
Corpus-Driven Task Design and Its Effectiveness in Developing Pragmatic Competence among IT English Learners

Mamajonova Manzura Ne'matjon qizi

manzurawebster94@gmail.com

PhD student,

Uzbekistan State World Languages Univeristy

Annotation

This article presents a corpus-driven system of 24 pedagogical tasks designed to develop both pragmatic competence and grammatical accuracy among university-level learners of English for Information Technology (EIT) purposes. The task system draws on a diverse range of authentic corpora – including the Corpus of Contemporary American English (COCA), the British National Corpus (BNC), the Michigan Corpus of Academic Spoken English (MICASE), the Enron Email Corpus, Sketch Engine, Stack Overflow, and GitHub – to ground instruction in real-world professional communication. Task types span dialogues, scenario-based interactions, role-plays, professional email and report writing, online commenting, synchronous chat simulations with international IT specialists, inductive corpus analysis, and learner text production, all anchored in domain-specific IT terminology and grammatical structures. The theoretical framework integrates data-driven learning (DDL), task-based language teaching (TBLT), and interlanguage pragmatics. Drawing on evidence from a pre-/post-test quasi-experimental design, the study demonstrates statistically significant gains in pragmatic awareness, sociopragmatic sensitivity, and targeted grammatical accuracy following task implementation. The findings suggest that multi-corpus task architectures offer a principled and scalable approach to ESP pedagogy, with particular value in contexts where learners require intercultural communicative competence for professional collaboration.

Keywords

Corpus-driven task design, pragmatic competence, data-driven learning, English for IT, TBLT, interlanguage pragmatics, Sketch Engine, COCA, MICASE

Pragmatik kompetensiya takomillashtirishda korpusga asoslangan topshiriqlar dizaynining samaradorligi (IT yo'nalishidagi ingliz tili o'rganuvchilari misolida)

Mamajonova Manzura Ne'matjon qizi

manzurawebster94@gmail.com

PhD doktorant,

O'zbekiston davlat jahon tillari universiteti

Annotatsiya

Mazkur maqolada Axborot Texnologiyalari uchun ingliz tili (English for Information Technology – EIT) yo'nalishida tahsil olayotgan oliy ta'lim talabalari orasida pragmatik kompetensiya hamda grammatik aniqlikni rivojlantirishga qaratilgan, korpus ma'lumotlariga asoslangan 24 ta pedagogik topshiriqdan iborat tizim taqdim etiladi. Topshiriqlar tizimi haqiqiy professional kommunikatsiyaga asoslangan o'qitishni ta'minlash maqsadida turli autentik til korpuslariga – Corpus of Contemporary American English (COCA), British National Corpus (BNC), Michigan Corpus of Academic Spoken English (MICASE), Enron Email Corpus, Sketch Engine, Stack Overflow hamda GitHub manbalariga tayanadi. Topshiriq turlari dialoglar,

vaziyatga asoslangan interaksional, rolli o'yinlar, professional elektron xatlar va hisobotlar yozish, onlayn izoh qoldirish, xalqaro IT mutaxassislari bilan sinxron chat simulyatsiyalari, induktiv korpus tahlili hamda o'quvchilar tomonidan matn yaratish faoliyatlarini qamrab oladi. Barcha topshiriqlar IT sohasiga xos terminologiya va grammatik strukturalar asosida ishlab chiqilgan. Nazariy asos sifatida ma'lumotlarga asoslangan o'qitish (Data-Driven Learning – DDL), vazifaga asoslangan til o'qitish (Task-Based Language Teaching – TBLT) hamda intertil pragmatikasi yondashuvlari integratsiyalashgan. Pre-test va post-testga asoslangan kvazi-eksperimental tadqiqot dizayni natijalari topshiriqlar joriy etilgandan so'ng o'quvchilarning pragmatik xabardorligi, sotsiopragmatik sezgirligi hamda maqsadli grammatik aniqligida statistik jihatdan ahamiyatli o'sish kuzatilganini ko'rsatadi. Natijalar ko'p korpusli topshiriqlar arxitekturasi ESP pedagogikasi uchun nazariy asoslangan va kengaytirilishi mumkin bo'lgan samarali yondashuv ekanini, ayniqsa professional hamkorlik uchun madaniyatlararo kommunikativ kompetensiya zarur bo'lgan ta'lim sharoitlarida yuqori ahamiyat kasb etishini ko'rsatadi.

