

DEVELOPING SPEECH ACT COMPETENCE IN ENGLISH THROUGH AI-GENERATED MULTIMODAL TASKS

Erkinova Diyorabonu Muzaffar qizi

Independent researcher

makhkamovadiyora2005@gmail.com

Abstract. This article examines the development of EFL students' speech act competence through AI-generated multimodal tasks. Speech act competence is understood as the ability to recognize communicative intentions and produce contextually appropriate requests, apologies, refusals, complaints, suggestions, offers, and other illocutionary acts. The study proposes a pedagogical framework in which generative artificial intelligence produces combinations of text, audio, image, dialogue, and role-based scenarios. These tasks require learners to interpret social status, interpersonal distance, imposition, tone, and non-verbal cues before selecting an appropriate linguistic strategy. The article argues that AI-supported multimodality can expand pragmatic input, diversify practice, provide adaptive feedback, and strengthen learners' ability to connect linguistic form with communicative function in interaction.

Keywords: speech act competence, artificial intelligence, multimodal tasks, pragmatics, EFL, illocutionary force, contextual appropriateness, feedback.

Introduction. The expansion of artificial intelligence in language education has created new possibilities for teaching not only linguistic forms but also the social and pragmatic dimensions of communication. In English as a foreign language (EFL) contexts, learners often acquire grammar and vocabulary without gaining sufficient experience in choosing expressions that fit particular interpersonal situations. A grammatically correct utterance may still sound excessively direct, impolite, distant, or inappropriate when the relationship between speakers, communicative purpose, or degree of imposition is ignored. Therefore, speech act competence should be treated as

a central component of communicative ability rather than an optional extension of grammar instruction.

The theoretical foundation of speech act competence originates in speech act theory. Austin demonstrated that utterances do not merely describe reality; they can perform actions such as promising, requesting, warning, apologizing, or declaring¹. This perspective established the relationship between linguistic expression and communicative action. Searle further systematized speech acts by explaining the conditions under which utterances function as particular illocutionary acts and by classifying communicative acts according to their functions. For language pedagogy, learners must understand not only what a sentence means but also what action a speaker intends to perform through it.

In EFL learning, the same communicative intention can be realized through several linguistic forms. A request may be expressed through an imperative, a modal question, a hint, or a conventionally indirect formula. A refusal may involve a direct negative response, an explanation, regret, postponement, or an alternative suggestion. The appropriateness of each option depends on social status, familiarity, institutional context, cultural expectations, and the seriousness of the imposition. Kasper and Rose emphasize that second-language pragmatic development is shaped by opportunities for input, interaction, instruction, and participation in communicative contexts². Classroom instruction should therefore expose learners to varied, context-sensitive speech act realizations.

Traditional textbook exercises often provide limited pragmatic variability. Learners may complete a dialogue, select a polite phrase, or memorize formulaic expressions, yet such activities do not always represent the interaction among language, social relations, tone, facial expression, gesture, and situational context. Taguchi notes that L2 pragmatics instruction benefits from systematic attention to

¹ Austin J.L. How to Do Things with Words. – Oxford: Clarendon Press, 1962. – 166 p.

² Searle J.R. Speech Acts: An Essay in the Philosophy of Language. – Cambridge: Cambridge University Press, 1969. – 203 p.

pragmatic forms, meanings, strategies, materials, and opportunities for practice. Consequently, speech act instruction should move beyond isolated sentence-level exercises toward tasks in which learners interpret, compare, produce, and evaluate communicative choices. Multimodality provides an appropriate basis for this shift. Kress explains that meaning is produced through multiple semiotic modes, including language, image, sound, gesture, and spatial organization. In real communication, an apology such as “*I’m sorry*” may communicate sincerity, irritation, embarrassment, or formality depending on intonation, facial expression, timing, and gesture. A refusal may appear polite in written form but become dismissive when accompanied by a sharp tone. Multimodal tasks can therefore help learners recognize that pragmatic meaning is distributed across verbal and non-verbal resources.

Generative AI strengthens this potential by enabling teachers and learners to create contextual versions of the same speech act³. AI can generate workplace requests, academic apologies, informal refusals, customer-service complaints, peer suggestions, or invitations across different levels of formality. It can also modify variables such as speaker status, familiarity, urgency, and communicative goal. Kohnke, Moorhouse, and Zou identify dialogue generation, individualized practice, feedback, and adaptable language-learning activities among the educational affordances of generative AI tools. These affordances are relevant because exposure to varied situations can make the relationship between context and linguistic choice more visible.

