



PDF Download
3776734.3794484.pdf
24 March 2026
Total Citations: 0
Total Downloads: 83

 Latest updates: <https://dl.acm.org/doi/10.1145/3776734.3794484>

SHORT-PAPER

Towards a Usage-Based Pedagogy for Second Language Learning with Robots

EVA VERHELST, Ghent University, Ghent, VOV, Belgium

TONY BELPAEME, Ghent University, Ghent, VOV, Belgium

Open Access Support provided by:

Ghent University

Published: 16 March 2026

[Citation in BibTeX format](#)

HRI '26: 21st ACM/IEEE International
Conference on Human-Robot Interaction
March 16 - 19, 2026
Scotland, Edinburgh, UK

Conference Sponsors:
SIGAI
SIGCHI

Towards a Usage-Based Pedagogy for Second Language Learning with Robots

Eva Verhelst
eva.verhelst@ugent.be
Ghent University - imec
Ghent Belgium

Tony Belpaeme
tony.belpaeme@ugent.be
Ghent University - imec
Ghent Belgium



Figure 1: Examples of use-based language tutoring with social robots [1, 27].

Abstract

Recent advances in generative AI and social robotics have opened new possibilities for robot-assisted language learning, yet integrating these technologies in pedagogically sound ways remains challenging. This paper matches theories of language learning to the design of autonomous robot tutors. Usage-based language learning, learning in context, Self-Determination Theory and Dual Coding Theory lend themselves to being operationalised for Robot-Assisted Language Learning. We present a proof-of-concept shared story-building system, in which a learner co-creates a story with a robot tutor. The system leverages large language models for dynamic content generation, automatic speech recognition for learner input, and image generation to provide multimodal scaffolding. By embedding vocabulary, adapting to learner input, and avoiding explicit corrections, the system aligns with usage-based and interactionist theories of language acquisition. We discuss the technological enablers and barriers — such as large language model adaptability and automatic speech recognition limitations — and propose directions for future work. This work contributes to the growing field of AI-powered social robots in education, demonstrating how theory-driven design can enhance engagement and learning outcomes.

CCS Concepts

- **Computing methodologies** → **Natural language processing**;
- **Human-centered computing** → Natural language interfaces; •
- Computer systems organization** → *Robotics*.

Keywords

Robot-assisted Language Learning, Generative Artificial Intelligence, Storytelling, Conversational Practice

ACM Reference Format:

Eva Verhelst and Tony Belpaeme. 2026. Towards a Usage-Based Pedagogy for Second Language Learning with Robots. In *Companion Proceedings of the 21st ACM/IEEE International Conference on Human-Robot Interaction (HRI Companion '26)*, March 16–19, 2026, Edinburgh, Scotland, UK. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3776734.3794484>

1 Introduction

In recent years, impressive leaps in the performance of generative AI have opened up many opportunities in language learning. Generative AI now enables chatbots to converse fluently in a second language, making language practice accessible from home. This, together with language learning apps such as Duolingo, has made language learning more accessible than ever. However, users often struggle to maintain consistent engagement, limiting their L2 proficiency gains [13]. Meanwhile, robot-assisted language learning (RALL) has sought to address these gaps, traditionally relying on teleoperated robots to compensate for technological limitations [20]. The advent of generative AI might allow for more autonomous robots in research, further bridging the gap in bringing RALL to classrooms. Connecting large language models (LLMs) to social robots could replicate the experience of practising with a fluent speaker. Yet, two key questions arise: Is this approach pedagogically optimal, and does current technology truly support it?

To address these questions, we synthesise theories of language learning with the technological requirements for autonomous robot tutors. We begin by examining pedagogical principles that could inform interaction design, followed by an assessment of the technological enablers and barriers. Finally, we present a proof-of-concept interaction: a shared story-building task. Here, the robot leads the conversation by initiating a story and dynamically adapting to the learner's input, thereby bridging theory and technology.



