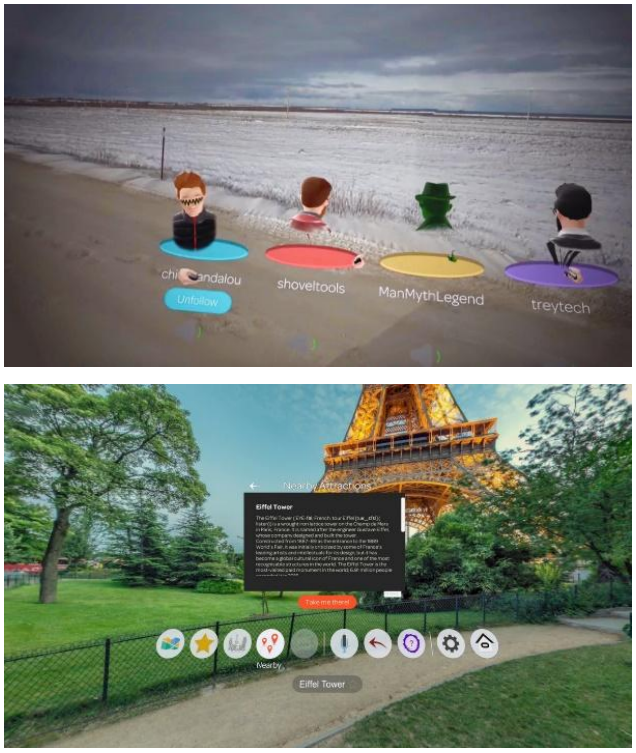

Figure 1. VR speaking environment used during the speaking tasks

Top: Multi-user speaking environment in Meta Quest 3S

Bottom: Wander application based on Google Street View

2.3 Data

Five datasets were constructed from the 13 complete instructional cycles.

The first dataset consisted of speech measures extracted from the Pre-task and Post-task recordings. Because recording durations varied across learners, speech rate measured in words per minute (WPM) was included to facilitate comparison across performances. Measures included duration, total words, word types, type-token ratio (TTR), and WPM.

The second dataset comprised DDL inquiry items extracted from learners' worksheets. Across the 13 instructional cycles, learners generated 30 corpus consultation items, each representing an independently selected linguistic question.

The third dataset consisted of a Reuse Analysis, which traced each worksheet item from the learner's original Pre-task production through corpus consultation to the Post-task performance. This analysis examined not only whether learners reused corpus-supported expressions directly but also whether they adapted or transferred the linguistic knowledge in subsequent speech.

The fourth dataset analysed the quality of learners' corpus searches. Search Quality evaluated whether learners formulated appropriate search terms, obtained relevant corpus evidence, and interpreted concordance lines appropriately.

The fifth dataset examined noticing, operationalised as the observations written in learners' worksheets after corpus consultation. These comments were analysed to determine the depth of linguistic awareness demonstrated during DDL.

2.4 Coding of DDL Inquiry Types

Each worksheet item was first classified according to the primary linguistic issue investigated. Following iterative examination of the worksheet data, five categories were identified:

- Collocation/Phraseology
- Lexis
- Grammar
- Preposition
- Discourse/Pragmatics

Classification was based on the principal focus of the learner's corpus inquiry rather than the final revised expression. When multiple issues were involved, the category that best represented the learner's stated purpose was selected.

2.5 Reuse Analysis

Code	Category	Definition
0	No evidence of application	No evidence that the expression or linguistic knowledge investigated through DDL appeared in the Post-task performance.
1	Attempted application	The learner attempted to apply the corpus finding, but the resulting expression was incomplete, inappropriate, or inaccurate.
2	Direct application	The learner reproduced the revised expression from the worksheet almost verbatim.
3a	Linguistic adaptation	The learner appropriately adapted the corpus-supported lexical,

Code	Category	Definition
		collocational, or grammatical pattern to a different linguistic context.
3b	Functional transfer	The learner addressed the same communicative problem using a different but appropriate expression rather than the worksheet revision itself.
C	Confirmed retention	Corpus consultation confirmed that the original expression was acceptable, and the learner appropriately retained it in the Post-task performance.