AI-generated multimodal tasks may be organized around noticing, interpretation, comparison, production, and reflection⁴. Learners can observe an image or video frame representing a situation, listen to an exchange, and identify the intended speech act. They can then examine alternative AI-generated responses and determine which option best matches the speakers’ relationship and purpose. In production,

³ Kasper G., Rose K.R. Pragmatic Development in a Second Language // Language Learning. – 2002. – Vol. 52, Supplement 1. – 362 p.

⁴ Taguchi N. Teaching Pragmatics: Trends and Issues // Annual Review of Applied Linguistics. – 2011. – Vol. 31. – P. 289–310.

learners formulate their own response or continue an AI-mediated role-play. Finally, they compare their answer with alternative realizations and explain why one strategy is more appropriate. This sequence shifts attention from formally correct English toward reasoned pragmatic choices.

A major pedagogical advantage of AI is its capacity to vary one contextual factor while keeping communicative intention constant. Learners may practice requesting an extension from a close classmate, a lecturer, and an administrator. Although the content remains similar, the expected level of directness, mitigation, address terms, and explanation changes. Such contrastive tasks make sociopragmatic relationships explicit and encourage learners to notice how choices respond to status and distance.

AI-generated language, however, should not be treated as an unquestionable pragmatic norm. Responses may reflect dominant varieties of English or incomplete interpretations of context. The instructional value of AI lies in providing material for analysis rather than replacing teacher judgment. Teachers remain responsible for selecting suitable scenarios, checking plausibility, encouraging comparison, and discussing cultural variation. Learners should evaluate AI output critically and justify their preferred speech act realization. Used in this way, AI-generated multimodal tasks can support context-sensitive, reflective, and flexible speech act competence.

Table 1. Framework of AI-Generated Multimodal Tasks for Developing Speech Act Competence.

Speech act	Contextual focus	AI-generated multimodal task	Student activity	Expected competence outcome
<i>Requests</i>	Status, distance, degree of imposition	AI-generated workplace, academic, and peer scenarios combining dialogue, image, and audio	Compare direct and indirect requests and select the most appropriate form	Context-sensitive requesting strategies

<i>Apologies</i>	Responsibility, seriousness, interpersonal relationship	Audio-visual scenarios presenting different levels of offence and emotional response	Identify apology components and produce an appropriate repair strategy	Appropriate expression of apology and responsibility
<i>Refusals</i>	Power relations, politeness, face management	AI role-play involving invitations, requests, and offers	Compare direct refusal, explanation, regret, and alternative responses	Strategic and polite refusal competence
<i>Complaints</i>	Formality, dissatisfaction, institutional context	Customer-service and academic complaint scenarios with tone and facial-expression cues	Reformulate overly direct complaints and justify linguistic choices	Appropriate management of dissatisfaction
<i>Suggestions and advice</i>	Familiarity, authority, involvement	AI-generated peer, teacher–student, and workplace dialogues	Select modal, hedged, or direct strategies according to context	Flexible suggestion and advice strategies
<i>Offers and invitations</i>	Social distance, relationship, communicative purpose	Interactive AI scenarios combining text, visual prompts, and simulated responses	Produce, accept, or decline offers and invitations appropriately	Adaptive interpersonal speech act performance

The table illustrates how AI-generated multimodal tasks can connect particular speech acts with contextual variables, semiotic resources, and learner decisions. Requests and refusals receive attention because they often require careful management of social distance, status, and imposition⁵. In these tasks, students do not reproduce one fixed formula; instead, they compare alternatives and explain why a specific realization is more appropriate in a given setting. Apology tasks emphasize sincerity,

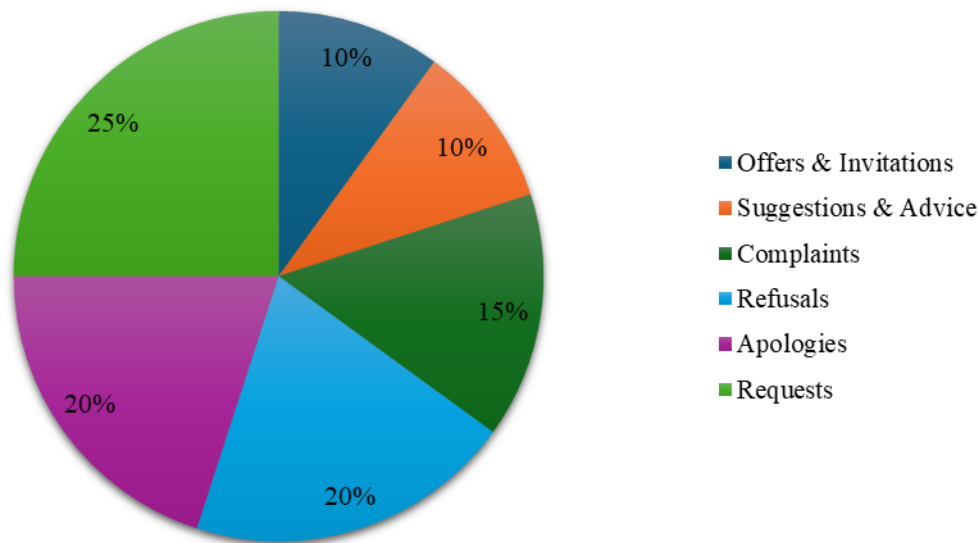
⁵ Kress G. Multimodality: A Social Semiotic Approach to Contemporary Communication. – London–New York: Routledge, 2010. – 232 p.