Kalit so'zlar Korpusga asoslangan topshiriqlar dizayni, pragmatik kompetensiya, ma'lumotlarga asoslangan o'qitish, IT uchun ingliz tili, TBLT, intertil pragmatikasi, Sketch Engine, COCA, MICASE

Эффективность корпусно-ориентированного проектирования заданий в развитии прагматической компетенции у изучающих английский язык для сферы информационных технологий

Мамаджонова Манзура Неъматжон кизи
manzurawebster94@gmail.com
PhD-докторант,
Узбекский государственный университет
мировых языков

Аннотация В данной статье представлена корпусно-ориентированная система из 24 педагогических заданий, направленных на развитие прагматической компетенции и грамматической точности у студентов высших учебных заведений, изучающих английский язык для информационных технологий (English for Information Technology – EIT). Система заданий основана на использовании разнообразных аутентичных языковых корпусов – Corpus of Contemporary American English (COCA), British National Corpus (BNC), Michigan Corpus of Academic Spoken English (MICASE), Enron Email Corpus, Sketch Engine, Stack Overflow и GitHub, что позволяет строить обучение на основе реальной профессиональной коммуникации. Типы заданий включают диалоги, ситуационно-ориентированные взаимодействия, ролевые игры, написание профессиональных электронных писем и отчетов, онлайн-комментирование, синхронные чат-симуляции с международными IT-специалистами, индуктивный корпусный анализ и создание текстов обучающимися. Все задания опираются на специализированную IT-терминологию и соответствующие грамматические структуры.

Теоретическая основа исследования объединяет обучение на основе данных (Data-Driven Learning – DDL), обучение на основе выполнения задач (Task-Based Language Teaching – TBLT) и межъязыковую прагматику. Результаты квазиэкспериментального исследования с использованием pre- и post-тестирования демонстрируют статистически значимое повышение уровня прагматической осведомлённости, социопрагматической чувствительности и целевой грамматической точности после внедрения разработанной системы заданий. Полученные результаты свидетельствуют о том, что многокорпусная архитектура заданий представляет собой теоретически обоснованный и масштабируемый подход к преподаванию ESP, особенно эффективный в образовательных условиях, требующих формирования межкультурной коммуникативной компетенции для профессионального сотрудничества.

Ключевые слова

Корпусно-ориентированное проектирование заданий, прагматическая компетенция, обучение на основе данных, английский язык для IT, TBLT, межъязыковая прагматика, Sketch Engine, COCA, MICASE

Introduction

The intersection of corpus linguistics and language pedagogy has generated sustained scholarly interest over the past three decades (Johns, 1991; Boulton & Cobb, 2017; Pérez-Paredes, 2022). Within this tradition, data-driven learning (DDL) positions learners as active researchers who interrogate authentic language data to discover grammatical patterns, collocational preferences, and pragmatic conventions embedded in real use (Johns, 1991; Gilquin & Granger, 2010). Yet despite a robust empirical record demonstrating DDL efficacy for vocabulary and grammar instruction (Lee & Lin, 2019; Zhong & Wakat, 2023), its application to pragmatic competence – particularly in English for Specific Purposes (ESP) contexts – remains comparatively underexplored (Bardovi-Harlig, Mossman, & Su, 2017; Bouzekria et al., 2023).

This gap carries particular weight in the domain of IT English education. IT professionals operate in highly specialised discourse communities characterised by their own genre conventions, terminological registers, and interactional norms (Stack Overflow, GitHub,

Slack channels, technical documentation). Learners preparing for such environments require not merely grammatical accuracy but pragmalinguistic fluency: the ability to select contextually appropriate speech acts, navigate power asymmetries in professional correspondence, manage face-threatening acts in collaborative code review, and calibrate formality in asynchronous written communication with international colleagues. These competencies are poorly addressed by conventional coursebook materials, which tend to present decontextualised grammar rules and idealised dialogues rather than the messy, genre-specific texts professionals actually encounter (Wu, 2024; Hashemi & Daneshfar, 2020).

The present study addresses this gap through the design, implementation, and evaluation of a system of 24 corpus-driven tasks targeting pragmatic competence and grammatical accuracy among undergraduate IT majors. The task system is distinctive in three respects. First, it draws on a deliberately heterogeneous corpus ecology – encompassing general reference corpora

(COCA, BNC), academic spoken English (MICASE), specialised professional corpora (Enron Email Corpus, Stack Overflow, GitHub), and a powerful corpus querying platform (Sketch Engine) – to expose learners to the full register range of IT professional discourse. Second, the tasks encompass a typologically diverse sequence of activities, from inductive corpus analysis and scenario-based role-play to professional email composition and international chat simulation, thereby operationalising the full input-to-output cline of DDL (Chujo et al., 2018; Rasikawati, 2019). Third, the system integrates explicit instruction in domain-specific IT terminology alongside pragmatic and grammatical objectives, responding to calls for integrated ESP pedagogy (Friginal, 2013).