Table 1: Coding scheme for Application of DDL Learning

Application of DDL learning was analysed by comparing three sources of evidence: the Pre-task recording, the completed worksheet, and the Post-task recording. Each worksheet item was assigned one of six mutually exclusive codes, as shown in Table 1.

Unlike binary reused/not reused classifications frequently employed in DDL research, this framework recognises adaptive application and justified retention as evidence of learning in spoken production.

2.6 Search Quality

Each corpus search was coded according to the relevance of the search terms, the appropriateness of the corpus evidence, and the extent to which the evidence supported the learner's evaluation or revision, as shown in Table 2.

Code	Description
0	No usable corpus evidence or irrelevant search
1	Weak or poorly targeted search producing limited support
2	Relevant search with partially appropriate evidence
3	Well-targeted search providing clear contextual evidence for evaluation or revision

Table 2: Coding scheme for Search Quality

This analysis focused on learners' corpus consultation behaviour rather than the linguistic outcome itself.

2.7 Noticing Analysis

Learners' written observations in the DDL worksheets were analysed to examine the depth of linguistic noticing, using the coding scheme presented in Table 3.

Code	Description
0	No explicit noticing or unsupported conclusion
1	Surface evaluation (e.g., "more natural", "more common")
2	Explicit comparison or identification of a linguistic pattern
3	Generalisable or context-sensitive linguistic explanation

Table 3: Coding scheme for Noticing Analysis

This framework was intended to distinguish simple evaluative comments from deeper reflections involving form, meaning, frequency, register, or communicative appropriateness.

2.8 Analysis

Speech measures were analysed descriptively by comparing Pre-task and Post-task performances across the 13 complete instructional cycles. DDL inquiry types, Search Quality, Noticing, and Reuse Analysis were summarised using frequency distributions and illustrative examples. Rather than treating verbatim reproduction as the sole indicator of learning, the analyses examined how corpus-supported knowledge was subsequently integrated into learners' spoken production through direct application, linguistic adaptation, functional transfer, or confirmed retention. This integrated analysis allowed the complete instructional sequence—from VR speaking to corpus consultation and subsequent speaking—to be examined within the same instructional cycles.

3. Results

3.1 Speech Performance Across the Complete DDL Cycles

Speech performance was examined using the 13 complete instructional cycles. Table 4 summarises changes in duration, total words, word types, type-token ratio (TTR), and words per minute (WPM) between the Pre-task and Post-task performances.

As shown in Table 4, mean duration increased slightly from 234.8 to 241.1 seconds. Mean total words increased from 255.3 to 292.9, and mean word types increased from 130.3 to 148.2. Mean TTR decreased slightly from 0.516 to 0.505, whereas mean speech rate increased from 67.2 to 74.4 WPM. Thus, the clearest descriptive changes were increases in speech quantity and speech rate, while lexical diversity as measured by TTR remained broadly stable. Eight of the thirteen instructional cycles showed an increase in total words, whereas ten showed an increase in WPM. Because the instructional sequence combined task repetition with corpus consultation, these changes should be interpreted as outcomes of the complete VR–DDL cycle rather than as isolated effects of DDL.

Measure	Pre	Post	Mean change
Duration (seconds)	234.8	241.1	+6.2
Total words	255.3	292.9	+37.6
Word types	130.3	148.2	+17.8
TTR	0.516	0.505	−0.011
WPM	67.2	74.4	+7.3

Table 4: Mean Pre–Post speech measures across the 13 complete DDL cycles

3.2 Linguistic Focus of DDL Inquiry

Thirty learner-selected DDL inquiries were identified across the 13 instructional cycles. Each inquiry was classified according to its principal linguistic focus.

As shown in Figure 2, corpus consultation was dominated by phraseological concerns. Collocation and phraseology

accounted for the majority of learner inquiries, whereas grammar, prepositions, and discourse-related issues were comparatively infrequent. Learners commonly searched for naturally occurring multiword expressions such as be well known for, fun rides, before taking a Shinkansen, packed with people, and enjoy a view of. These findings suggest that learners primarily used corpus consultation to improve the naturalness and communicative appropriateness of spoken descriptions rather than merely to correct grammatical errors.

