EP-JOURNALS

GLOBAL JOURNAL OF EDUCATION, FINANCE & MANAGEMENT

Available on: https://gjefm.ep-journals.org/ E-ISSN: 3117-6771, P-ISSN: 3117-6763

Global Journal of Education, Finance and Management

Volume: 02 Issue: 10 October 2026

A Proposed Framework for Learning Spontaneous Arabic Speaking Skill from Vocabulary Recognition

Zainur Rijal Abdul Razak

Fakulti Pengajian Bahasa Utama, Universiti Sains Islam Malaysia, 71800 Nilai, Negeri Sembilan, Malaysia

Published Date: October 28, 2026 **DOI: 10.65150/EP-gjefm/V2E10/2026-03**

ABSTRACT: Students may recognise Arabic vocabulary by reading passages and vocabulary tests while remaining unable to retrieve and use the same words during spontaneous speech. This discrepancy raises a pedagogical question concerning the organisation and accessibility of lexical knowledge rather than vocabulary size alone. This article develops a research problem around the proposition that teaching familiar words through frequent, communicatively useful phrases may improve their productive accessibility more effectively than teaching them as isolated lexical items. It proposes a framework connecting corpus-informed phrase selection, spaced retrieval, and varied speaking tasks. The framework distinguishes recognition, controlled productive recall, context-sensitive phrase use, and transfer to unrehearsed communication. Existing research has investigated formulaic language instruction and oral fluency, while studies of collocation learning have examined distributed practice and retrieval schedules. Consequently, the proposed contribution concerns the conditions under which recognised Arabic vocabulary becomes available for spontaneous production, rather than a general claim that phrase instruction is beneficial. A factorial intervention is proposed to distinguish the effects of instructional unit and retrieval schedule while maintaining comparable speaking opportunities. This study argues that the strongest evidence would involve improved access to previously recognised vocabulary in new communicative situations. It presents testable propositions and methodological safeguards for investigating this transition without equating corpus frequency, memorised reproduction, or faster speech with communicative competence.

KEYWORDS: Arabic language education, productive vocabulary, formulaic language, corpus-informed instruction, spontaneous speech

INTRODUCTION

A learner can know what an Arabic word means without being able to use it when speaking. In a reading activity, the written form supplies a cue, the surrounding sentence narrows interpretation, and the learner may have time to compare possible meanings. During conversation, these supports are reduced. The learner must begin with an intended message, select suitable vocabulary, construct an utterance, and articulate it while responding to another speaker. A word that appears familiar on paper may therefore remain unavailable at the moment of communication (Nation, 2013; Webb, 2008). This problem deserves investigation because vocabulary achievement can be assessed in ways that conceal difficulties in productive access. Matching words with definitions, selecting translations, and identifying correct meanings provide evidence of particular aspects of lexical knowledge. They do not directly establish that students can retrieve those words without seeing them (Webb, 2008). Nation's (2013) account of vocabulary knowledge provides a foundation for distinguishing receptive and productive demands rather than treating knowledge of a word as a single achievement.

The educational concern is therefore more precise than saying that students have weak vocabulary. Some students may lack the necessary lexical knowledge while others may possess partial knowledge that is difficult to access under time pressure. Still others may retrieve individual words but struggle to combine them appropriately (Conklin & Schmitt, 2012; Nation, 2013) because these situations require different instructional responses. Additional wordlists might address unfamiliar vocabulary while doing little to resolve difficulty assembling familiar words into usable speech. This present article examines one possible response that is teaching vocabulary through frequent phrases selected from relevant Arabic corpus evidence, revisiting these phrases through spaced retrieval, and requiring their use in varied speaking tasks (Karpicke & Roediger, 2008; McGuire & Larson-Hall, 2017). The central proposition is that productive accessibility depends partly on how lexical knowledge is organised and practised. A familiar word encountered repeatedly within useful combinations may become easier to retrieve in connection with a communicative purpose (Conklin & Schmitt, 2012; Ellis, 2012).

However, the general relationship between phrase instruction and speaking performance is already a research topic. Nergis (2021), for example, compared instruction involving corpus-derived formulaic sequences with single academic vocabulary in an English-language programme. The phrase group achieved an advantage on a global speech-rate measure, with effects maintained at delayed assessment. A new study should therefore move beyond simply asking whether teaching phrases can improve fluency. A stronger problem concerns whether phrase-based instruction improves the spontaneous accessibility of Arabic words that learners already recognise, and whether this advantage depends on distributed retrieval practice. This formulation connects an identifiable learning difficulty with an instructional mechanism and a measurable

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

outcome. It also creates room for an informative negative result where phrase learning might improve controlled recall without improving unrehearsed speaking (Thomson, 2017). This article therefore was carried out based on this problem to identify provisional research gap and propose a design capable of testing competing explanations.

