Materiallar to'plami · Conference Proceedings
"ZAMONAVIY TADQIQOTLAR VA BARQAROR RIVOJLANISH" respublika ilmiy-amaliy konferensiyasi
04 August 2026
UDK 37
PEDAGOGIKA

INTEGRATING CORPUS-BASED TECHNOLOGIES INTO ESP INSTRUCTION FOR THE IMPROVEMENT OF PRAGMATIC COMPETENCE

Manzura Ne'matjon qizi Mamajonova*
PhD student, Uzbekistan State World Languages University
✉ manzurawebster94@gmail.com

ABSTRACT

English for Specific Purposes (ESP) instruction has traditionally prioritized technical vocabulary and grammatical accuracy, often leaving pragmatic competence - the ability to use language appropriately in professional and academic contexts underdeveloped. This article examines how corpus technologies, including concordancers, specialized corpora, and data-driven learning (DDL), can be integrated into ESP instruction to foster learners' pragmatic awareness of speech acts, hedging, politeness, and discourse markers. Drawing on recent research, the article outlines a cyclical pedagogical framework moving learners from corpus exploration to contextualized production, presents a comparative table of corpus-based tools, and discusses pedagogical implications, challenges, and future research directions. The analysis suggests that, despite certain technical and representativeness limitations, corpus-based approaches offer ESP learners authentic, discipline-specific input that supports the noticing and internalization of pragmatic norms.

Keywords

corpus linguistics, pragmatic competence, English for Specific Purposes (ESP), data-driven learning, concordancing, discourse markers, corpus-based instruction

DOI: 10.5281/zenodo.21743064

INTRODUCTION

English for Specific Purposes (ESP) instruction targets the current and future academic or occupational needs of learners, focusing on the language, genres, and skills required to meet those needs [9]. In practice, however, ESP courses often concentrate on specialized vocabulary and text structure, while the pragmatic dimension of communication - how language is used appropriately given a particular audience, purpose, and context receives only marginal attention.

Corpus linguistics offers a partial solution: by analyzing authentic, discipline-specific texts, learners can observe how professionals actually hedge, request, apologize, or signal stance, rather than relying on idealized textbook dialogues. This article examines how corpus technologies - concordancing software, specialized ESP corpora, and data-driven learning (DDL) -can be integrated into ESP instruction to support pragmatic competence. It reviews recent research, proposes a cyclical classroom framework, compares several corpus tools, and discusses pedagogical implications and limitations.

Pragmatic competence is generally understood as comprising two dimensions: pragmalinguistic knowledge -the relationship between linguistic forms and their communicative functions -and sociopragmatic knowledge - understanding of the contextual and cultural variables that determine which forms are appropriate [1]. Both are difficult to teach because the mapping between form, function, and context shifts according to register, culture, and the relationship between interlocutors.

In ESP settings this difficulty is compounded by discourse-community specificity: a hedge appropriate in a medical case report may be too tentative in an engineering proposal, and a casual business-email request

may be too direct in a formal client letter. Learners therefore need exposure not only to general pragmatic norms but to those of their target discipline.

Corpus linguistics - the study of language through large electronic collections of naturally occurring texts - provides a methodological basis for identifying such patterns. Data-driven learning (DDL) invites learners to act as researchers: they search corpora for a given form, observe concordance lines, and inductively formulate hypotheses about usage. Reviews of DDL research report that the approach helps learners move beyond rule-based instruction toward discovery-based learning, though its benefits depend heavily on task design and learner support [6; 5].

While most DDL research has focused on vocabulary and grammar, a smaller body of work applies corpus methods to pragmatics specifically. Bardovi-Harlig, Mossman and Su demonstrated that corpus-based instruction on conventional expressions and pragmatic routines improved learners' recognition and production of these forms more than instruction without corpus exposure [2], indicating that corpora can reveal pragmatic routines and politeness strategies otherwise difficult to notice in textbook input.

Corpus-based approaches and tools for ESP pragmatics.

Several strands of practice support pragmatic development in ESP classrooms. The first is data-driven learning, in which learners query a corpus for a target expression - hedging verbs such as "suggest" in academic writing, or mitigating phrases such as "I was wondering if" in business requests - and examine concordance lines to infer function. Guided, scaffolded tasks are considerably more effective than unstructured browsing, especially for learners new to corpus tools [7].

The second strand involves specialized ESP corpora compiled from authentic professional or academic genres, such as business emails, medical case notes, or conference presentations. Freely available software that batch-converts documents into searchable text has made such compilation increasingly feasible, allowing both teachers and learners to become researchers of their own target discourse community [10].