The predominance of phraseological searches indicates that learners perceived spoken communication primarily as a problem of lexical patterning rather than isolated grammar rules. Instead of asking whether an expression was grammatically correct, learners frequently investigated how native speakers conventionally described places, scenery, historical sites, and tourist attractions.

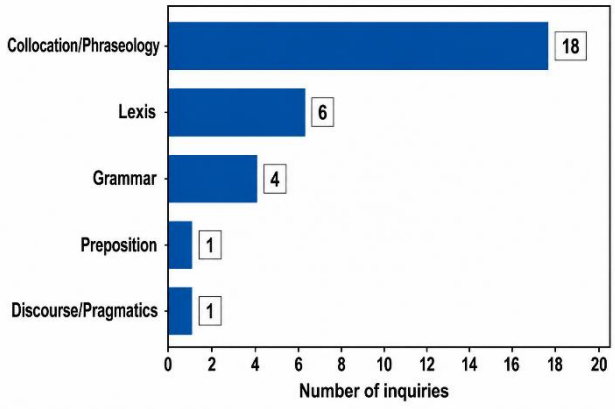

Figure 2. Distribution of DDL inquiry types (n = 30)

3.3 Application of DDL Learning

Each of the 30 worksheet items was traced from the Pre-task performance through corpus consultation to the Post-task performance using the coding framework described above.

Code	Category	n	%
0	No evidence of application	10	33.3
1	Attempted application	1	3.3
2	Direct application	4	13.3
3a	Linguistic adaptation	9	30.0
3b	Functional transfer	4	13.3
C	Confirmed retention	2	6.7
Total		**30**	**100.0**

Table 5: Distribution of Application of DDL Learning codes (n = 30)

Table 5 summarises the distribution of application codes. Overall, 19 of the 30 items (63.3%) demonstrated successful application or confirmed retention (Codes 2, 3a, 3b, or C). Direct application accounted for only four items (13.3%), whereas linguistic adaptation and functional transfer together accounted for thirteen items (43.3%), indicating that learners typically incorporated corpus-supported knowledge flexibly rather than reproducing worksheet expressions verbatim.

Table 6 presents ten illustrative trajectories covering all six application codes. The two Code 0 examples show that learners sometimes retained their original wording despite identifying corpus-supported alternatives. The sole Code 1 item involved an inaccurate attempt to apply the revised comparison. The Code 2 examples demonstrate near-verbatim incorporation of worksheet revisions, whereas the Code 3a items show adaptation of a corpus-supported word or phrase to a new sentence. In the Code 3b examples, learners addressed the original communicative problem through a different appropriate formulation. Finally, corpus consultation confirmed electronic noise as acceptable, and the learner retained the expression in the subsequent performance. Together, these examples demonstrate why a binary reused/not reused distinction would fail to capture the range of outcomes observed in spoken DDL.

These findings indicate that evaluating spoken DDL solely through exact reuse would substantially underestimate learning. Spoken performance required learners to integrate corpus-supported knowledge into newly constructed utterances rather than simply repeating previously prepared expressions.

Code	Selected issue / DDL outcome	Evidence in Post-task performance
0	*Minato Mirai is located near Yokohama Station* → worksheet revision: *situated near Yokohama Station*	The learner retained *located near Yokohama Station*; the corpus-derived alternative was not applied.
0	*Owakudani was created by a volcanic eruption* → worksheet revision: *formed by volcanic activity*	The learner retained *created by a volcanic eruption about 3,000 years ago*.
1	*said to come in as fast as a galloping horse* → *said to come in faster than a galloping horse*	The Post contained *said to come in faster than you learn*, an inaccurate or unclear attempted application.
2	*before Shinkansen* → *before taking a Shinkansen*	*You can buy the Ekiben before taking a Shinkansen.*
2	*cool and fantastic feeling* → *cool and great feeling*	*...blue and white colors to create a cool and great feeling.*
3a	*rides* → corpus example *fun rides*	*...roller coasters and other fun rides...*
3a	*famous for its waterfront scenery* → *well known for its waterfront scenery*	*This area is well known for its waterfront scenery, modern buildings, shopping areas, and entertainment.*
3b	Avoid repetition of *many colors* → *many other types of colors*	The learner did not reproduce the worksheet revision but listed specific colours instead.