RESEARCH PROBLEM

Vocabulary recognition occurs when learners identify a lexical form or interpret its meaning when the form is supplied. Productive accessibility concerns their ability to retrieve a suitable form when beginning with a meaning, intention, or communicative situation. The distinction is especially important when the research question concerns spontaneous speaking (Nation, 2013; Webb, 2008). The strength of recognition varies, as learners may select a correct translation through elimination, recognize a word only in a familiar passage, or understand it reliably across several contexts. These performances should not be treated as equivalent (Nation, 2013; Webb, 2008). Before investigating a recognition–production discrepancy, researchers must establish that recognition is sufficiently robust to make the discrepancy meaningful.

Productive accessibility encompasses distinct achievements, including retrieving a word accurately after a long delay, producing it quickly in a translation exercise, and using it appropriately while negotiating a classroom task (Fang et al., 2024; Nation, 2013). The proposed study should therefore examine a sequence of increasingly demanding performances rather than divide vocabulary knowledge into a simple known–unknown classification. The target problem can be operationalised as a recurring pattern in which students demonstrate reliable recognition of selected vocabulary but show slow, unsuccessful, or inappropriate access to that vocabulary during production. This pattern must be documented through baseline assessment and should not be assumed merely because teachers report that students hesitate when speaking. Distinguishing observable speech behaviour from underlying lexical accessibility also requires care when interpreting fluency measures (Suzuki et al., 2021; Tavakoli & Uchihara, 2020).

Knowing a word does not necessarily provide knowledge of its usual combinations, grammatical environment, or discourse function (Nation, 2013). A learner may retrieve an appropriate lexical item but remain uncertain about the preposition, verb pattern, or surrounding expression needed to complete the message. Phrase instruction could address this difficulty by presenting words together with recurrent linguistic relationships. Instead of repeatedly learning an item as an isolated translation equivalent, students would encounter it within expressions associated with particular meanings and purposes. Such instruction could support both lexical choice and the construction of a larger utterance (Ellis, 2012; McGuire & Larson-Hall, 2017; Nergis, 2021). Nevertheless, not every frequent sequence should be treated as a psychologically stored unit. Corpus recurrence describes patterns in recorded language; it does not establish how an individual learner represents those patterns (Conklin & Schmitt, 2012; Gablasova et al., 2017). The proposed research should therefore use “corpus-informed phrases” as an instructional description and reserve claims about holistic retrieval for evidence capable of supporting them.

REVIEW OF THE LITERATURE

3.1 Formulaic instruction and oral performance

Research on formulaic language provides a reason to investigate phrase instruction, but it also limits claims of novelty. Nergis’s (2021) intervention is particularly relevant because it involved a comparison between multiword sequences and single vocabulary items, with communicative oral practice in both conditions. The proposed Arabic study cannot claim originality simply from combining corpus-selected phrases and speaking activities. Instead, it should investigate a more specific issue, whether instructional organisation changes access to vocabulary already available receptively. This shifts attention from general improvement in speaking to a defined transition in lexical performance. It also makes baseline lexical knowledge central to participant and item selection.

The distinction matters theoretically because gains arising mainly from learning previously unknown vocabulary concern vocabulary acquisition, whereas faster and more appropriate use of previously recognised items provides evidence relevant to productive accessibility. Both outcomes matter, but they answer different questions. The distinction between vocabulary recognition and productive availability provides the foundation for the proposed research. Nation (2013) treats vocabulary knowledge as multidimensional, while Webb (2008) demonstrates that estimates of receptive and productive vocabulary differ according to the knowledge assessed and scoring criteria. Consequently, recognising a word should not be interpreted as evidence that a learner can retrieve it during spontaneous communication. The proposed study should establish this discrepancy directly through separate recognition and production measures.

