

Corpus-Informed Development of Pedagogical Discourse Competence in Foreign Language Teacher Education

Nilufar Ulmasovna Ochilova

Doctoral Researcher, Bukhara State Pedagogical Institute, PhD in Philological Sciences, Uzbekistan

Received: 31 December 2025 **Accepted:** 23 January 2026 **Published:** 28 February 2026

ABSTRACT

This design study proposes a corpus-informed cycle for developing pedagogical discourse competence among pre-service foreign language teachers. The approach uses a small, ethically curated classroom corpus not as a source of model sentences but as evidence for noticing relationships among pedagogical function, linguistic realization, sequential position, and learner uptake. Corpus-pedagogic synthesis produced five stages: formulating a functional inquiry, retrieving contrasting episodes, interpreting concordance in context, designing a discourse alternative, and testing it through microteaching. An evidence card safeguards against frequency-based prescriptions. The result is a professional inquiry method that connects authentic classroom data with situated rehearsal while preserving contextual and ethical judgment.

Keywords: Classroom corpus, data-driven learning, teacher education, pedagogical discourse, concordance, learner uptake.

INTRODUCTION

Pre-service teachers are often given lists of useful classroom phrases, yet such lists detach language from timing, learner response, and pedagogical purpose. A phrase such as ‘Can you say more?’ may open reasoning, delay necessary clarification, or publicly expose an unprepared learner depending on its position and delivery. Professional learning therefore requires access not only to forms but to episodes in which forms do interactional work.

Classroom corpora can make recurrent teacher and learner discourse searchable. However, frequency alone is not a pedagogical norm. The most common expression may reflect a narrow instructional routine, and a low-frequency move may be particularly valuable in a specific contingency. This study asks how corpus work can support contextualized noticing and rehearsal rather than the imitation of frequent teacher talk.

The dissertation problem isolated in this article is the use of decontextualized classroom phrase lists that hide timing,

sequence, learner response, and pedagogical function. The scope is therefore deliberately restricted: it is not a general orchestration framework or a frequency-based prescription, but a disciplined way to learn from a small classroom corpus. This boundary prevents the recurring term pedagogical discourse competence from turning separate publications into paraphrases of the same claim. It also makes the article accountable for one identifiable conceptual or design outcome.

The analytical unit was defined as function, linguistic realization, sequential position, local context, and learner uptake. Sources were not grouped merely because they used similar terminology. Each proposition was coded for the professional decision it explains, the opportunity it creates or removes for a learner, the evidence by which it can be examined, and the contextual limitation that qualifies the claim. This function-first procedure reduced category inflation and preserved the sequential character of classroom discourse.

The design outcome is a corpus inquiry card and

contrastive concordance set. Its content validity was examined against three requirements: direct correspondence with the focal problem, visibility in a feasible pre-service teacher task, and capacity to support a subsequent instructional decision. The framework is therefore presented as an analytically reasoned prototype. It does not convert conceptual synthesis into an unsupported claim of intervention effectiveness.

METHODS

A design-oriented literature synthesis brought together classroom corpus research, data-driven learning, language teacher education, and interaction analysis. The proposed corpus consists of consented, de-identified transcripts of short classroom episodes tagged by pedagogical function, participation format, learner level, sequential position, and observable learner response. Sensitive personal data and evaluative labels are excluded.

The design was tested conceptually against four criteria: functional searchability, preservation of sequential context, explicit uncertainty, and transfer to performance. A unit was retained only if the query could return contrasting realizations of the same pedagogical purpose and if users could inspect the turns before and after the focal item.

The synthesis differentiated the focal problem into formulating a functional inquiry, retrieving contrasting episodes, restoring interactional context, and designing and testing an alternative. These dimensions should not be read as interchangeable competences. Each captures a different decision point and produces a different kind of evidence. Their connection lies in the movement from interpreting a situation to creating an opportunity, observing its consequence, and deciding what must be retained or redesigned.