Code	Selected issue / DDL outcome	Evidence in Post-task performance
3b	*I very like soccer → I really like soccer*	*I like Camp Nou because the atmosphere is amazing.*
C	*electronic noise* was confirmed as acceptable	*You can also hear machine sounds or electronic noises in the background.*

Table 6: Illustrative Pre–DDL–Post trajectories

3.4 Search Quality

The quality of learners' corpus consultation was analysed for all thirty worksheet items.

Table 7 summarises the distribution of Search Quality codes. More than half of the searches (53.3%) were classified as highly targeted, and a further seven (23.3%) produced generally appropriate corpus evidence despite minor limitations. Consequently, 23 of the 30 searches (76.7%) yielded corpus evidence that was considered useful for evaluating or revising learners' language.

Less successful searches were typically characterised by overly general search terms, misspellings, or searches that produced examples only loosely related to the learner's intended meaning.

Code	Category	n	%
0	No usable search evidence	2	6.7
1	Weak or poorly matched search	5	16.7
2	Relevant but partially targeted search	7	23.3
3	Targeted and contextually appropriate search	16	53.3
Total		30	100.0

Table 7: Distribution of Search Quality codes (n = 30)

3.5 Noticing

Learners' written observations were analysed to examine the depth of noticing that accompanied corpus consultation.

Code	Category	n	%
0	No explicit or supported noticing	7	23.3
1	Surface-level noticing	7	23.3
2	Explicit comparison or pattern noticing	8	26.7
3	Contextualised or generalisable noticing	8	26.7
Total		30	100.0

Table 8: Distribution of Noticing codes (n = 30)

Table 8 summarises the results. More than half of the observations (53.3%) demonstrated explicit linguistic comparison or broader contextual understanding (Codes 2 and 3). Learners frequently commented on collocational preferences, register differences, grammatical environments, or recurring phraseological patterns identified through corpus consultation. These observations suggest that corpus consultation promoted active evaluation of language use rather than simple acceptance of concordance examples.

3.6 Relationships Among Search Quality, Noticing, and Application

Cross-tabulation of the three datasets showed that successful corpus searches did not automatically result in successful application during the subsequent speaking task. Although most high-quality searches (Codes 2 and 3) eventually led to successful application, several well-executed searches nevertheless resulted in Code 0 because the investigated expression did not become relevant during the Post-task performance or because learners selected an alternative formulation while speaking.

Similarly, deeper noticing did not guarantee direct reuse. Instead, learners who demonstrated contextualised understanding often adapted corpus-supported expressions to fit the developing discourse. These findings reinforce the importance of distinguishing adaptive application from verbatim repetition when evaluating spoken DDL.

Overall, the three analyses converged on a consistent interpretation. Learners generally conducted effective corpus searches, documented meaningful linguistic observations, and subsequently incorporated corpus-supported knowledge into their speech. However, this incorporation most often occurred through adaptation or functional transfer rather than exact reproduction of worksheet expressions.

4. Discussion

This study investigated how learner-generated spoken corpora could support data-driven learning (DDL) following VR speaking activities. By restricting the analyses to 13 complete instructional cycles, the study examined the entire instructional sequence—from initial speaking through corpus consultation to subsequent performance—using a consistent dataset. Rather than evaluating speech gains alone, the analyses combined speech measures, learner-selected DDL inquiries, application patterns, search quality, and noticing to provide a more comprehensive picture of learning.

