



## AI-assisted task-based language teaching: EFL learners' acquisition of collocations through ChatGPT and movie language

Francesca Poli<sup>a</sup> and Valentina Morgana<sup>b</sup>

<sup>a</sup>Faculty of Linguistic Sciences and Foreign Literatures, Università Cattolica del Sacro Cuore, , [francesca.poli@unicatt.it](mailto:francesca.poli@unicatt.it) and <sup>b</sup>Università Cattolica del Sacro Cuore, , [valentina.morgana@unicatt.it](mailto:valentina.morgana@unicatt.it)

How to cite: Poli, F.; Morgana, V. (2025). AI-assisted task-based language teaching: EFL learners' acquisition of collocations through ChatGPT and movie language. In Y. Choubsaz, P. Díez-Arcón, A. Gimeno-Sanz, V. Morgana, A. C. Murphy & F. L. Seracini (Eds.), *Advancing CALL: New research agendas - EUROCALL 2025 Short Papers*. <https://doi.org/10.4995/EuroCALL2025.2025.21279>

---

### Abstract

*This study investigates the potential of generative artificial intelligence (AI), specifically ChatGPT primed with data from the American Movie Corpus (AMC), to support the acquisition of high-frequency spoken collocations in a general English context. Thirty-four B2-level Italian university students were randomly assigned to guided AI, unguided AI, or non-AI control conditions and completed a ten-week task-based language teaching (TBLT) programme. Eight target collocations (e.g., good thing, little problem, get home) were selected based on frequency and statistical association measures ( $t$ -score > 2;  $MI$  > 3) in the AMC. Learner output was analysed using a linear mixed-effects model to compare collocation frequency at pretest, posttest and delayed posttest. While no statistically significant differences emerged between instructional conditions, the findings point to important trends that suggest potential benefits of AI-mediated tasks. These results are interpreted with caution, given small, unequal groups and a conservative bigram operationalisation that may underestimate more flexible patterns. While the integration of corpus-informed input into AI-mediated tasks remains promising, sustained exposure, stronger focus-on-form, and broader pattern coding may be required to detect gains.*

**Keywords:** *GenAI; spoken collocations; task-based language teaching; corpus-informed pedagogy.*

---

## 1. Introduction

Collocational competence is a core component of fluent spoken English, as it enhances naturalness, clarity, and idiomaticity. Yet for many learners, collocations are often considered difficult to acquire, especially in spontaneous speech (Nation, 2013; Bestgen & Granger, 2014; Durrant, 2014; Lundell, 2020; among others). This difficulty is compounded in general English contexts, where lexical diversity and idiomatic use are critical but not easily taught through traditional classroom methods.

Generative AI tools such as ChatGPT could be potential mediators for collocation development: these tools can simulate meaningful dialogue, provide exposure to authentic patterns, and scaffold learner production through interaction. In this study, we investigate whether guided or unguided use of ChatGPT — specifically, a version trained on movie dialogues from the American Movie Corpus (AMC; Forchini, 2021) — can enhance learners' spoken collocational proficiency in general English.

We build on recent research that combines task-based language teaching (TBLT), corpus-informed language input, and generative AI to create authentic learning conditions. This paper reports on a three-group design (guided, unguided, control) with Italian university students and measures the frequency and accuracy of eight high-frequency collocations over time.

We ask:

1. Can ChatGPT – trained on a corpus of American movie dialogues (AMC; Forchini, 2021) – effectively foster the acquisition of high-frequency spoken collocations?
2. How do guided AI-mediated activities compare with unguided interactions in enhancing collocational competence?

Within this framework, TBLT approach was chosen for its learning-by-doing, goal-oriented, and student-centred approach. It supports language learning through real tasks that mirror everyday use, helping students engage with the language in context. Instead of focusing directly on form, it relies on the idea of “noticing” (Schmidt, 1990) as learners become aware of useful language patterns while working on meaningful tasks. This approach aligns perfectly with the use of technology, especially AI, as both are based on similar principles, such as active learning, clear goals, and scaffolding (González-Lloret & Ortega, 2014; Kim & Namkung, 2024).

## **2. Literature review**

### **2.1. Spoken collocations in general English**

Collocations — frequently co-occurring word pairs — are central to fluent speech and idiomaticity (Firth, 1957; Sinclair, 1991; Pawley & Syder, 1983). However, research consistently shows that collocations remain difficult to acquire, even at advanced levels (Granger, 1998b; Laufer & Waldman, 2011).