The first dimension, formulating a functional inquiry, establishes the professional purpose of the episode. A novice must distinguish the curricular intention from the discourse action that is supposed to enact it. Without this distinction, an apparently communicative activity may still distribute all interpretive authority to the teacher, while a concise intervention may create substantial room for learner reasoning.

The second dimension, retrieving contrasting episodes, concerns the alternatives available at a consequential

moment. The relevant question is not whether a preferred form appears, but why that option fits the learners, timing, relationship, and task. Requiring at least two plausible alternatives makes tacit assumptions visible and enables the student teacher to anticipate both an intended opportunity and a foreseeable cost.

The third dimension, restoring interactional context, turns the planned decision into inspectable performance. Consider comparing several occurrences of 'Can you say more?' to determine when it opens reasoning, delays repair, or places an unprepared learner under pressure. The move cannot be evaluated from wording alone. Its adequacy depends on sequential placement, delivery, the learner's available response route, and what actually happens in the following turn or action.

The final dimension, designing and testing an alternative, makes redesign part of competence rather than a remedial activity after failure. A student teacher may retain the original move, modify its timing, combine it with another mode, or reject it after examining evidence. In each case, the decision must be justified by the learning purpose and by an observed learner consequence rather than by personal preference.

The primary evidence is a traceable chain from corpus episode to interpretation, design choice, microteaching trial, and revised claim. This formulation guards against counting visible techniques as proof of competence. A question, recast, pause, visual support, or feedback statement becomes professionally meaningful only when its function and the learner opportunity that follows can be reconstructed from the episode.

RESULTS

Stage one is a functional inquiry. Instead of searching for a word, the student frames a professional question such as how teachers invite elaboration after a partial answer or repair an instruction misunderstood by several learners. This prevents technology from determining the pedagogical problem.

Stage two retrieves contrasting episodes. Concordance lines remain linked to expanded transcripts, task purpose, participation format, and learner level. Students compare a move that appears to open learner meaning with one that closes or redirects it. Stage three interprets patterns cautiously: lexical bundles, modality, pronouns, question

design, wait time notation, and uptake are examined together.

Stage four requires the student to construct two alternatives for a new scenario and justify them with corpus evidence and pedagogical reasoning. Stage five tests one alternative in microteaching, where a peer introduces an unanticipated learner response. The student then compares the predicted effect with the observed next turn and revises the evidence

card.

The evidence card records query, selected episodes, contextual similarities and differences, tentative pattern, designed alternative, and performance outcome. It explicitly prohibits claims such as ‘teachers should always use X’ unless supported by function, context, and learner evidence.

Table 1. Corpus-informed inquiry cycle

Stage	Student action	Required evidence	Safeguard
Functional inquiry	Frame a pedagogical discourse problem	Purpose and learner need	No word-first searching
Contrast	Retrieve different episodes	Expanded transcript and metadata	No isolated concordance
Interpret	Relate form, sequence, and uptake	Before/after turns	Tentative claims
Design	Create and justify alternatives	Comparable corpus cases	More than one option
Test	Use an option in microteaching	Observed next turn	Revise after mismatch

The specific contribution is turning corpus evidence into contextual professional noticing rather than imitation of frequent forms. This distinguishes the article from the other studies in the publication set. It neither reclaims their models nor repeats their assessment constructs; it develops the single dissertation problem assigned to this paper and states what falls outside its explanatory reach.

In teacher education, a corpus inquiry card and contrastive concordance set can organize a compact cycle of transcript comparison, individual design, peer challenge, microteaching, and evidence-based revision. The tutor supplies a consequential situation rather than a model answer. Students first articulate the purpose, then compare alternatives and predict learner opportunities before performance. Post-performance discussion begins with a selected piece of evidence, which limits impressionistic praise or criticism.

Context sensitivity remains essential. The same linguistic form can perform different work with another age group, proficiency level, institutional norm, or interactional history. Consequently, the framework is not a script or a frequency checklist. Its educational value lies in making the grounds and consequences of a choice discussable

while preserving the teacher's responsibility for situated judgment.