This work is licensed under a Creative Commons Attribution 4.0 International License.
HRI Companion '26, Edinburgh, Scotland, UK
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2321-6/2026/03
<https://doi.org/10.1145/3776734.3794484>

2 Theories in Language Learning

Learning a second language is far more than memorising vocabulary or mastering basic phrases. It is a multifaceted endeavour that requires understanding and integrating multiple layers of linguistic structure. Successful language learning involves grasping the phonological rules that govern sound patterns, the morphological principles that shape word formation, the syntactic frameworks that organise sentences, and the semantic systems that convey meaning. Beyond these, learners must also navigate pragmatic rules, which dictate how language is used appropriately in social and cultural contexts. Together, these interconnected components form the foundation of communicative competence, making the teaching of second languages a complex challenge, on which many varied pedagogical perspectives exist. Some of these lend themselves well to the implementation in AI-powered technology. To design an effective autonomous robot tutor, we must first understand which language learning theories can be operationalised in a social robot. Below, we explore key theories that inform our approach — focusing on those that align with interactive, context-rich, and learner-centred practice.

2.1 Usage-Based Language Learning

Usage-based theories of second language acquisition assume that language learning is driven primarily by the learner’s active engagement with the target language in meaningful contexts. This relies on active use in context, with very little formal instruction on the full linguistic system of the target language. The learner, instead, is expected to infer the underlying “rules” of the language through interaction and contextualised practice [6]. In educational settings, this often translates into experiential activities — such as visiting a shop to order groceries in the target language — that provide authentic opportunities for language use and pattern recognition.

2.2 Feedback in Language Learning

The “no negative feedback” hypothesis in language learning, also known as the *poverty of the stimulus* argument, finds its origin in the study of language acquisition of young children. It posits that children learn a language without consistent corrections for grammatical errors [3]. Proponents argue that since children produce imperfect language and parents rarely provide explicit negative or corrective feedback, learners must have an innate ability to acquire language based on positive examples alone. This hypothesis has been debated, with some studies suggesting that parental reformulations, or recasts, and other implicit corrections do provide some negative evidence [23]. Still, as a concept in second language learning, it makes some very powerful assumptions which resonate with usage-based language learning. If indeed learners can learn the grammar of a language through exposure to — often imperfect — samples of the target language together with context and weak feedback through use in multimodal context, then this does offer a viable alternative to formal language education.

2.3 Narrative and Learner Agency in Language Learning

Storytelling is one of the oldest and most effective tools for language learning. It provides a contextualised, engaging, and multimodal framework for practising language skills, especially when

supported by illustrations and non-verbal cues. Stories create immersive environments where learners encounter language in meaningful contexts, fostering comprehension, vocabulary retention, and communicative output [16].

Learner agency — the ability to influence or co-create the narrative — further enhances the effectiveness of storytelling. Self-Determination Theory posits that autonomy increases intrinsic motivation [22], and this principle extends to language learning. When learners can choose the direction in which a story develops, they become more engaged and invested in the interaction. This aligns with interactionist and task-based theories, which emphasise the role of meaningful choice in deepening language practice and output [25]. Additionally, narrative transportation frameworks [9] suggest that co-constructing a story — where the learner collaborates with a tutor or peer to build the narrative — strengthens engagement and comprehension [14]. This process not only encourages linguistic creativity but also provides opportunities for implicit feedback and vocabulary reinforcement through repeated exposure in a structured yet flexible context.

2.4 Multiple Representations of Learning Content

There is ample theoretical and empirical support for using illustrations to support vocabulary learning. From Paivio’s Dual Coding Theory [5] to Mayer’s Cognitive Theory of Multimedia Learning [18], all argue that there are two or more representational systems (i.e., verbal and imagery), which suggests that word-picture or context-rich imagery boosts memory and learning. Many studies empirically confirm that visual-rich input leads to better vocabulary learning than text-only conditions (e.g., Feng & Webb, 2020 [7]).

2.5 Adaptation of Difficulty

It has long been known that in any learning, ensuring a good match between the difficulty of the content to the skills of the student is crucial to ensure effective education. This concept is most well-known as Vygotsky’s *zone of proximal development*: where the task is just too difficult for the student to solve on their own, but it is achievable with some help [4]. This too has been proven in robot-assisted language learning, where it was found that matching the robot’s language level to that of the student increases learning gains and can increase engagement [20, 28].