Much of this evidence comes from written corpora, leaving a gap in our understanding of spoken collocational competence (Myles, 2015). Spoken collocations play unique roles in turn-taking, stance-marking, and building rapport in conversation. Yet, learner speech often exhibits a lower density and diversity of these expressions compared to native speech (Bestgen & Granger, 2014; Altenberg, 1998). The underuse of high-frequency collocations (as indicated by t-score) and infrequent but strongly associated ones (as indicated by MI) can contribute to less natural and more effortful interaction.

Against this backdrop, our study focuses on eight frequent collocations from the American Movie Corpus that exemplify natural conversational English: *good thing*, *little problem*, *right thing*, *get home*, *time for*, *know truth*, *just got*, and *really appreciate*. We examine how targeted interaction with ChatGPT — either guided or unguided — can mediate their acquisition in spoken production.

### **2.2. TBLT and collocation development**

TBLT emphasises meaningful, goal-oriented interaction and can be aligned with technology-enhanced settings (Ellis, 2003; Schmidt, 1990). In AI-mediated task cycles, models such as ChatGPT can provide exemplars, reformulations, and rehearsal opportunities (González-Lloret and Ortega, 2014); however, benefits depend on task design and how learners are guided to notice and reuse target patterns (Kim & Namkung, 2024; Huang et al., 2024).

At the same time, successful AI integration depends on strong pedagogical alignment, institutional backing, and targeted teacher training (Kim & Namkung, 2024). Without these elements, there is a risk that technology will function merely as a superficial add-on, rather than as a meaningful component of task-based learning.

As far as collocations are concerned, learner corpus research suggests that they are more likely to be acquired when embedded in communicative tasks that reflect real-life discourse demands (Durrant & Schmitt, 2009). In this light, spoken collocations such as “really appreciate” or “just got” benefit from repetition and contextualisation across diverse scenarios, such as those that can be generated by AI tools. Building on this, we primed AI

interactions with collocations observed in the AMC to increase the salience and recycling of conversational language.

### 2.3. Guided vs unguided learning with AI

A key variable in AI-mediated instruction is the extent of guidance provided to learners. In guided conditions, learners receive structured support, such as explicit instructions, targeted prompts, metalinguistic explanations, or scaffolded tasks that attract the learners' attention to specific forms or language features.

On the other hand, unguided learning engages learners in self-directed or exploratory interactions with minimal external support (cf. Ellis, 2003; Hulstijn, 2001). In AI-mediated environment, unguided learning can involve activities such as open-ended writing, unstructured dialogues, or problem-solving with no focus on specific language targets.

The growing use of AI in language learning prompts important considerations regarding the degree of instructional support learners require to take full advantage of these tools. On the one hand, AI can replicate aspects of guided instruction by supplying corrections, alternative phrasings, or explanations whenever requested. On the other hand, it also facilitates open-ended exploration, allowing learners to experiment, iterate, and reflect at a self-directed pace. Previous studies indicate that the relative benefits of guided versus unguided AI use may be influenced by factors such as the learner's proficiency level, the complexity of the task, and the intended pedagogical outcomes (Liu et al., 2025; Sako, 2024; Huang et al., 2024).

As AI becomes more embedded in language education, it is increasingly important to examine how varying degrees of scaffolding interact with learner autonomy and AI mediation. Such understanding is essential for optimising learning outcomes and for creating tasks that balance support with independence.

### 2.4. The AMC and target collocations

The AMC is a purpose-built database of scripted dialogue drawn from a balanced selection of 50 American films produced between the mid-20th century and the early 2020s. It was compiled to capture the breadth of language used in popular cinematic narratives, covering multiple decades, genres, and communicative settings. The corpus comprises over half a million words, representing a diverse range of social interactions, registers, and communicative functions. Its design aimed to reflect the linguistic diversity of mainstream films while maintaining representativeness across key variables such as release decade, film genre, and gender balance among speakers.

One of the corpus's strengths lies in its focus on naturally flowing, colloquial English as scripted for performance. Although fictional, movie dialogue is often carefully crafted to sound authentic, making it a rich source of everyday conversational structures, pragmatically salient expressions, and high-frequency collocations. At the same time, its representativeness requires caution, and future work should triangulate with spontaneous corpora where possible.