The speech measures showed modest improvements in fluency following the instructional cycle. Mean speech rate, total words, and word types all increased, whereas TTR remained relatively stable. This pattern is consistent with Skehan's (2009) distinction among fluency, complexity, accuracy, and lexis, suggesting that improvement in one dimension of performance does not necessarily entail parallel improvement in another. These findings suggest that learners generally became able to produce more language at a faster rate during the second VR performance. However, because the design incorporated task repetition, review of one's own performance, and corpus consultation, these improvements should not be attributed solely to DDL. Instead, they represent the combined outcome of the complete VR–DDL cycle. This interpretation is consistent with previous research demonstrating that repeated task performance itself promotes greater fluency (Bygate, 2001), while corpus consultation provides opportunities

for refining linguistic accuracy and phraseological choice.

The DDL inquiries themselves provide further insight into learners' language development. Most searches focused on collocations and phraseological expressions rather than grammar. Learners typically investigated expressions such as well known for, fun rides, packed with people, and before taking a Shinkansen, indicating that they were primarily concerned with producing natural and audience-appropriate descriptions rather than simply correcting grammatical errors. This finding supports previous observations that corpus consultation is particularly valuable for developing phraseological competence and formulaic language, areas that are often difficult to master through explicit grammar instruction alone.

Perhaps the most important finding concerns how corpus-supported learning appeared in subsequent speech. Direct reproduction of worksheet revisions accounted for only a small proportion of the analysed items, whereas linguistic adaptation and functional transfer occurred considerably more frequently. Learners often modified corpus-supported expressions to fit new contexts or solved the original communicative problem using different but appropriate language. These findings suggest that evaluating DDL only through exact reuse substantially underestimates learning, particularly in speaking, where language must be produced online in response to communicative demands. The present coding framework therefore provides a broader account of learning by recognising adaptation, transfer, and confirmed retention alongside verbatim reuse.

The analyses of Search Quality and Noticing further support this interpretation. Most learners conducted relevant corpus searches and demonstrated explicit observations regarding lexical, grammatical, or phraseological patterns. Nevertheless, successful searches did not always result in immediate application during the Post-task performance. In several cases, learners appeared to understand the corpus evidence but chose different expressions while speaking or encountered no opportunity to reuse the investigated item. Spoken production therefore involves not only acquiring corpus-supported knowledge but also selecting and integrating that knowledge dynamically during communication. This finding reinforces the value of analysing the entire learning process rather than relying solely on before-and-after performance measures.

The instructional design adopted in this study also contributes to current DDL research. Rather than presenting teacher-selected corpus tasks, learners generated their own questions from difficulties encountered during VR speaking. Their own spoken production thus became a learner-generated spoken corpus that served as the starting point for subsequent inquiry. This sequence created a close connection between communicative need, corpus consultation, and immediate reuse, illustrating how DDL can be embedded naturally within task-based speaking activities instead of functioning as an isolated reference exercise.

Several limitations should be acknowledged. First, the number of complete instructional cycles was relatively small because analyses were restricted to learners who completed the entire instructional sequence. This decision prioritised methodological consistency over sample size. Second, the study did not include a comparison condition without corpus consultation, making it impossible to isolate the contribution of DDL from task repetition. Finally, the analyses focused on short-term application immediately following corpus consultation. Longer-term retention and transfer across different speaking tasks remain important questions for future research.

Despite these limitations, the findings demonstrate the pedagogical potential of learner-generated spoken corpora for integrating VR speaking and DDL. Rather than functioning primarily as a tool for correcting errors, corpus consultation supported learners in refining phraseological choices, developing more natural spoken expressions, and flexibly incorporating corpus-informed knowledge into subsequent communication. The study therefore suggests that learner-generated spoken corpora provide a productive framework for linking immersive speaking activities with evidence-based language learning.

5. Conclusion

This study examined how learner-generated spoken corpora can support data-driven learning (DDL) within a VR-based speaking environment. By analysing 13 complete instructional cycles, the study integrated evidence from speech performance, learner-selected corpus inquiries, worksheet analyses, and subsequent speaking to investigate the complete learning process rather than isolated learning outcomes.