The article is organised as follows. Section 2 reviews the theoretical foundations of the study. Section 3 describes the corpus resources employed. Section 4 presents the task system and its design rationale. Section 5 reports the methodology and findings of the evaluation study. Section 6 discusses implications and concludes.

Theoretical Framework

Pragmatic Competence in Second Language Contexts

Pragmatic competence is widely understood as encompassing two inter-related dimensions: pragmalinguistic knowledge, which involves the linguistic resources available for performing speech acts, and sociopragmatic knowledge, which involves the social and cultural conditions governing appropriate language use in context (Leech, 1983; Thomas, 1983). For second language learners, these dimensions are particularly susceptible to cross-linguistic interference and transfer, a phenomenon extensively documented under the rubric of interlanguage pragmatics (Kasper & Blum-Kulka, 1993). Research in this area has consistently found that even proficient L2 speakers may exhibit pragmatic failure – producing grammatically well-formed utterances that are contextually

inappropriate – unless explicit pragmatic instruction is provided (Taguchi, 2015; LoCastro, 2012).

A growing body of evidence indicates that pragmatic competence is teachable and benefits from instruction that combines awareness-raising with authentic communicative practice (Jeon & Kaya, 2006; Taguchi & Roever, 2017). Task-based approaches are particularly well-suited to this end: by embedding pragmatic targets within communicative tasks that simulate real-world interactional demands, learners receive both the input and the interactional opportunities necessary to acquire contextually sensitive language use (Kim & Taguchi, 2015; Youn, 2018). Role-play, scenario simulation, and email writing tasks have been shown to promote pragmatic development by requiring learners to calibrate their linguistic choices to specific interlocutors, purposes, and settings (Qin, Jia, & Ren, 2024; Pan, 2023).

Data-Driven Learning and Corpus-Based Instruction

The term data-driven learning was coined by Tim Johns (1991) to describe an inductive, concordance-based approach to grammar and vocabulary instruction in which learners analyse corpus evidence rather than consult prescriptive rules. The foundational premise – that authentic text data constitutes the most reliable guide to how language actually functions – aligns with usage-based theories of language acquisition and positions the learner as an active hypothesis-tester rather than a passive recipient of grammatical rules (Ellis, Römer, & O'Donnell, 2016).

Systematic reviews of DDL research have demonstrated its effectiveness across a range of linguistic targets, including lexical collocation, grammatical constructions, writing proficiency, and, to a lesser extent, pragmatic routines (Boulton & Cobb, 2017; Pérez-Paredes, 2022; Al-Lawati, in Pérez-Paredes, 2022). Crosthwaite and Steeples (2022) demonstrated DDL efficacy for grammar comprehension, while Bouzekria et al. (2023)

found significant improvements in pragmatic routine production among Algerian EFL learners following corpus-based instruction using MICASE materials. Yilmaz and Koban Koç (2020) similarly documented gains in formulaic sequence production following corpus-informed instruction of pragmatic routines in an EFL setting. These findings collectively support the position that corpus-based instruction constitutes a viable and effective approach to pragmatic development when tasks are carefully designed to direct learner attention to target forms in context.

A key conceptual distinction within DDL concerns direct versus indirect uses of corpora (Flowerdew, 2009; Chambers & O'sullivan, 2004). In direct DDL, learners interact with corpus data themselves, formulating queries and analysing concordance output. In indirect DDL, the teacher uses corpus data to inform material design, constructing tasks based on corpus evidence without requiring learners to access the corpus independently. Both approaches have pedagogical value; the present study employs a hybrid model in which some tasks engage learners directly with COCA and Sketch Engine interfaces, while others are built from corpus-derived materials prepared by the instructor.

Task-Based Language Teaching and Grammar Integration

Task-based language teaching (TBLT) provides the overarching pedagogical framework within which the corpus-driven activities are situated. In TBLT, language learning is organised around communicative tasks – purposeful activities that require learners to use language to achieve a non-linguistic outcome – with form-focused instruction embedded within or around task performance (Ellis, 2018; Ellis et al., 2019). This framework accommodates both incidental acquisition through task engagement and intentional learning through pre-task and post-task form-focus phases (Skehan, 1998).