3 Social Robots for Language Learning

The previous section gave an overview of key concepts in the theoretical basis of language learning that are relevant to robot-assisted language learning. It paints a picture of an ideal tutoring robot, that understands the learner perfectly, engages them in interesting conversation and teaches a language along the way. Naturally, this entails substantial technical challenges, some of which have been addressed, while others remain open. In this section, we first motivate the use of social robots as tutors. We then discuss the required underlying technologies, highlighting both their current limitations and how we address them, as well as their strengths and how we aim to leverage them.

3.1 Learning with Social Robots

While computer-based apps can offer high-quality written or spoken language interaction, they lack the embodied social presence that underpins natural human communication. Language learning is inherently social, situated, and multimodal and social robots uniquely afford these channels. The robot's physical and social embodiment evokes stronger engagement, motivation, and linguistic alignment than screen-based agents, even when learners interact with a combination of a screen and a robot [12, 15]. Meta-analyses confirm that learners interacting with robots show greater retention and willingness to engage with learning compared to interventions using tablets [2].

3.2 ASR for Feedback

Formal education focuses heavily on providing feedback on learners' output. To implement feedback in a conversational robot, we need to know exactly what the student has said. Unfortunately, automatic speech recognition (ASR) systems, though showing promising results [29], often fail in language learning contexts, where speakers are non-fluent or have a non-native accent. For young children, studies have found that only up to 60% of speech is correctly recognised [11], while different accents of the same language (both native and non-native) have been shown to generate very different performances [8] and languages other than English lead to suboptimal results [19].

Therefore, in a learning interaction with a social robot, we cannot expect perfect speech transcriptions. As ASR introduces errors into the transcriptions, providing feedback using this input would be counterproductive. Luckily, language theories support avoiding explicit negative feedback, so we can focus on delivering implicit feedback, negating the need for exact transcriptions.

3.3 ASR for Meaningful Conversation

As discussed in the previous section, ASR for language learners is suboptimal. To ensure a successful, engaging conversation, it is important for the learner to feel understood. We argue that combining ASR with LLMs in a conversation that is well-grounded in context might solve that problem, leading to a robust conversation, even with suboptimal speech recognition. LLMs are known for having an impressive semantic understanding of language, and have recently been shown to exhibit what seems to be pragmatic understanding [24]. People rely heavily on pragmatics and context to understand speech, even when speech is difficult to understand in noisy environments [30] — just as suboptimal ASR performance. Therefore, we choose to design the robot's interaction so that even if only some of the learner's speech is correctly transcribed, LLMs can use the context of the conversation to fill in the gaps. The design of the interaction facilitates this, as the robot takes the lead in storytelling and asks open questions embedded in the story's context. We have shown the effectiveness of this technique within this interaction in previous work [26].

3.4 LLMs for Dynamic Content Generation

Since the advent of LLMs, they allow us to dynamically generate text about any topic. We leverage this flexibility in different ways to meet the pedagogical requirements discussed before.

While combining a robot with an LLM to have conversations in any language is now easily achievable, that might not be the pedagogically best way to learn that language. What we can do is build a structured interaction guided by pedagogical theories, in which we want to teach a certain language concept, and dynamically generate the content that is needed for this interaction, making sure it contains the concept we are teaching. Section 4 details how we have done this in an example application.

Additionally, we allow the user to meaningfully influence both the topic and the narrative progression, matching the contents to the learner's interests. Given the topic that the learner chooses for the story, LLMs allow us to generate a story fitting to this topic. Throughout the story, the learner's spoken answers to the questions are transcribed as best as possible and used to steer the narrative any way the student likes.

Finally, as learners can be expected to form ungrammatical sentences at times, we prompt the LLM to take the input as transcribed by the ASR and incorporate it as closely as possible into its answer. This not only adapts the content of the story to the user's output, but also provides an example of how the learner's utterance can be used in a grammatical sentence, whether or not the learner made any mistakes. This way, explicit negative feedback is avoided, following the *poverty of the stimulus* argument, while a corrective response in the form of a recast is used.