Formulaic-language research offers a plausible explanation for the potential value of phrase instruction. Ellis (2012) discusses formulaic language in relation to frequency and acquisition, while Conklin and Schmitt (2012) review its processing characteristics. However, processing advantages should not be assumed for every learner or every recurrent expression. Instructional studies provide mixed evidence for example, McGuire and Larson-Hall (2017) reported improvements in spoken fluency, whereas Thomson (2017) found gains in cued recall without a significant advantage in speaking fluency over the comparison condition. These differences strengthen the need to distinguish learning expressions from transferring them into speech.

3.2 Distributed practice and collocation learning

Distributed practice has also been investigated in multiword learning. Yamagata et al. (2023) compared practice schedules for verb–noun collocations and included assessment of unstudied combinations sharing a node word. Their findings make transfer beyond directly practised material an important consideration for a new investigation. Fang et al. (2024) distinguished several outcomes when examining retrieval schedules including explicit knowledge, automatized explicit knowledge, and implicit knowledge. Their results varied across measures and assessment times. This supports the methodological principle that an intervention may affect different aspects of lexical knowledge differently where a single posttest cannot represent the entire learning process. The proposed study would extend this concern to spoken Arabic performance by examining whether differences associated with instructional unit and practice distribution remain evident when learners formulate unfamiliar messages, using controlled recall as one outcome and communicative transfer as a more demanding test. Retrieval and distributed practice provide complementary foundations for the intervention: Karpicke and Roediger (2008) demonstrate the importance of retrieval for retention, while Cepeda et al. (2006)

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

synthesise evidence concerning distributed verbal practice. Although these studies justify investigating retrieval schedules, their implications for spontaneous Arabic speaking require direct assessment within the proposed intervention.

Arabic collocation learning already has a research presence in Malaysia. Asbulah et al. (2019) study of students in public universities examined collocation-learning strategies through questionnaire data. This provides relevant contextual grounding but reported strategy use does not establish the causal effectiveness of a particular instructional treatment. A new study should therefore avoid describing Arabic phrase learning as an untouched field. Its contribution would be methodological and explanatory, directly observing recognised vocabulary, manipulating instruction and retrieval conditions, and assessing subsequent use in unrehearsed speech. Corpus-informed materials likewise require careful interpretation. Boulton and Cobb (2017) examine corpus use in language learning, Gablasova et al. (2017) address the identification and interpretation of collocations, and Zaki (2017) demonstrates applications within Arabic teaching. For speaking assessment, Lambert et al. (2017) make task repetition relevant to treatment control, while Tavakoli and Uchihara (2020) and Suzuki et al. (2021) support distinguishing multiword use, temporal performance, and perceived fluency. Together, these sources justify assessing unfamiliar tasks through several complementary measures.

3.3 Word productivity during speaking

The initial idea is best described as a working theoretical proposition where lexical knowledge becomes more accessible for speaking when learners practise retrieving words within useful combinations and communicative contexts. It becomes a stronger theory only when its concepts, mechanisms, boundary conditions, and disconfirming evidence are specified. The proposed account comprises three potentially complementary mechanisms: phrase instruction may provide information about lexical combinations, retrieval practice may strengthen access from meaning to linguistic form, and varied speaking tasks may connect retrieved material with different communicative demands, although their combined presence does not establish that all three are necessary. The account also recognises limitations arising from phrases that are too difficult, inappropriate for the target register, or irrelevant to the speaking task; repeated retrieval failures without adequate feedback; and practice so narrow that learners reproduce trained expressions without being able to adapt them. These conditions should form part of the explanation.

A word taught in a phrase is connected to neighbouring words and has a larger meaning. These connections could reduce uncertainty during production by providing a partially available linguistic structure. The learner would still need to select the expression and adapt it to the intended message. This proposed advantage does not imply that isolated-word learning has no value. Single-word instructions may efficiently establish meaning, pronunciation, and form. The question is whether it provides sufficient preparation for the particular outcome under investigation and appropriate production under communicative pressure. A useful comparison therefore begins with similar target words and equivalent opportunities for meaningful use. It asks whether deliberately organising initial learning around recurrent phrases creates an additional advantage. This makes the argument a testable instructional comparison rather than a general rejection of word-focused learning.

Reading a phrase while repeating it aloud keeps the linguistic form available, whereas retrieving it from a meaning or situational cue requires the learner to reconstruct or access that form from memory. The framework therefore emphasises retrieval direction, suggesting that instruction aimed at spontaneous production should include practice beginning with communicative intent rather than a visible Arabic answer. Retrieval difficulty must nevertheless remain manageable, as excessively difficult attempts may lead to repeated errors or disengagement. Spacing raises the further question of whether separating retrieval opportunities affects retention and subsequent accessibility, which the proposed study should investigate by varying practice distribution while keeping the number of planned attempts approximately constant to avoid attributing the effects of additional practice to spacing.