Future empirical work should collect a traceable chain from corpus episode to interpretation, design choice, microteaching trial, and revised claim across repeated episodes. Expert review can establish relevance and clarity, but developmental sensitivity requires longitudinal comparison of student decisions. Learner response, student-teacher explanation, and mentor observation should be triangulated so that neither external scoring nor self-report becomes the sole warrant for a claim of growth.

The local theoretical basis also clarifies why corpus frequency cannot be treated as a pedagogical norm. Safarov’s pragmalinguistic account places meaning in the relationship among intention, addressee, and context; therefore, a recurrent teacher phrase must be restored to its interactional episode before interpretation. Jalolov, Makhkamova, and Ashurov view English language teaching methodology through purposeful classroom activity, while Jalolov and colleagues link instructional technology to the learner action elicited by a task. These Uzbek perspectives justify the article’s shift from collecting frequent expressions to comparing how discourse forms function in situated classroom sequences.

DISCUSSION

The cycle builds on corpus-based teacher development while shifting the learner from consumer of frequency lists to investigator of pedagogical action. Farr's work demonstrates the value and challenges of corpus-informed instruction in teacher education; the present design gives the inquiry a functional and sequential focus. It also aligns with evidence-led reflective practice by requiring claims to point to transcript data.

A small local corpus can increase relevance to the institutional context, but it may reproduce limited routines or normalize ineffective practices. Contrasting episodes, external benchmark samples, and an explicit critical question about power and participation are therefore required. The method also depends on robust consent, de-identification, access control, and deletion schedules.

This study does not report learning gains. The designed cycle should be evaluated through changes in the diversity and contextual fit of students' discourse alternatives, the quality of evidence use, and transfer from corpus analysis to unprompted teaching.

CONCLUSION

A five-stage corpus-informed cycle was developed to connect authentic classroom discourse with pre-service teacher rehearsal. Functional queries, contextualized concordance, contrasting episodes, designed alternatives, and microteaching prevent corpus frequency from becoming prescription. The approach treats corpus evidence as a prompt for professional judgment, not as a substitute for it.

REFERENCES

1. Farr F. Evaluating the use of corpus-based instruction in a language teacher education context // *Language Awareness*. 2008. Vol. 17(1). P. 25-43. DOI: 10.2167/la414.0.
2. Farr F. *The discourse of teaching practice feedback*. London: Routledge, 2011.
3. O'Keeffe A., McCarthy M., Carter R. *From Corpus to Classroom*. Cambridge: Cambridge University Press, 2007.
4. Boulton A. Data-driven learning: taking the computer out of the equation // *Language Learning*. 2010. Vol. 60(3). P. 534-572.
5. Walsh S. *Exploring Classroom Discourse*. London: Routledge, 2011.
6. Walsh S., Mann S. Doing reflective practice: a data-led way forward // *ELT Journal*. 2015. Vol. 69(4). P. 351-362.
7. Seedhouse P. *The Interactional Architecture of the Language Classroom*. Malden, MA: Blackwell, 2004.
8. Council of Europe. *CEFR Companion Volume*. Strasbourg, 2020.
9. Cirocki A., Hallet W. *Teacher Education and Professional Competences*. Cambridge University Press, 2024. DOI: 10.1017/9781009350235.003.
10. Farr F., Farrell A., Riordan E. Corpora for language teacher education: reflections and future directions // *TESOL Quarterly*. 2024. DOI: 10.1002/tesq.3281.
11. Safarov Sh. *Pragmalingvistika*. - Toshkent: O'zbekiston Milliy ensiklopediyasi, 2008.
12. Jalolov J.J., Makhkamova G.T., Ashurov Sh.S. *English Language Teaching Methodology*. - Tashkent: Fan va texnologiya, 2015.
13. Jalolov J., Mahkamova G., Mamadayupova V., Ismatova L. *Ingliz tili o'rgatish texnologiyasi*. - Toshkent: TDPU, 2010. - 96 b.