3.5 Language Level-Aligned LLMs

Matching the language in a lesson to the level of the student greatly improves learning, so ensuring the generated robot speech fits the student will lead to greater learning gains. This can be done by prompting the model to generate language of a certain style, providing it with details about the learner, such as their age, proficiency and native language. Others have explored achieving this with combinations of supervised finetuning and reinforcement learning [17], aligning generated text with the CEFR levels. This allows for easier incorporation into a lesson plan and ensures effective learning. Independent of which method is used, the LLM-generated language can be matched to that of the student, either by measuring the learner's fluency along the way or by starting from an existing estimation of the learner's skill level.

3.6 Image Generation for Multiple Representations

Improvements in image generation models allow us to use them in language learning. Going from a turn in a conversation — a piece of a story in our example — to an image is easily obtained with image generation models that can be prompted, especially when using an LLM in between to come up with a fitting description of that image. This allows us to show a learner multiple representations of the content of the conversation, improving the accessibility of learning. Even when a learner's language skill might be lacking to fully grasp the meaning based on the foreign language alone, images present another opportunity to the learner to reach understanding. Combining generated images with text can teach a language without relying on translations, definitions or explanations of words, letting users learn from context provided in multiple means.

4 Shared Story-Building System

To illustrate how language learning theories can be used to strengthen the pedagogical validity of a language learning interaction with a social robot, we propose our own implementation. We implemented this as a shared story-building activity with a Furhat robot. The robot initiates a story in the target language (L2) on a topic freely chosen by the learner. Then, the story unfolds jointly: the robot poses open-ended questions about the narrative (e.g., “Where does she go next?”, “What should she do with this treasure?”). The learner’s responses are incorporated into the following story segments. The robot’s narration is powered by an LLM, which is prompted to tell part of a story accompanied by a question. Additionally, during every part of the story, the LLM generates a description of the scene of the current part of the story. This description is given to a text-to-image model. The resulting image is shown to the student while the robot is telling the accompanying story part. For better learner comprehension of the story, captions are shown on the same screen as the images, synchronised with the robot’s speech.

The automatic speech recognition and text-to-speech are the Furhat robot’s default, relying on Microsoft Azure. The images are generated using a Stable Diffusion model running on a local server [21]. GPT-4o is used as LLM, accessed through its API.

4.1 Story Generation

The LLM is guided by a system prompt that defines the context: an interactive storytelling exercise in which the robot acts as a language tutor for language learners. The model is instructed to pose a question every few sentences and to add an image description to their response. This description, not shown to the learner, is used to generate the accompanying image. The model is instructed that the questions should influence the story.

For the first turn, the user prompt includes the topic chosen by the user, how many turns are left in the story and which vocabulary words should be used (see Section 4.3 for more details).

For the remaining turns, the user message includes the recognised user speech, with the instruction that it could be wrong and incomplete (due to sub-par speech recognition). The model is instructed to continue the story while incorporating the user’s answer, by both using it as a recast in the answer and by having it steer the narrative of the story. The vocabulary that should be used is added to the prompt. The model is again instructed to add a description of an image and ask a question at the end of the turn.

After each exchange, the robot acknowledges the learner’s response. This serves both to confirm that the contribution was understood and to increase the feeling of interaction. The robot speech is captioned on the same screen as the images, optimising for understanding by exposing the learner to the text in multiple ways.

4.2 Image Generation

Throughout the story, images depicting scenes of the story are shown on screen. This serves to contextualise and ground the story visually, improving language understanding and increasing engagement. These images are generated based on a description generated by the LLM. To ensure proper parsing and consistent generation of these descriptions, structured outputs are used. This is a feature of the OpenAI API (and similar LLMs), where the output adheres to a schema that can be easily defined. Here, this schema consists

of three strings: the story without a question, the question and the image description. This image description is then sent to a Stable Diffusion model [21] running on local servers at our institute. To ensure a consistent style between images and to avoid overly realistic images, the description is appended with a style descriptor (“from a children’s illustrated book”). Images are added to the story as an extra representation of the learning content, ensuring that next to the text with captions (both verbal), there is also a representation in the form of imagery.