The third strand draws on interactive pragmatics platforms that share the same goal of exposing learners to contextualized input. A recent systematic review of technology-integrated L2 pragmatics research identified self-access platforms such as Words at Work, organized around realistic workplace scenarios, and Dancing with Words, used to teach apology routines; both supported explicit, contextualized learning through interactive tasks [8]. The same review noted that corpus-based instruction, alongside computer-mediated communication and simulated dialogue systems, was among the effective technological applications, though still comparatively underused [8].

Table 1 compares several widely available corpus tools and platforms by type and pragmatic focus within ESP instruction.

Table 1. Corpus-based tools and their applications in ESP pragmatics instruction

Tool / Resource	Type	Pragmatic Focus	Example Application in ESP
AntConc	Free concordancer	Frequency, collocation, key-word-in-context	Locating hedging devices (e.g., "might", "perhaps") in business correspondence
Sketch Engine	Corpus query system	Word sketches, collocation networks	Analyzing politeness markers in legal and administrative correspondence
COCA (Corpus of Contemporary American English)	General reference corpus	Register comparison	Comparing formal and informal request strategies across professional registers
BAWE (British Academic Written English)	Specialized academic corpus	Academic hedging and stance markers	Teaching cautious claims and stance-taking in EAP research writing
Words at Work	Interactive pragmatics platform	Workplace speech acts	Simulating professional requests, complaints, and clarifications
Dancing with Words	Self-access pragmatics website	Apology strategies	Practicing L2 apology routines in service-encounter scenarios

General reference corpora such as COCA are useful for comparing register-specific variation, while specialized corpora such as BAWE let learners focus narrowly on academic genres. Purpose-built pragmatics platforms complement these tools by embedding input within realistic scenarios and providing explicit feedback, which raw concordance data alone cannot offer.

A cyclical framework for corpus-based pragmatic instruction.

Building on the approaches above, this article proposes a five-stage cyclical framework for integrating corpus technologies into ESP pragmatics instruction, illustrated in Figure 1. The cycle begins with compiling a small, discipline-specific corpus from authentic texts relevant to learners' field. Learners then use concordancing software to search for target pragmatic expressions, generating concordance lines for noticing recurring patterns the hedges, boosters, or politeness markers accompanying a speech act. This is followed by guided analysis, in which learners hypothesize about the contextual factors conditioning form choice, typically in pairs with teacher scaffolding. The cycle concludes with contextualized production role-plays, simulated correspondence, or writing with feedback feeding back into corpus refinement for subsequent cycles.

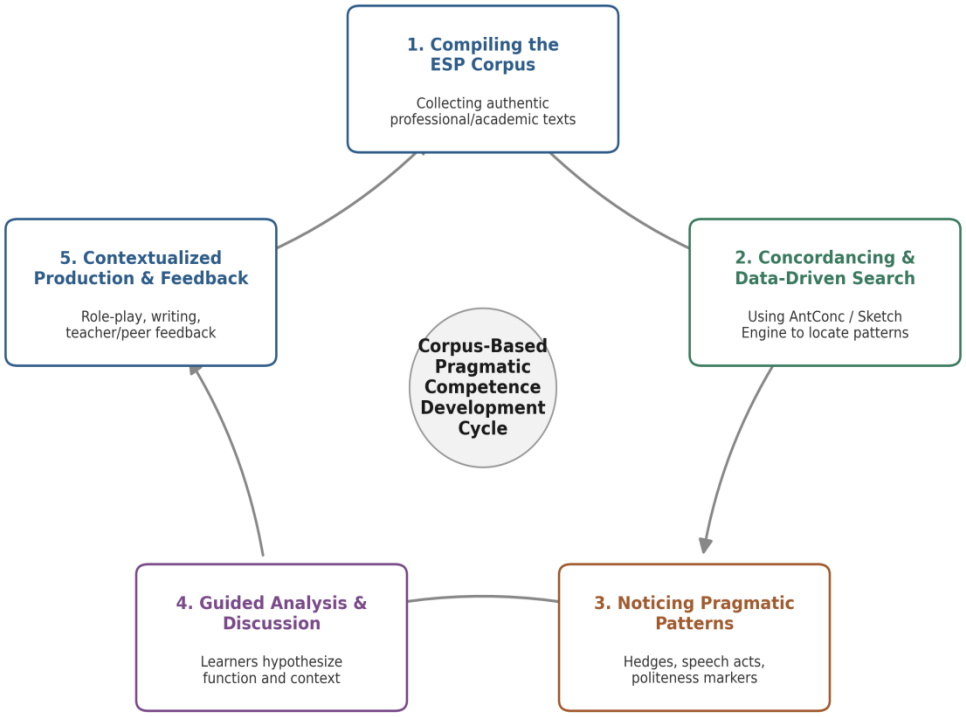


Figure 1. A cyclical framework for corpus-based development of ESP learners' pragmatic competence