Integrating grammar instruction with pragmatic task work is essential in ESP contexts.

As Friginal (2013) argues, corpus-based grammar instruction allows teachers to present grammatical structures not as abstract rules but as lexico-grammatical patterns instantiated in genre-specific texts, thereby connecting form to communicative function. For IT learners, structures such as conditional constructions, passive voice in technical documentation, modal verbs for hedging in code review, and discourse markers in asynchronous communication are not merely grammatical targets but pragmatic resources with discipline-specific conventions of use (Biber et al., 1999; Davies, as cited in Friginal, 2013).

Corpus Resources Employed

The task system is anchored in a deliberately diverse corpus ecology, designed to expose learners to the full register range of IT professional discourse. Each corpus was selected for its specific pedagogical contribution to the system.

The Corpus of Contemporary American English (COCA), developed by Mark Davies at Brigham Young University, contains over one billion words of American English drawn from spoken, fiction, magazine, newspaper, academic, television, and web genres (Davies, 2012). Its size and genre balance make it particularly suitable for tasks requiring learners to examine how IT-related vocabulary and grammar patterns function across registers – for example, comparing the use of technical conditionals in academic versus informal web discourse, or examining modal frequency in professional correspondence.

The British National Corpus (BNC), comprising 100 million words of late-twentieth-century British English in both spoken and written modes, provides a complementary resource for examining variation in professional register across national varieties of English. For IT learners engaged in international communication, awareness of such variation is pragmatically relevant (Frankenberg-Garcia et al., 2019, as cited in Pérez-Paredes et al., 2025).

The Michigan Corpus of Academic Spoken English (MICASE) offers approximately 1.7 million words of transcribed academic speech from the University of Michigan, encompassing lectures, seminars, office-hour consultations, and study group interactions. For IT learners preparing for collaborative academic and professional environments, MICASE provides authentic models of informal spoken interaction in technical domains – a genre largely absent from published EFL materials. Bardovi-Harlig et al. (2017) demonstrated MICASE's efficacy as a source of corpus-informed pragmatic instruction, a finding replicated by Bouzekria et al. (2023) in EFL classroom settings.

The Enron Email Corpus, consisting of approximately 600,000 professional emails, represents one of the most extensively studied naturally-occurring collections of workplace written communication (Klimt & Yang, 2004). For the purposes of pragmatic instruction, it offers an unparalleled resource for examining how professional writers negotiate requests, manage status differentials, hedge claims, and structure multi-party communicative events in asynchronous electronic discourse – competencies directly relevant to IT professionals.

Stack Overflow and GitHub represent the primary discourse communities in which IT professionals participate globally. Stack Overflow's question-and-answer corpus exposes learners to the distinctive genre conventions of technical knowledge exchange: how questions are formulated, how solutions are presented, how upvoting and commenting conventions index community membership. GitHub's issue tracker and pull request commentary constitute a specialised genre of collaborative code review, with its own face-threat mitigation strategies and evaluative discourse patterns. Including these platforms as corpus sources responds directly to the argument for needs-driven ESP corpus design (Flowerdew, 2015).

Sketch Engine, a commercial corpus analysis platform hosting over 400 pre-compiled corpora and supporting user-built corpus construction (Kilgarriff et al., 2014), was employed both as a tool for learner-facing DDL tasks and as an instructor resource for identifying frequent IT collocations, word sketches, and thematic vocabulary patterns. Its word sketch functionality – which visualises the typical grammatical behaviour of a target word – was particularly valuable for designing tasks targeting technical collocation and nominal phrase structure.

The Task System: Design Principles and Task Typology

Design principles

The 24-task system was developed according to four interlocking design principles. First, authenticity: all tasks are grounded in or derived from authentic corpus data, ensuring that language models reflect real professional practice rather than idealised textbook norms (Widdowson, 2000; Timmis, 2010). Second, progressivity: tasks are sequenced to move from receptive corpus analysis through controlled production to free communicative performance, operationalising the observe-hypothesise-experiment cycle central to DDL pedagogy (Johns, 1991; Chujo et al., 2018). Third, integration: each task addresses pragmatic, grammatical, and terminological objectives simultaneously, reflecting the holistic character of communicative competence in professional contexts. Fourth, ecological validity: task scenarios are drawn from genuine IT professional situations – debugging sessions, project planning emails, code review interactions, technical documentation – to ensure transfer potential.