4.3 Vocabulary Insertion

To align learner interactions with the curriculum, we illustrate a method for embedding lesson content into the narrative, focusing on vocabulary words. To increase the chance of the words being picked up by the learners, we insert them into the story multiple times each. To ensure that all words are used the right amount of times and that they are spread throughout the story, we maintain an updated list of which words should still be used and how often. This list is then added to the prompt, as well as the amount of words that should be used in this turn. As the words used in a story often appear in a conjugated or inflected form, we used lemmatisation from the spaCy [10] library to compare the lemmatised words in the story against the lemmas of the words in the vocabulary list.

5 Conclusion

We explored the intersection of pedagogical theory and technology in robot-assisted language learning. By synthesising insights from usage-based language learning, feedback mechanisms, learner agency, and adaptive difficulty, we have identified key principles that guide the design of autonomous robot tutors. The shared story-building system presented here demonstrates how these principles can be operationalised in practice, leveraging the strengths of large language models, social robots, and dynamic content generation to create engaging, context-rich language learning experiences.

Our proof-of-concept interaction, where a social robot leads a collaborative story-building task, shows how technology can bridge theory and practice. By dynamically adapting to learner input, integrating multimodal representations, and aligning language complexity with the learner’s proficiency, the system addresses core challenges in language acquisition. Importantly, it avoids over-reliance on explicit feedback, instead using implicit corrections and contextual scaffolding to support learning, in line with the *poverty of the stimulus* argument and usage-based theories. While technological limitations remain, such as the accuracy of automatic speech recognition for non-native speakers, we design our proof-of-concept to leverage the strengths of related technology, such as LLMs’ context awareness for masking poorly performing ASR.

With this work, we hope to inform on the use of generative AI in social robots in a pedagogically sound way. The proposed system is theorised to lead to more engaged learners and — in the long term — increased learning gains. Future work includes further development of the proposed system and testing it — first in a lab environment and later on in a real-world education setting.

Acknowledgments

This research is funded by the Research Foundation Flanders (FWO Vlaanderen, 1S50425N) and the Flanders AI Research 2 initiative.