This structure reflects two principles from the DDL literature: learners benefit most from corpus work embedded in a clear pedagogical sequence rather than free exploration [7], and pragmatic competence develops most reliably when noticing is followed by guided, meaningful production rather than passive recognition [3; 4].

The framework has several practical implications. Teachers do not need commercial or highly technical corpora to begin: a modest corpus of twenty to thirty authentic texts from learners' target genre internship reports, sample correspondence, published case studies already yields useful concordance data with free tools such as AntConc. Worksheets should target specific pragmatic categories rather than open-ended search comparing concordance lines for direct versus mitigated requests, or sorting hedged versus unhedged claims in abstracts making the contrast explicit rather than left to unguided inference.

Teacher preparation matters considerably: teachers need sustained opportunities to develop corpus literacy before they can design effective tasks [7], so institutions should budget time for training alongside acquiring software and materials. Finally, noticing activities are most effective when paired with production tasks such as role-play or simulated correspondence, since pragmatic competence is ultimately demonstrated in use, not recognition alone.

Corpus-based pragmatics instruction faces several limitations. Learners without prior corpus training often struggle with the interpretive demands of reading concordance lines a difficulty independent of general proficiency, related

instead to unfamiliarity with corpus output. Careful scaffolding and gradual withdrawal of teacher support have been recommended to manage this [5].

A second limitation concerns representativeness: whether a corpus adequately represents the target domain, and how design decisions affect the patterns that emerge, remain persistent concerns for small, teacher-compiled ESP corpora [10]. Third, pragmatics-specific corpora remain scarce; most large corpora were designed for lexico-grammatical rather than pragmatic analysis, so speech acts are rarely annotated and must be located manually.

Finally, broader reviews call for more longitudinal, socially situated research and closer study of how learner differences interact with technological affordances an agenda that applies directly to corpus-based ESP pragmatics and remains largely unaddressed [8].

CONCLUSION

Corpus technologies give ESP instructors a practical means of exposing learners to authentic, discipline-specific pragmatic patterns that conventional materials rarely convey. Combining concordancing tools, specialized corpora, and interactive pragmatics platforms within a cycle that moves from noticing to guided analysis to contextualized production can help learners request, hedge, apologize, and signal stance more accurately. The approach has limitations. It requires corpus literacy, depends on corpus representativeness, and still lacks well-developed, pragmatics-annotated ESP resources but future work on accessible discipline-specific corpora, teacher training, and automated annotation could make corpus-based instruction a standard component of ESP teaching.

REFERENCES

1. Bachman, L. F. (1990). Fundamental considerations in language testing. Oxford: Oxford University Press.

2. Bardovi-Harlig, K., Mossman, S., & Su, Y. (2017). The effect of corpus-based instruction on pragmatic routines. Language Learning & Technology, 21(3), 76–103.

3. Kerber, N., Shea, J., & Tecedor, M. (2023). “Sorry, that’s all I know!”: A study on web-based pragmatic instruction for novice learners. Foreign Language Annals, 56(3), 645–666.

4. Timpe-Laughlin, V., Green, A., & Oh, S. (2021). Raising pragmatic awareness: A think-aloud study. System, 98, 102470.

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

6. Chen, M., & Flowerdew, J. (2018). A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom. International Journal of Corpus Linguistics, 23(3), 335–369.

7. Ma, Q., Yuan, R., Cheung, L. M. E., & Yang, J. (2022). Teacher paths for developing corpus-based language pedagogy: A case study. Computer Assisted Language Learning, 1–32.

8. Qi, X., & Chen, Z. (2025). A systematic review of technology integration in developing L2 pragmatic competence. Education Sciences, 15(2), 172.

9. Jablonkai, R. R., & Csomay, E. (Eds.). (2022). *The Routledge handbook of corpora and English language teaching and learning.* Taylor & Francis.

10. Miller, D., & Biber, D. (2015). Evaluating reliability in quantitative vocabulary studies: The influence of corpus design and composition. International Journal of Corpus Linguistics, 20(1), 30–55.