For this study, the AMC served as the input base for priming ChatGPT to respond using naturalistic conversational English. The eight target collocations — *good thing*, *little problem*, *right thing*, *get home*, *time for*, *know truth*, *just got*, and *really appreciate* — were selected for their high frequency in the AMC (cf. Methodology). These items appear across a variety of pragmatic contexts in the corpus, ranging from expressions of evaluation (*good thing you called*) to problem-solving (*we've got a little problem*), ethical stance (*do the right thing*), and personal narrative (*I just got back from work*).

The integration of AMC-derived collocations into the AI-mediated tasks ensured that learners were exposed to lexically and pragmatically authentic language that had been observed in multiple, varied interactional contexts. Because the AMC spans decades and genres, the collocations were not tied to a single setting but occurred in workplace exchanges, family interactions, friendships, and confrontational dialogue. This diversity not only increased the likelihood of transfer to learners' own communicative situations but also supported the incidental

recycling of these expressions during task performance.

The priming process involved embedding sample AMC collocation instances into ChatGPT prompts, which encouraged the model to generate responses containing these forms in ways that were contextually appropriate and idiomatically natural. In this way, the AMC acted as a bridge between corpus-informed language pedagogy and AI-mediated task-based learning, grounding the interaction in authentic, frequent, and pragmatically salient lexical patterns.

### **3. Methodology**

This research adopted a mixed-methods, quasi-experimental approach to explore how guided and unguided ChatGPT-mediated tasks influence the acquisition of spoken collocations among upper-intermediate (B2) university learners. The broader project involved 75 participants, from which a subset of 34 was selected for the present analysis based on the completeness and quality of their recorded outputs. All participants were first- or second-year undergraduates in foreign language programmes, and their proficiency in English was confirmed using the Oxford Placement Test, administered one week before the study commenced.

Over a ten-week period, each learner generated seven text-based outputs through interactions with ChatGPT 4o (Premium version): a pretest, an immediate posttest, a delayed posttest, and four task-based exchanges (T1–T4), resulting in 178 data points. The control group (non-AI) used ChatGPT solely for the pretest, posttest, and delayed posttest; their task outputs were produced instead with the aid of internet resources or other non-AI digital tools. The design compared three instructional modes: a guided AI condition, an unguided AI condition, and a control condition without AI-supported task performance.

Participants were allocated at random to one of three instructional conditions. In the guided group (N = 15), learners engaged with ChatGPT through structured activities incorporating targeted prompts, attention-directing questions, and clearly defined objectives aimed at encouraging the use of the selected collocations. The unguided group (N = 10) undertook the same set of tasks but worked with ChatGPT in a more open-ended manner, without prompts or scaffolding specifically designed to highlight the target language items. The control group (N = 10) also completed the same tasks but without ChatGPT or any other AI tools. While excluded from AI support, these students were permitted to use alternative digital resources such as websites, search engines, and spell checkers.

The tasks, grounded in Ellis's (2023) TBLT framework, simulated real-world communicative situations but did not explicitly present or require the target collocations, allowing participants to notice and use them as part of task performance. Examples included requesting help to communicate a difficult or surprising situation to a friend, making a decision, or resolving a conflict. Time allocation was consistent across groups: 60 minutes for the four tasks combined, and 90 minutes each for the pretest, posttest, and delayed posttest. All ChatGPT-mediated exchanges were archived and transcribed for subsequent quantitative and qualitative analysis.

The eight target collocations were chosen for their high frequency in the AMC and their communicative value in everyday English. Selection was based not only on raw frequency but also on statistical association strength, ensuring that the items represented meaningful, recurrent lexical partnerships. Specifically, candidates were retained if they met two statistical thresholds: a t-score greater than 2, indicating high overall occurrence and stability, and a mutual information (MI) score greater than 3, signalling a stronger-than-chance co-occurrence relationship.

To track usage in learner output, only exact lemma-based matches were counted. This meant that the core collocation could appear in any inflected form and still be included in the analysis — for example, *get home* encompassed variants such as *got home* or *getting home*, provided that the lexical pairing remained intact. This approach ensured that both the frequency and the flexibility of the target expressions could be measured while avoiding overcounting of unrelated sequences. However, this may underestimate acquisition for items that frequently occur as extended colligational frames (e.g., *just got + X: just got paid / just got easier / just go in*). We therefore flag this measurement limitation and a direction for future pattern-based coding that allows one token to

the right of the node.