References

- [1] Saya Amioka, Ruben Janssens, Pieter Wolfert, Qiaoqiao Ren, Maria Jose Pinto Bernal, and Tony Belpaeme. 2023. Limitations of audiovisual speech on robots for second language pronunciation learning. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 359–367.
- [2] Tony Belpaeme, James Kennedy, Arathi Ramachandran, Brian Scassellati, and Fumihide Tanaka. 2018. Social robots for education: A review. *Sci Robot* 3, 21 (15 8 2018), eaat5954. doi:10.1126/scirobotics.aat5954
- [3] R. Brown and C. Hanlon. 1970. Derivational complexity and order of acquisition in child speech. In *Cognition and the development of language*, J. R. Hayes (Ed.). Wiley, New York, 11–53.
- [4] Seth Chaiklin et al. 2003. The zone of proximal development in Vygotsky's analysis of learning and instruction. *Vygotsky's educational theory in cultural context* 1, 2 (2003), 39–64.
- [5] James M Clark and Allan Paivio. 1991. Dual coding theory and education. *Educational psychology review* 3, 3 (1991), 149–210.
- [6] Nick C Ellis and Stefanie Wulff. 2020. Usage-based approaches to L2 acquisition. In *Theories in second language acquisition*. Routledge, 63–82.
- [7] Yanxue Feng and Stuart Webb. 2020. Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition* 42, 3 (2020), 499–523.
- [8] Calbert Graham and Nathan Roll. 2024. Evaluating OpenAI's Whisper ASR: Performance analysis across diverse accents and speaker traits. *JASA Express Letters* 4, 2 (2024).
- [9] Melanie C Green. 2021. Transportation into narrative worlds. In *Entertainment-education behind the scenes: Case studies for theory and practice*. Springer International Publishing Cham, 87–101.
- [10] Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python. (2020). doi:10.5281/zenodo.1212303
- [11] Ruben Janssens, Eva Verhelst, Giulio Antonio Abbo, Qiaoqiao Ren, Maria Jose Pinto Bernal, and Tony Belpaeme. 2024. Child speech recognition in human-robot interaction: Problem solved?. In *International Conference on Social Robotics*. Springer, 476–486.
- [12] James Kennedy, Paul Baxter, and Tony Belpaeme. 2015. The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction*. 67–74.
- [13] Matt Kessler, Shawn Loewen, and Talip Gönülal. 2025. Mobile-assisted language learning with Babbel and Duolingo: comparing L2 learning gains and user experience. *Computer Assisted Language Learning* 38, 4 (2025), 690–714.
- [14] Jacqueline Kory and Cynthia Breazeal. 2014. Storytelling with robots: Learning companions for preschool children's language development. In *The 23rd IEEE international symposium on robot and human interactive communication*. IEEE, 643–648.
- [15] Daniel Leyzberg, Samuel Spaulding, Mariya Toneva, and Brian Scassellati. 2012. The physical presence of a robot tutor increases cognitive learning gains. In *Proceedings of the annual meeting of the cognitive science society*, Vol. 34.
- [16] Claudio Rezende Lucarevski. 2016. The role of storytelling on language learning: A literature review. *Working Papers of the Linguistics Circle* 26, 1 (2016), 24–44.
- [17] Ali Malik, Stephen Mayhew, Christopher Piech, and Klinton Bicknell. 2024. From tarzan to tolkien: Controlling the language proficiency level of llms for content generation. In *Findings of the Association for Computational Linguistics: ACL 2024*. 15670–15693.
- [18] Richard E Mayer. 2024. The past, present, and future of the cognitive theory of multimedia learning. *Educational Psychology Review* 36, 1 (2024), 8.
- [19] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning*. PMLR, 28492–28518.
- [20] Natasha Randall. 2019. A survey of robot-assisted language learning (RALL). *ACM Transactions on Human-Robot Interaction (THRI)* 9, 1 (2019), 1–36.
- [21] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis With Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10684–10695.
- [22] Richard M Ryan and Edward L Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist* 55, 1 (2000), 68.
- [23] Matthew Saxton. 2000. Negative evidence and negative feedback: Immediate effects on the grammaticality of child speech. *First Language* 20, 60 (2000), 221–252.
- [24] Settaluri Lakshmi Sravanthi, Meet Doshi, Tankala Pavan Kalyan, Rudra Murthy, Pushpak Bhattacharyya, and Raj Dabre. 2024. PUB: A Pragmatics Understanding Benchmark for Assessing LLMs' Pragmatics Capabilities. *arXiv preprint arXiv:2401.07078* (2024).
- [25] Merrill Swain. 2000. The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. *Sociocultural theory and second language learning* 97, 1 (2000), 97–114.
- [26] Eva Verhelst and Tony Belpaeme. 2025. Large Language Models Cover for Speech Recognition Mistakes: Evaluating Conversational AI for Second Language Learners. In *Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction*. 1705–1709.
- [27] Paul Vogt, Rianne Van Den Berghe, Mirjam De Haas, Laura Hoffman, Junko Kanero, Ezgi Mamus, Jean-Marc Montanier, Cansu Oranç, Ora Oudgenoeg-Paz, Daniel Hernández García, et al. 2019. Second language tutoring using social robots: a large-scale study. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Ieee, 497–505.
- [28] Jacqueline Kory Westlund and Cynthia Breazeal. 2015. The interplay of robot language level with children's language learning during storytelling. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction extended abstracts*. 65–66.
- [29] Wayne Xiong, Lingfeng Wu, Fil Alleva, Jasha Droppo, Xuedong Huang, and Andreas Stolcke. 2018. The Microsoft 2017 conversational speech recognition system. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 5934–5938.
- [30] Sheryl R Young. 1990. Use of dialogue, pragmatics and semantics to enhance speech recognition. *Speech Communication* 9, 5-6 (1990), 551–564.

Received 2025-12-08; accepted 2026-01-12