The findings suggest three main conclusions. First, learners generally produced longer and faster spoken performances following the VR–DDL instructional cycle, although these improvements should be interpreted as the combined effects of task repetition, self-review, and corpus consultation rather than DDL alone. Second, learners primarily used corpus consultation to investigate collocations and phraseological expressions instead of isolated grammatical forms, indicating that communicative needs during speaking naturally directed attention toward formulaic language. Third, successful learning was most frequently reflected through linguistic adaptation and functional transfer rather than verbatim reproduction of corpus examples. These findings demonstrate that evaluating spoken DDL solely through exact reuse substantially underestimates how corpus-supported knowledge is incorporated into spontaneous speech.

Methodologically, the study also illustrates the value of analysing learner-generated spoken corpora. Rather than relying on teacher-selected target items, learners generated their own corpus inquiries directly from communicative problems encountered during VR speaking. This learner-centred sequence created an explicit connection between language production, corpus consultation, and subsequent language use, providing a practical model for integrating

DDL into communicative speaking instruction.

Although the number of complete instructional cycles was necessarily limited, restricting the analyses to learners who completed the entire instructional sequence ensured that all reported findings were based on the same dataset. Future research should examine larger samples, include comparison groups without corpus consultation, and investigate whether corpus-supported learning transfers to new speaking tasks over longer periods.

Overall, the present study suggests that learner-generated spoken corpora constitute a promising pedagogical resource for linking VR speaking activities with evidence-based corpus consultation. Rather than functioning simply as a tool for correcting errors, DDL supported learners in developing more natural phraseological choices and in flexibly integrating corpus-informed knowledge into subsequent spoken communication.

6. Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP23K00530 and the 2026 Short-Term Research Project of the College of Economics, Aoyama Gakuin University.

7. Bibliographical References

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

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 23–48). Longman.

Crosthwaite, P. & Steeples, B. (2024). Data-driven learning with younger learners: Exploring corpus-assisted development of the passive voice for science writing with female secondary school students. *Computer Assisted Language Learning, 37*(5–6):1166–1197.

Dhimolea, T. K., Kaplan-Rakowski, R., & Lin, L. (2022). A systematic review of research on high-immersion virtual reality for language learning. *TechTrends, 66*(5):810–824.

Geluso, J., & Yamaguchi, A. (2014). Discovering formulaic language through data-driven learning: Student attitudes and efficacy. *ReCALL, 26*(2):225–242.

Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 3–33). John Benjamins.

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

Kaplan-Rakowski, R. & Gruber, A. (2023). The impact of high-immersion virtual reality on foreign language anxiety when speaking in public. *Smart Learning Environments, 10*:46.

Lusta, A., Demirel, O., & Mohammadzadeh, B. (2023). Language corpus and data driven learning (DDL) in language classrooms: A systematic review. *Heliyon, 9*(12):e22731.

O'Keeffe, A. (2021). Data-driven learning: A call for a broader research gaze. *Language Teaching, 54*(2):259–272.

Satake, Y. (2024). Enhancing English language skills through motivation: A case study using the Wander app for virtual direction-giving tasks. In *INTED2024 Proceedings* (pp. 1493–1502). IATED.

Satake, Y. (2025). The effects of immersive VR tour guiding on English learning. In B. Lacy, M. Swanson, & P. Lege (Eds.), *Moving JALT Into the Future: Opportunity, Diversity, and Excellence: JALT2024 Conference Proceedings* (pp. 328–337). Japan Association for Language Teaching.

Satake, Y., Yamamoto, S., & Obari, H. (2024). Effects of English-speaking lessons in virtual reality on EFL learners' confidence and anxiety. In *Frontiers in technology-mediated language learning* (pp. 26–40). Routledge.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*(4):510–532.

Ueno, S. & Takeuchi, O. (2023). Effective corpus use in second language learning: A meta-analytic approach. *Applied Corpus Linguistics, 3*(3):100076.

Yudintseva, A. (2023). Virtual reality affordances for oral communication in English as a second language classroom: A literature review. *Computers & Education: X Reality, 2*:100018.

8. Language Resource References

Davies, M. (2008). *The Corpus of Contemporary American English (COCA)*. English-Corpora.org. https://www.english-corpora.org/coca/