Learner production was uploaded into Sketch Engine (Kilgariff et al., 2015). A lemma-based SQL search verified the presence and frequency of each of the eight collocations. We modelled changes in collocational frequency across three timepoints (Pretest, Posttest, Delayed Posttest) and three groups (guided AI, unguided AI, control) using a linear mixed-effects model (LME) with Group and Time as fixed effects and Participant as a random intercept to account for repeated measures. Given the unequal, small groups, we emphasise estimates and confidence intervals over dichotomous decisions and interpret null effects as inconclusive rather than definitive.

#### 4. Results

The linear mixed-effects model comparing collocation frequency at Pretest, Posttest, and Delayed Posttest revealed a significant main effect of Time,  $F(6, 197.8) = 6.382, p < .001$ , indicating that frequency of collocation use changed over time. Pairwise estimates showed a general reduction of frequency from Pretest to Delayed Posttest ( $Estimate = -0.2408, p = .022$ ), while no difference was observed between Posttest and Delayed. No significant main effect of Group ( $p = .282$ ) or Time  $\times$  Group interaction ( $p = .476$ ) was found, suggesting that changes in collocation use occurred consistently across instructional conditions. These findings indicate a general decline in frequency over time, regardless of whether learners engaged in guided AI tasks, unguided AI interactions, or non-AI-based instruction. Random intercepts for participants accounted for approximately 7% of the total variance ( $ICC = 0.069$ ), suggesting results were largely driven by the task or time points, rather than by learner differences.

Overall, we did not detect measurable gains in the use of the eight target collocations over ten weeks, irrespective of guided or unguided AI mediation. We are cautious in interpreting this pattern for three reasons. First, group sizes were small and unequal, which reduces precision and power in between-group contrasts. Second, our conservative bigram operationalization likely undercounted more flexible realisations, particularly for items like *just got*, which commonly occur in extended frames. Third, while AMC-informed priming increases exposure to conversational patterns, scripted speech may still differ from spontaneous speech, so observable effects may require longer exposure, more attention to form, or pattern-wider coding to surface gains.

Pedagogically, our findings suggest positioning generative AI as part of a broader instructional ecosystem: corpus-informed pre-task input, guided noticing of collocational patterns, structured rehearsal, and targeted feedback may be necessary to translate exposure into production gains. Extending the duration and intensity of tasks and expanding coding to include near-neighbour frames (e.g., *just got + [POS]*, *really appreciate + [NP /that-clause]*) could yield a more sensitive measure of development.

#### 5. Conclusion

This study set out to examine whether guided or unguided interactions with ChatGPT, primed on authentic conversational data from the American Movie Corpus, could foster the acquisition of high-frequency spoken collocations in a general English context. Although the linear mixed-effects analysis revealed no statistically significant effects for either instructional condition, the findings contribute to a growing body of work highlighting that collocational development—particularly in spontaneous speech—is a gradual process that depends on sustained, repeated exposure and practice (e.g., Durrant & Schmitt, 2009; Laufer & Waldman, 2011). The absence of measurable gains over a ten-week period may be attributable to several factors, including the design constraints (small, unequal groups) and measurement choices (conservative bigram matching).

Despite these limitations, the study makes a novel contribution by integrating corpus-informed collocations from the AMC into AI-mediated task-based learning for general English—an area that has received far less attention than EAP-focused AI-collocation research. By embedding high-frequency, pragmatically salient expressions into guided and unguided ChatGPT interactions, the design explored how large language models can deliver conversational input grounded in authentic usage patterns across diverse communicative contexts.

Pedagogically, the results caution against assuming that short-term AI-mediated exposure, even when corpus-informed, will automatically lead to measurable gains. Instead, they suggest that generative AI should be positioned as part of a broader instructional ecosystem, ideally supplemented with varied communicative practice and teacher-led or peer-mediated feedback. Future research should lengthen exposure and adopt coding that captures extended colligations, thereby providing a fuller test of whether AI-mediated, corpus-informed tasks can accelerate collocational development in spontaneous speech.