An expression becomes educationally useful when a learner can select and adapt it appropriately, but accurate reproduction in a single classroom dialogue provides limited evidence of this ability, as transfer requires some change in the situation, interlocutor, information, or linguistic environment. The framework therefore proposes the following relationship:

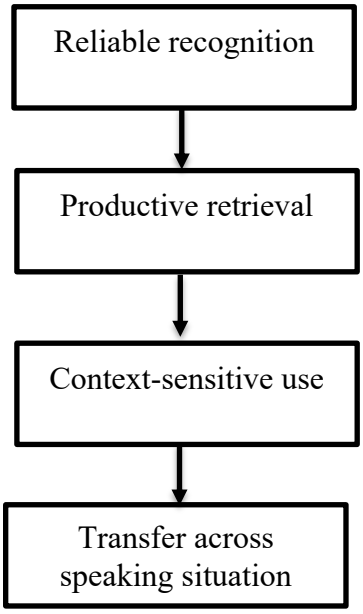


Figure 1: Flow of speaking retrieval

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

These stages are analytically distinct rather than a guaranteed developmental sequence. A learner may progress unevenly, and instruction may improve one stage without changing another. The study becomes theoretically informative by examining where improvement occurs and where it stops.

RESEARCH QUESTIONS AND PROPOSED DESIGN

The principal research question is whether corpus-informed phrase instruction improves the spontaneous use of recognised Arabic vocabulary more than single-word instruction when exposure and speaking opportunities are comparable?

In this regard, three supporting questions clarify the mechanism:

1. Does spaced retrieval produce different outcomes from concentrated retrieval for controlled recall and delayed speaking?
2. Does the effect of phrase instruction depend on retrieval schedule?
3. Do improvements extend to unfamiliar speaking tasks while preserving accuracy and communicative appropriateness?

The first proposition predicts an advantage for phrase instruction in appropriate target-vocabulary use during unfamiliar tasks, expressing a directional hypothesis rather than finding the analysis must confirm. The second predicts that spacing may support retention beyond the immediate posttest, with a stronger delay than immediate effect being compatible with this proposition, although interpretation would depend on assessment reliability and actual practice participation. Meanwhile, the third concerns interaction. Phrase instruction might benefit more from spacing than single-word instruction, but interaction should not be assumed. Independent effects would support a different account: instructional organisation and distribution could make separate contributions. The fourth proposition concerns retrieval as a possible explanatory pathway. Improvements in timed productive access may accompany improvements in spontaneous use. Such an association would be consistent with the framework, but it would not by itself establish causal mediation. Finally, a phrase advantage limited to rehearsed material would qualify the theory. It would suggest successful learning of particular expressions without sufficient evidence of broader communicative transfer.

A two-by-two factorial intervention would provide the clearest test of the proposed mechanisms. The design would cross instructional units with retrieval schedule.

Table 1: **Proposed design for spontaneous speaking skills**

Condition	Instructional unit	Retrieval schedule
A	Corpus-informed phrases	Spaced
B	Corpus-informed phrases	Concentrated
C	Individual target words	Spaced
D	Individual target words	Concentrated

All conditions would include comparable speaking tasks. This is necessary because a group receiving substantially more speaking practice might improve for reasons unrelated to phrase organisation. The study would include baseline assessment, immediate posttesting, and delayed posttesting. A teaching period of several weeks could provide sufficient opportunities to implement the treatments, but its duration should be determined through piloting and practical access rather than presented as theoretically optimal.

A suitable population would be intermediate learners of Arabic as an additional language who can understand basic instructional discourse and participate in short speaking tasks. Malaysian university students would provide a plausible setting, subject to access and documented need. Participant selection should establish both general proficiency and the specific lexical discrepancy. Students need not display identical vocabulary profiles, but the study should record which items each learner recognises and can already produce. Random assignment would strengthen causal interpretation, but where instruction involves intact classes, class-level allocation may be necessary, preferably with several classes per condition to distinguish instructional effects from class characteristics. Sample size should be determined through a power analysis accounting for the planned primary outcome, the smallest educationally meaningful effect, clustering, and expected attrition, as dividing a convenient total into four groups would not ensure adequate power, particularly for detecting interaction.