Task typology

The 24 tasks are distributed across six functional categories. The first category, Corpus Analysis Tasks (Tasks 1-5), engages learners directly with COCA, BNC, and Sketch Engine to investigate lexico-grammatical patterns relevant to IT communication.

Illustrative tasks include comparative concordance analysis of modal verbs (must, should, might, need to) across technical documentation and informal online forums; frequency-based examination of IT-specific collocations in COCA academic versus web subcorpora; and word sketch analysis in Sketch Engine to map the typical syntactic environments of high-frequency IT terminology. These tasks develop corpus literacy alongside metalinguistic awareness of register variation, a precondition for informed pragmatic choice-making (Han, 2024; Pérez-Paredes et al., 2025).

The second category, Dialogue and Scenario Tasks (Tasks 6-10), draws on MICASE transcripts and Enron email data to present learners with authentic interactional scenarios requiring pragmatic decision-making. Learners analyse naturally-occurring exchanges – a senior engineer explaining a bug to a junior colleague, a project manager declining a client request – before reconstructing parallel dialogues in which they must navigate the same face-threatening acts using contextually appropriate mitigation strategies. The MICASE data is used specifically to sensitise learners to the informal register features of professional spoken interaction, including hedging devices, discourse markers, and listener-response tokens, which are pragmatically salient but rarely taught explicitly (Bardovi-Harlig et al., 2017).

The third category, Role-Play Tasks (Tasks 11-13), positions learners as participants in simulated IT workplace interactions: a code review meeting in which one reviewer must give critical feedback to a peer, a technical support call in which the specialist must explain an error message to a non-technical user, and a team standup meeting requiring learners to report progress, obstacles, and next steps. These tasks directly assess sociopragmatic sensitivity by requiring learners to calibrate formality, directness, and technical register to specific interlocutors and power relations. Role-play has been demonstrated to be among

the most effective task types for developing pragmatic production capacity (Youn, 2018).

The fourth category, Professional Writing Tasks (Tasks 14-18), targets email and report genres using the Enron corpus and GitHub pull request commentary as input models. Tasks include drafting a formal request email to a line manager for additional development resources, composing a technical incident report following a server failure, writing a peer code review comment that balances critical evaluation with positive affect, and revising an informationally dense technical summary for a non-specialist audience. These tasks develop pragmatic control of written register alongside grammatical accuracy in complex nominal structures, passive constructions, and hedging expressions characteristic of professional IT discourse (Biber et al., 1999; Qin et al., 2024).

The fifth category, Online Interaction and Commentary Tasks (Tasks 19-21), uses Stack Overflow and GitHub data as both input and production contexts. Learners analyse community-upvoted responses on Stack Overflow to identify the discourse features of expert answers – organisation, hedging, code citation conventions, meta-commentary – before producing their own responses to simulated technical questions. A GitHub-based task requires learners to participate in a structured commenting simulation on a fictional pull request, practising the direct but mitigation-sensitive pragmatic norms of open-source code review. These tasks develop awareness of community-specific pragmatic conventions, which are opaque to learners without explicit instruction.

The sixth category, Intercultural Communication Tasks (Tasks 22-24), simulates asynchronous and synchronous written communication between learners and fictional international IT specialists from diverse L1 backgrounds. Drawing on telecollaborative frameworks in intercultural foreign language education (Vyatkina & Belz, 2006; O'Dowd, as cited in Vyatkina & Belz, 2006), these tasks

require learners to manage intercultural pragmatic challenges: negotiating meeting times across time zones, clarifying ambiguous technical specifications with a non-native-speaking partner, and responding diplomatically to face-threatening feedback from an international colleague. Particular attention is given to the pragmalinguistic strategies – request modification, stance marking, positive politeness – documented as culturally variable in cross-linguistic pragmatics research (Kasper & Blum-Kulka, 1993; Taguchi & Roever, 2017).

Methodology and Findings

Research Context and Participants

The task system was implemented over a 12-week semester with 64 undergraduate IT majors at a state university, divided into an experimental group ($n = 32$) receiving corpus-driven task instruction and a control group ($n = 32$) receiving conventional ESP instruction based on a published IT English coursebook. All participants were studying English as a foreign language in a non-English-dominant national context, with B1–B2 proficiency levels (CEFR) as determined by a pre-study placement assessment.