## References

- Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 101–122). Oxford University Press.
- Bestgen, Y., & Granger, S. (2014). Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of Second Language Writing*, 26, 28–41. <https://doi.org/10.1016/j.jslw.2014.09.004>
- Biskup, D. (1990). Some remarks on combinability: Lexical collocations. In J. Arabski (Ed.), *Foreign language acquisition papers*, (pp. 31–44). Uniwersytet Slaski.
- Cowie, A. P. (1992). Multiword lexical units and communicative language teaching. In P. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 1–12). Palgrave Macmillan.
- Durrant, P. (2014). Corpus frequency and second language learners' knowledge of collocations: A meta-analysis. *International Journal of Corpus Linguistics*, 19(4), 443–477. <https://doi.org/10.1075/ijcl.19.4.01dur>
- Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics in Language Teaching*, 47(2), 157–177. <https://doi.org/10.1515/iral.2009.007>
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- Forchini, Pierfranca (Ed.). 2021. The American Movie Corpus: A tool for the development of spoken lexico-grammatical competence. EDUCatt.
- Firth, J. R. (1957). *Papers in linguistics 1934-1951*. Oxford University Press.
- González-Lloret, M., & Ortega, L. (2014). Towards technology-mediated TBLT: An introduction. In M. González-Lloret & L. Ortega (Eds.), *Technology-mediated TBLT: Researching technology and tasks* (pp. 1–22). John Benjamins.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and lexical phrases. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 145–160). Oxford University Press.
- Gyllstad, H. (2005). Words that go together well: Developing test formats for measuring learner knowledge of English collocations. (The department of English in Lund: Working papers in linguistics, vol. 5). <http://www.englund.lu.se/images/stories/pdf-files/workingspapers/vol05/gyllstad-wp-05.pdf>
- Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian learners cope with English vocabulary. *International Journal of Applied Linguistics*, 4(2), 237–258. <https://doi.org/10.1111/j.1473-4192.1994.tb00065.x>

- Huang, X., Wu, Y., & Dou, A. (2024). AI-enhanced task-based language teaching: Fostering personalized college English learning. *Frontiers in Educational Research*, 7(10), 204-209. <https://doi.org/10.25236/FER.2024.071033>
- Hulstijn, J. H. (2001). The use of computer technology in experimental studies of second language acquisition: A survey of some techniques and some ongoing studies. *Language Learning & Technology*, 3(2), 33-45.
- Jiang, F., & Hyland, K. (2025). Metadiscursive nouns in academic argument: ChatGPT vs. student practices. *Journal of English for Academic Purposes*, 75, 101514. <https://doi.org/10.1016/j.jeap.2025.101514>
- Kim, H., & Namkung, J. M. (2024). Methodological characteristics in technology-mediated task-based language teaching research: Current practices and future directions. *Annual Review of Applied Linguistics*, 44, 1-23. <https://doi.org/10.1017/S0267190524000096>
- Laufer, B., & Waldman, T. (2011). Verb–noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, 61(2), 647–672. <https://doi.org/10.1111/j.1467-9922.2010.00621.x>
- Liu, X. J., Wang, J., Zou, B. (2025). Evaluating an AI speaking assessment tool: Score accuracy, perceived validity, an oral peer feedback as feedback enhancement. *Journal of English for Academic Purposes*, 75, 101505. <https://doi.org/10.1016/j.jeap.2025.101505>
- Lundell, F. F. (2020). Formulaicity. In N. Tracy-Ventura & M. Paquot (Eds.), *The Routledge handbook of second language acquisition and corpora* (pp. 370–381). Routledge. <https://doi.org/10.4324/9781351137904-32>
- Myles, F. (2015). Second language acquisition theory and learner corpus research. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 309-331). Cambridge University Press. <https://doi.org/10.1017/CBO971139649414.016>
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. John Benjamins.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–226). Routledge.
- Poli, F. (2024). Learner adverb+ adjective collocations: A study on frequency and collocationality. *L'Analisi Linguistica e Letteraria*, 32(3). <http://www.analisilinguisticaeletteraria.eu/index.php/ojs/article/view/648>
- Sako, T. (2024). Enhancing critical thinking through AI-assisted collaborative task-based learning: A case study of prospective teachers in Japan. *Journal of English Language Teaching and Linguistics*, 9(2), 157-170. <http://dx.doi.org/10.21462/jeltl.v9i2.1319>
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>
- Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 1–22). John Benjamins.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

Zhang, X., & Umeanowai, K. O. (2025). Exploring the transformative influence of artificial intelligence in EFL context: A comprehensive bibliometric analysis. *Education and Information Technologies*, 30(3), 3183-3198. <https://doi.org/10.1007/s10639-024-12937-z>

Zou, B., Guan, X., Shao, Y., & Chen, P. (2023). Supporting speaking practice by social network-based interaction in Artificial Intelligence (AI)-assisted language learning. *Sustainability*, 15(4), 2872. <https://doi.org/10.3390/su15042782>