Baseline assessment should separate recognition from production without allowing one test to supply answers for another. Productive testing should generally precede recognition testing, or equivalent item sets should be counterbalanced. Recognition could be assessed through more than one context. Productive access could be examined through meaning prompts, situational prompts, and short speaking tasks. An item-level profile would identify vocabulary that is recognised but not reliably produced. This approach avoids assuming that every hesitation represents a retrieval failure. It also permits analysis of whether treatment effects differ for initially unknown words, recognised words, and words already available productively. The primary analysis should focus on the category that matches the research problem.

CORPUS-INFORMED MATERIALS AND INSTRUCTION

Corpus selection should align with the communicative domain, so an intervention targeting spoken Modern Standard Arabic for university discussion should use materials that closely represent that context. A smaller, clearly documented collection may provide a stronger basis for phrase selection than a larger resource dominated by unrelated written genres, since size alone does not establish suitability. The study should therefore report corpus composition, provenance, permissions, genre coverage, and relevant processing decisions. As no Arabic phrase should be labelled frequent without analysis, the present article provides neither an invented phrase list nor unsupported frequency figures; actual teaching targets must be selected from verified data and evaluated for pedagogical suitability.

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

Candidate phrases should be examined for recurrence, distribution, meaning, function, and adaptability. Frequency concentrated in a single source may reflect a speaker's habits or a specialised topic rather than broad usefulness. Human review should determine whether candidates are complete enough to teach, appropriate for the register, and manageable at the learners' proficiency level. Reviewers should also identify phrases whose use requires substantial grammatical knowledge beyond the intended treatment. Arabic processing requires explicit decisions about orthographic variation, attached elements, inflection, and the relationship between surface forms and broader phrase patterns. These decisions affect counts and matching. Original examples should remain traceable so that computational grouping can be checked against linguistic context.

Both instructional groups should study the same core vocabulary, with phrase conditions organising teaching around recurrent combinations and their functions and word conditions organising initial teaching around individual meanings and forms. Because words naturally enter combinations during speaking practice, the comparison is best described as phrase-organised versus word-organised vocabulary instruction within a common speaking curriculum. Planned exposure, retrieval attempts, feedback, and instructional time should be documented, with differences reported transparently because the additional words and relationships within phrases may prevent perfect equivalence between conditions. A materials audit would record the target vocabulary, phrase length, supporting words, anticipated difficulty, and exposure schedule, while piloting could identify items that create excessive imbalance or fail to elicit meaningful production.

A phrase-based lesson could begin with a short contextual encounter and an explanation of meaning and communicative function, followed by examination of the target combination and retrieval from a meaning or situational cue. Subsequent activities would require learners to adapt the expression by changing a participant, time reference, stance, or communicative goal where linguistically appropriate, before completing a speaking task in which the expression is useful but not displayed as a required answer. Feedback should distinguish lexical selection, phrase formation, and communicative appropriateness, with delayed feedback during fluency-focused interaction preserving task continuity and immediate correction during controlled practice helping to prevent repeated rehearsal of inaccurate forms.

RETRIEVAL SCHEDULES AND SPEAKING TASKS

The spaced conditions would distribute retrieval attempts across separate sessions, whereas concentrated conditions would place an equivalent planned number within a shorter period, with both receiving initial instruction and corrective feedback. The intervals should be justified by the instructional calendar and pilot evidence, and the time between the final practice attempt and each assessment should be reported without claiming that one schedule is optimal before comparing alternatives. Exposure logs and careful scheduling of common speaking tasks are essential because unplanned encounters with target items during ordinary lessons could effectively introduce distributed practice into concentrated conditions. The central design challenge is therefore to provide equal speaking opportunities while maintaining a meaningful difference in target retrieval distribution; if classroom conditions cannot sustain this distinction, a narrower experiment or revised research question would be preferable to an unsupported claim of treatment control. Training tasks should encourage learners to exchange information, explain choices, or resolve a problem, making the target vocabulary relevant without turning it into a compulsory script. Tasks could vary in topic familiarity, interactional structure, and information distribution, provided these variations remain manageable so that difficulties do not arise primarily from unfamiliar content. Parallel tasks should be piloted to ensure comparable linguistic and cognitive demands, while assessment should differ meaningfully from rehearsed tasks to evaluate transfer beyond practice.