responsibility, and repair strategies, while complaint tasks focus on balancing dissatisfaction with interpersonal appropriateness. Suggestions, advice, offers, and invitations provide opportunities to distinguish levels of directness and involvement.

The multimodal dimension is essential because pragmatic interpretation may change when verbal content is combined with voice quality, facial expression, gesture, or situational imagery. AI can generate parallel scenarios in which one factor changes, allowing learners to observe how pragmatic choices shift across contexts. The teacher can then direct attention to particular forms, discuss cultural differences, and prevent learners from treating AI output as a universal standard⁶. Accordingly, the proposed task system integrates receptive noticing with productive performance: students first identify communicative intention, then evaluate alternatives, produce their own response, and finally revise it through reflective feedback. Such sequencing supports the gradual transformation of speech act knowledge into context-sensitive communicative competence.

Figure 1. Recommended Distribution of AI-Generated Multimodal Speech-Act Tasks

⁶ Kohnke L., Moorhouse B.L., Zou D. ChatGPT for Language Teaching and Learning // RELC Journal. – 2023. – Vol. 54, No. 2. – P. 537–550.



The pie chart presents a recommended, rather than empirical, distribution of AI-generated multimodal tasks for instructional planning. Requests receive the largest proportion because they permit systematic manipulation of status, distance, imposition, and mitigation. Apologies and refusals are also emphasized because learners must coordinate linguistic strategies with interpersonal sensitivity and contextual expectations. Complaints receive a moderate share, while suggestions, advice, offers, and invitations provide supplementary practice in less confrontational but still context-dependent acts. The distribution is intended as a flexible design template rather than a fixed syllabus. Teachers may adjust the proportions according to learners' proficiency, curricular objectives, diagnostic results, and local communicative needs. Its main methodological purpose is to ensure balanced exposure while allocating greater practice to speech acts requiring more intensive sociopragmatic negotiation and strategic linguistic choice.

Conclusion. In conclusion, AI-generated multimodal tasks offer a promising methodological resource for developing EFL students' speech act competence because they combine contextual variation, pragmatic choice, multimodal input, interactive production, and reflective feedback within one learning environment. Their principal

value lies in moving instruction beyond memorized formulas toward the interpretation and production of contextually appropriate communicative actions. By manipulating social status, interpersonal distance, degree of imposition, tone, and situational purpose, AI can provide learners with repeated opportunities to compare alternative realizations of requests, apologies, refusals, complaints, suggestions, offers, and invitations.

At the same time, effective implementation requires pedagogical control. AI-generated responses should be evaluated critically rather than presented as unquestionable models of English pragmatics. Teacher mediation remains necessary for explaining sociocultural variability, identifying pragmatic mismatches, and guiding learners toward justified communicative decisions. Multimodal resources should likewise be selected purposefully so that audio, image, gesture, and visual context contribute directly to pragmatic interpretation.

The proposed approach therefore positions AI as a generator of communicative possibilities, not as a substitute for human judgment. When integrated with noticing, comparison, role-play, feedback, and reflection, AI-supported multimodal tasks can strengthen learners' ability to connect linguistic form with illocutionary force, interpersonal meaning, and situational appropriateness in English communication.

REFERENCES

1. Austin J.L. *How to Do Things with Words*. – Oxford: Clarendon Press, 1962. – 166 p.
2. Kasper G., Rose K.R. *Pragmatic Development in a Second Language* // *Language Learning*. – 2002. – Vol. 52, Supplement 1. – 362 p.
3. Kohnke L., Moorhouse B.L., Zou D. ChatGPT for Language Teaching and Learning // *RELC Journal*. – 2023. – Vol. 54, No. 2. – P. 537–550.
4. Kress G. *Multimodality: A Social Semiotic Approach to Contemporary Communication*. – London–New York: Routledge, 2010. – 232 p.

5. Searle J.R. *Speech Acts: An Essay in the Philosophy of Language*. – Cambridge: Cambridge University Press, 1969. – 203 p.
6. Taguchi N. Teaching Pragmatics: Trends and Issues // *Annual Review of Applied Linguistics*. – 2011. – Vol. 31. – P. 289–310.