Instruments and Assessment

Pragmatic competence was assessed using a multi-component battery administered pre- and post-instruction. The battery comprised a discourse completion test (DCT) targeting six speech act categories (requests, refusals, complaints, apologies, suggestions, and expressions of gratitude) instantiated in IT professional scenarios; a pragmatic judgment task requiring learners to evaluate the contextual appropriateness of matched response pairs; and a written production task in which learners composed a professional email under timed conditions. Grammatical accuracy was measured through a targeted error-correction task focusing on modal auxiliary constructions, passive voice, conditional clauses, and nominal group complexity – the four grammatical domains most prominently addressed in the task system.

Inter-rater reliability for the DCT scoring was established through independent rating by two trained assessors, achieving a Cohen's kappa of .84. Statistical analysis employed paired-samples t-tests for within-group comparisons and independent-samples t-tests for between-group comparisons, supplemented by effect size calculations (Cohen's d).

Findings

The experimental group demonstrated statistically significant improvements across all pragmatic competence measures. DCT scores increased from a pre-test mean of 62.4 ($SD = 9.3$) to a post-test mean of 78.1 ($SD = 7.6$), representing a large effect size ($d = 1.89$, $p < .001$). Gains were most pronounced in request modification and refusal strategies – precisely the speech act categories most extensively rehearsed through corpus analysis and role-play tasks – and in the email production task, where post-test responses exhibited markedly greater register sensitivity, as evidenced by reduced directness and increased use of hedging devices. Pragmatic judgment scores similarly improved significantly ($d = 1.42$, $p < .001$), indicating enhanced metalinguistic awareness of pragmatic appropriateness.

Grammatical accuracy gains were significant for all four targeted domains, with the largest effect sizes observed for modal auxiliary use ($d = 1.61$) and passive construction accuracy ($d = 1.44$). The control group showed modest but non-significant gains on all measures, consistent with general proficiency development over the semester. Between-group comparisons at post-test were significant across all measures ($p < .001$), confirming the superior effectiveness of corpus-driven task instruction relative to conventional coursebook instruction.

Qualitative analysis of learner post-task reflections revealed three recurrent themes. First, learners reported heightened awareness of register variation in professional English following corpus analysis tasks,

particularly the contrast between technical documentation and informal forum discourse. Second, role-play and scenario tasks were identified as particularly valuable for developing confidence in professional interaction, with several learners noting that the corpus-derived authenticity of the scenarios made the activities feel meaningfully connected to their future professional lives. Third, some learners initially reported difficulty navigating corpus interfaces, consistent with findings in the DDL literature regarding the steep learning curve associated with direct corpus consultation (Kennedy & Miceli, as cited in Pérez-Paredes et al., 2022); this difficulty attenuated significantly over the first four weeks of instruction.

Discussion and Implications

The findings reported above contribute to a growing body of evidence supporting the effectiveness of corpus-based instruction for pragmatic development in L2 contexts (Bardovi-Harlig et al., 2017; Bouzekria et al., 2023; Yilmaz & Koban Koç, 2020). Several aspects of the present study extend this literature in novel directions.

The multi-corpus design – spanning general, academic, professional, and community-of-practice corpora – appears to be a particularly important feature of the task system's effectiveness. By juxtaposing COCA academic data with Stack Overflow commentary, for example, learners are confronted with genuine register variation that complicates simplistic notions of "correct" professional English, fostering the sociopragmatic sensitivity required for flexible intercultural communication. This design principle responds to Flowerdew's (2015) argument for needs-driven ESP corpus construction and to the broader call for authentic, genre-specific language input in professional language education.

The integration of IT-specific terminology within pragmatic and grammatical tasks addresses a frequently noted limitation of existing ESP pragmatics instruction: the

tendency to treat pragmatic development as domain-neutral, when in reality the pragmatic norms of IT professional discourse are distinctively shaped by community-of-practice conventions (GitHub code review, Stack Overflow Q&A, Agile standups) that require explicit pedagogical attention. The present task system demonstrates that domain-specific pragmatic instruction can be effectively operationalised through corpus-driven task design without sacrificing the communicative authenticity that gives TBLT its pedagogical power.

The significant gains in grammatical accuracy, particularly for modal auxiliary use and passive voice, underscore the value of integrating form-focused instruction within communicative task sequences. As Friginal (2013) argues, corpus data allows grammar instruction to move beyond rule recitation to the demonstration of grammatical form as pragmatic resource – a reconceptualisation that appears to enhance both the salience and the transferability of grammatical knowledge. The present findings are consistent with DDL research demonstrating that corpus-based grammar instruction outperforms conventional approaches (Crosthwaite & Steeples, 2022; Zhong & Wakat, 2023).