Near transfer could involve a new situation requiring a familiar communicative function, whereas more demanding transfer could introduce an unfamiliar topic or interactional arrangement while retaining reasonable opportunities to use the learned vocabulary. Neither assessment should display target expressions, since providing a phrase bank would shift the task from spontaneous retrieval to supported selection. In this study, "spontaneous" could be operationally defined as unrehearsed performance on an unseen task following a standardised preparation period and without access to target-language prompts, rather than speech produced without any planning or prior learning. The primary outcome should reflect the research problem directly which is appropriate use of target vocabulary during unfamiliar speaking tasks. This could be assessed through opportunity-sensitive coding rather than raw frequency alone. A target word may be absent because the learner cannot retrieve it, because the task does not require it, or because an appropriate alternative is available. Researchers should therefore interpret non-use cautiously and design tasks that create plausible opportunities without forcing specific wording. Secondary outcomes would include recognition accuracy, controlled productive recall, retrieval latency, phrase appropriateness, and temporal fluency. The measurement plan should distinguish use of exact trained phrases from flexible use of their constituent vocabulary.

Temporal measures might include speech rate, articulation rate, silent-pause duration, and repair frequency, with each defined in advance alongside procedures for handling fillers, repetitions, and incomplete utterances. Speed should be assessed together with accuracy and task achievement, since faster speech containing unsuitable combinations does not necessarily indicate more effective productive access. Arabic assessment also requires register decisions aligned with the study's educational aims, including explicit coding procedures for regional forms, code-switching, and mixed-register responses when tasks target Modern Standard Arabic. Raters should be trained and, where feasible, unaware of treatment condition and assessment time, with a subset of recordings independently coded to establish agreement and documented procedures used to resolve disagreements.

The primary analysis should test instructional unit, retrieval schedule, assessment time, and relevant interactions using models that account for repeated observations and the structure of learners, items, tasks, and classes where applicable. Superiority between conditions should be evaluated through estimated differences, their uncertainty, and their educational importance, since significant improvement within one group alone does not establish an advantage over another. Baseline production and recognition should be incorporated appropriately, alongside examination of ceiling effects, missing data, and unequal participation, particularly attrition at delayed testing, when returning students may differ from those who do not return. The study should preregister its primary outcome and planned comparisons, while clearly identifying any additional analyses of proficiency differences or item characteristics as exploratory.

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

Corresponding Author: Zainur Rijal Abdul Razak

PEDAGOGICAL SIGNIFICANCE

The proposed framework allows several informative outcomes: a phrase advantage in immediate recall but not delayed speaking would suggest learning of expressions without demonstrating durable communicative accessibility, whereas an advantage on delayed unfamiliar tasks would provide stronger evidence of transfer. If the latter advantage were greater under spaced retrieval, it could suggest that instructional organisation and practice distribution work together, although this interaction would require direct statistical support. Equal spacing benefits across phrase and word conditions would instead suggest a general contribution of distributed retrieval, while still offering practical guidance for vocabulary teaching.

Similar improvement across all groups could reflect the contribution of common speaking practice, a weak instructional contrast, or insufficient statistical precision, with treatment records and confidence intervals helping to distinguish these possibilities. Increased use of exact phrases without broader lexical flexibility would identify a boundary between memorised reproduction and adaptable language use, warranting a qualified interpretation rather than a judgement of complete success or failure. Interpretation must therefore remain aligned with the measures: corpus recurrence supports material selection, improved recall indicates learning of tested items, and unfamiliar-task performance supports transfer within those task conditions, but none alone establishes a universal teaching principle.

The proposed study could help teachers diagnose why familiar vocabulary remains unavailable in speech. If recognition is secure but productive retrieval is weak, instruction may need more opportunities to begin with meaning and formulate language without a visible answer. If phrase organisation proves useful, teachers could supplement individual vocabulary study with recurrent combinations linked to communicative functions. The implication would be integration rather than replacement: which shows that students still need knowledge of individual words, pronunciation, grammar, and meaning. A further implication concerns assessment where courses that value speaking should include tasks requiring retrieval and adaptation, because recognition tests cannot directly demonstrate those outcomes. Short, repeated speaking assessments may reveal difficulties that written vocabulary scores overlook.

Several limitations would remain, as effects may depend on proficiency, task type, language background, phrase difficulty, and the Arabic variety taught, limiting the generalisability of a successful university intervention to young learners or advanced professional communication. Classroom contamination through shared materials or external learning resources also requires consideration, since participation logs can document some additional practice but are unlikely to ensure complete control. The framework also has theoretical limits: faster retrieval and improved speech would be consistent with stronger lexical accessibility but would not prove that phrases are stored and retrieved as indivisible units, a claim requiring more direct psycholinguistic evidence. Finally, publication potential depends on execution, as a carefully framed question cannot compensate for poorly documented materials, unreliable coding, insufficient sample size, or unsupported novelty claims; the strongest contribution would therefore be a transparent test of a specific learning mechanism.

CONCLUSION

The uncertain transition from recognising Arabic vocabulary to using it appropriately during spontaneous speech requires direct investigation because vocabulary familiarity, controlled recall, and communicative availability represent distinct educational outcomes. Corpus-informed phrase instruction offers a plausible response by connecting familiar words with recurrent combinations and communicative functions, while spaced retrieval may support continued access and varied speaking tasks may encourage adaptation beyond the original learning context; however, these relationships require empirical evidence. The central research question is therefore whether phrase-organised instruction improves the use of recognised vocabulary in unfamiliar speaking tasks under comparable practice conditions. A factorial design would help distinguish the effects of instructional organisation and retrieval distribution, while baseline, immediate, and delayed measures would capture performance at different points in learning. The proposed contribution is an account of when and how recognised vocabulary becomes productively accessible, supported by appropriate, retained use in unfamiliar situations and qualified by outcomes limited to recognition or rehearsed reproduction. Either pattern would advance understanding if the study were designed and reported transparently.

REFERENCES

1) Asbulah, L. H., Aladdin, A., Lubis, M. A., & Sahrim, M. (2019). Strategi pembelajaran kolokasi bahasa Arab dalam kalangan pelajar universiti awam Malaysia. *GEMA Online Journal of Language Studies*, 19(2), 139–159. https://doi.org/10.17576/gema-2019-1902-09
2) 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
3) Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354
4) Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, 32, 45–61. https://doi.org/10.1017/S0267190512000074
5) Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, 32, 17–44. https://doi.org/10.1017/S0267190512000025
6) Fang, N., Elgort, I., & Chen, Z. (2024). Effects of retrieval schedules on the acquisition of explicit, automatized-explicit, and implicit knowledge of L2 collocations. *Studies in Second Language Acquisition*, 46(3), 663–685. https://doi.org/10.1017/S0272263124000184
7) Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, 67(S1), 155–179. https://doi.org/10.1111/lang.12225
8) Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968. https://doi.org/10.1126/science.1152408
9) Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition*, 39(1), 167–196. https://doi.org/10.1017/S0272263116000085

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

Corresponding Author: Zainur Rijal Abdul Razak

10) McGuire, M., & Larson-Hall, J. (2017). Teaching formulaic sequences in the classroom: Effects on spoken fluency. *TESL Canada Journal*, 34(3), 1–25. https://doi.org/10.18806/tesl.v34i3.1271
11) Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.
12) Nergis, A. (2021). Can explicit instruction of formulaic sequences enhance L2 oral fluency? *Lingua*, 255, Article 103072. https://doi.org/10.1016/j.lingua.2021.103072
13) Suzuki, S., Kormos, J., & Uchihara, T. (2021). The relationship between utterance and perceived fluency: A meta-analysis of correlational studies. *The Modern Language Journal*, 105(2), 435–463. https://doi.org/10.1111/modl.12706
14) Tavakoli, P., & Uchihara, T. (2020). To what extent are multiword sequences associated with oral fluency? *Language Learning*, 70(2), 506–547. https://doi.org/10.1111/lang.12384
15) Thomson, H. (2017). Building speaking fluency with multiword expressions. *TESL Canada Journal*, 34(3), 26–53. https://doi.org/10.18806/tesl.v34i3.1272
16) Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, 30(1), 79–95. https://doi.org/10.1017/S0272263108080042
17) Yamagata, S., Nakata, T., & Rogers, J. (2023). Effects of distributed practice on the acquisition of verb–noun collocations. *Studies in Second Language Acquisition*, 45(2), 291–317. https://doi.org/10.1017/S0272263122000225
18) Zaki, M. (2017). Corpus-based teaching in the Arabic classroom: Theoretical and practical perspectives. *International Journal of Applied Linguistics*, 27(2), 514–541. https://doi.org/10.1111/ijal.12159

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

Corresponding Author: Zainur Rijal Abdul Razak