Several limitations of the present study warrant acknowledgement. The sample was drawn from a single institutional context, limiting the generalisability of findings across diverse EFL environments. The 12-week implementation period, while sufficient to demonstrate significant gains, does not permit conclusions about the long-term durability of pragmatic competence development. Future research should employ longitudinal designs and examine learner performance in authentic professional settings beyond the classroom. Additionally, the pragmatic assessment instruments, while multi-component, necessarily involve some artificiality; naturalistic data on learner pragmatic performance in genuine IT workplaces would provide valuable validation evidence.

Conclusion

This article has presented and evaluated a 24-task corpus-driven system designed to develop pragmatic competence, grammatical accuracy, and IT-specific communicative competence among EFL learners in technology-oriented academic programmes. By drawing on a diverse ecology of corpora – COCA, BNC, MICASE, the Enron Email Corpus, Sketch Engine, Stack Overflow, and GitHub – the task system exposes learners to the full register spectrum of IT professional discourse and situates language instruction within the genuine communicative contexts of their target domain. The task typology, which encompasses corpus analysis, dialogue reconstruction, role-play, professional writing, online commentary,

and intercultural communication simulation, operationalises the observe-hypothesise-produce cycle central to DDL pedagogy and the communicative task-centredness of TBLT.

The quasi-experimental evaluation study demonstrated significant gains in all targeted pragmatic and grammatical competences, with large effect sizes relative to a conventional coursebook comparison condition. These findings support the integration of corpus technologies into ESP pedagogy for IT contexts and contribute empirical evidence to the broader case for multi-corpus, multi-task pedagogical design as an effective approach to developing professional communicative competence in contemporary academic and workplace settings.

References:

1. Bardovi-Harlig, K., Mossman, S., & Su, Y. (2017). The effect of corpus-based instruction on pragmatic routines. *Language Learning & Technology*, 21(3), 76–103.
2. Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson.
3. Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348–393. <https://doi.org/10.1111/lang.12224>
4. Bouzekria, H., Mashaqba, B., Al Khalaf, E., & Huneety, A. (2023). Production of pragmatic routines by Algerian EFL learners: The effect of corpus-based instruction. *Ampersand*, 10, Article 100122. <https://doi.org/10.1016/j.amper.2023.100122>
5. Chambers, A., & O'sullivan, Í. (2004). Corpus consultation and advanced learners' writing skills in French. *ReCALL*, 16(1), 158–172. <https://doi.org/10.1017/S0958344004000400>
6. Chujo, K., Kobayashi, Y., & Uto, K. (2018). ESP vocabulary learning with data-driven learning in a corpus-based approach. *English Language Teaching*, 11(4), 1–14. <https://doi.org/10.5539/elt.v11n4p1>
7. Crosthwaite, P., & Steeples, R. (2022). Corpus-informed grammar instruction. *Language Teaching Research*, 26(5), 880–905. <https://doi.org/10.1177/13621688211023028>
8. Davies, M. (2012). *Expanding horizons in corpus linguistics: The potential of the Corpus of Contemporary American English*. In A. Hardie & R. McEnery (Eds.), *Corpus linguistics: Critical concepts in linguistics*. 47–72. Routledge.
9. Ellis, R. (2018). *Reflections on task-based language teaching*. Multilingual Matters.
10. Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2019). *Task-based language teaching: Theory and practice*. Cambridge University Press.
11. Ellis, N. C., Römer, U., & O'Donnell, M. B. (2016). *Usage-based approaches to language acquisition and processing*. Wiley-Blackwell.

12. Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics*, 14(3), 393–417. <https://doi.org/10.1075/ijcl.14.3.05flo>
13. Flowerdew, L. (2015). *Corpus-based approaches to language teaching and learning*. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes*. 335–358. Wiley-Blackwell.
14. Friginal, E. (2013). *Corpus linguistics for English teachers: New tools, online resources, and classroom activities*. Routledge.
15. Gilquin, G., & Granger, S. (2010). How can DDL be used in language teaching? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics*. 359–370. Routledge.
16. Han, S. (2024). Corpus-based qualitative analysis of pragmatic marker use in L2 learners. *Eurasian Journal of Applied Linguistics*, 10(3), 287–295.
17. Hashemi, M., & Daneshfar, S. (2020). Validity and reliability in pragmatic assessment. *TESOL Quarterly*, 54(2), 511–534. <https://doi.org/10.1002/tesq.568>
18. Jeon, E., & Kaya, T. (2006). *Effects of L2 instruction on interlanguage pragmatic development: A meta-analysis*. In J. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching*. 165–211. John Benjamins.
19. Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. *ELR Journal*, 4, 1–16.
20. Kasper, G., & Blum-Kulka, S. (Eds.). (1993). *Interlanguage pragmatics*. Oxford University Press.
21. Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). *The Sketch Engine: Ten years on*. *Lexicography*, 1(1), 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
22. Kim, Y., & Taguchi, N. (2015). Promoting task-based pragmatics instruction in EFL classroom contexts: The role of task complexity. *The Modern Language Journal*, 99(4), 656–677. <https://doi.org/10.1111/modl.12273>
23. Klimt, B., & Yang, Y. (2004). *Introducing the Enron corpus*. CEAS 2004 — First Conference on Email and Anti-Spam.
24. Leech, G. (1983). *Principles of pragmatics*. Longman.
25. Lee, P., & Lin, H. (2019). The effect of the inductive and deductive data-driven learning (DDL) on vocabulary acquisition and retention. *System*, 81, 14–25. <https://doi.org/10.1016/j.system.2018.12.011>
26. LoCastro, V. (2012). *Pragmatics for language educators: A sociolinguistic perspective*. Routledge.
27. Pan, T. (2023). Assessing pragmatic competence through naturally occurring data. *Language Assessment Quarterly*, 20(1), 58–76. <https://doi.org/10.1080/15434303.2022.2078499>
28. Pérez-Paredes, P., Ordoñana Guillamón, C., & Bueno Alastuey, M. C. (2022). L2 researchers' use of corpora for grammar and vocabulary. *Language Teaching Research*, 26(4), 566–587. <https://doi.org/10.1177/1362168819894708>
29. Pérez-Paredes, P., Curry, N., & McEnery, T. (2025). Corpus linguistics for language teaching and learning: A research agenda. *Language Teaching*, 58(1), 1–28. <https://doi.org/10.1017/S0261444824000156>
30. Qin, W., Jia, R., & Ren, W. (2024). Pragmatic competence in an email writing task: Influences of situation, L1 background, and L2 proficiency. *Written Communication*, 41(3), 412–438. <https://doi.org/10.1177/07410883241263543>

31. Rasikawati, R. (2019). Corpus-based data-driven learning to augment L2 students' vocabulary repertoire. *International Dialogues on Education*, 6(2), 83–98.
32. Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
33. Taguchi, N. (2015). Instructed pragmatics at a glance: Where instructional studies were, are, and should be going. *Language Teaching*, 48(1), 1–50. <https://doi.org/10.1017/S0261444814000263>
34. Taguchi, N., & Roever, C. (2017). *Second language pragmatics*. Oxford University Press.
35. Thomas, J. (1983). *Cross-cultural pragmatic failure*. *Applied Linguistics*, 4(2), 91–112. <https://doi.org/10.1093/applin/4.2.91>
36. Timmis, I. (2010). Corpora and English language teaching. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics*. 485–497. Routledge.
37. Vyatkina, N., & Belz, J. A. (2006). *A learner corpus-driven intervention for the development of L2 pragmatic competence in networked intercultural language study*. In K. Bardovi-Harlig, C. Félix-Brasdefer, & A. Omar (Eds.), *Pragmatics and language learning*. Vol. 11. 263–299. National Foreign Language Resource Center.
38. Widdowson, H. G. (2000). On the limitations of linguistics applied. *Applied Linguistics*, 21(1), 3–25. <https://doi.org/10.1093/applin/21.1.3>
39. Wu, Y. (2024). Challenges in pragmatic instruction for EFL learners. *TESOL Journal*, 15(2), e720. <https://doi.org/10.1002/tesj.720>
40. Yilmaz, N., & Koban Koç, D. (2020). Developing pragmatic comprehension and production: Corpus-based teaching of formulaic sequences in an EFL setting. *Dil ve Dilbilimi Çalışmaları Dergisi*, 16(2), 474–488. <https://doi.org/10.17263/jlls.712880>
41. Youn, S. J. (2018). Task design and validity evidence for assessment of L2 pragmatics in interaction. In N. Taguchi & Y. Kim (Eds.), *Task-based approaches to teaching and assessing pragmatics*. 217–246. John Benjamins.
42. Zhong, X., & Wakat, G. (2023). *Enhancing grammar proficiency of EFL learners through corpus-integrated lessons*. *Ampersand*, 11, Article 100139. <https://doi.org/10.1016/j.amper.2023.100139>