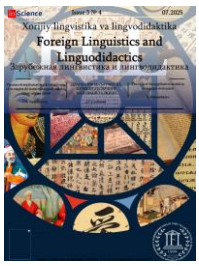
Xorijiy lingvistika va lingvodidaktika – Зарубежная лингвистика и лингводидактика – Foreign Linguistics and Linguodidactics
Journal home page:
https://inscience.uz/index.php/foreign-linguistics



Corpus-based discourse analysis: Automatic Speech Recognition (ASR) technologies and spoken corpus collection

Nargiza ASRorova[1]

Uzbekistan State World Languages University, Университет TIFT

ARTICLE INFO

Article history:
Received May 2025
Received in revised form
15 May 2025
Accepted 25 June 2025
Available online
15 July 2025

Keywords:
Automatic Speech Recognition (ASR),
speech transcription,
corpus,
spoken discourse analysis,
second language acquisition.

ABSTRACT

This study investigates the integration of Automatic Speech Recognition (ASR) technologies in the collection and analysis of spoken learner corpora, with a focus on L2 contexts. Employing a mixed-methods design, the research evaluates the effectiveness of ASR systems, specifically, Whisper and BERT-based models in producing accurate transcriptions and facilitating language acquisition. Quantitative results demonstrate high transcription accuracy, while qualitative data reveal that ASR-supported feedback significantly enhances learner engagement, pronunciation, and speaking proficiency. Technological limitations and ethical concerns related to data privacy and feedback mechanisms are also taken into account. Overall, the findings highlight the transformative potential of ASR technologies in language education by enabling scalable, real-time assessment and personalized feedback, while underscoring the need for continued refinement and equitable implementation across diverse learner populations.

DOI: https://doi.org/10.47689/2181-3701-vol3-iss4-pp1-8

Korpusga asoslangan diskurs tahlili: Nutqni avtomatik aniqlash (ASR) dasturlari va og‘zaki nutq korpusini to‘plash

Kalit so‘zlar:
Avtomatik nutqni aniqlash (ASR),
nutq transkripsiyasi,

ANNOTATSIYA

Ushbu tadqiqot Avtomatik nutqni aniqlash (ASR) texnologiyalarining ingliz tilini ikkinchi til sifatida (L2) o‘rganuvchilarning og‘zaki korpuslarini yig‘ish va tahlil qilish

[1] PhD student, Uzbekistan State World Languages University, Teacher, TIFT University

korpus,
og‘zaki diskurs tahlili,
ikkinchi tilni o‘zlashtirish.

jarayoniga integratsiyasini o‘rganadi. Turli metodlarni aralashtirish usulida olib borilgan mazkur tadqiqot ishida Whisper va BERT asosidagi modellar misolida ASR tizimlarining aniq transkripsiyalar yaratish va ikkinchi til o‘zlashtirish jarayonini oshirishdagi samaradorligi baholandi. Miqdoriy natijalar transkripsiya yuqori aniqlikda ekanligini ko‘rsatdi, tavsifiy jihatdan esa ASR ga asoslangan fikr-mulohazalar (feedback) talabalar faolligining, hamda talaffuz va og‘zaki nutq ko‘nikmalarining sezilarli darajada yaxshilanishiga olib kelishini aniqladi. Texnologiyaning kamchilik tomonlari, ma’lumotlar maxfiyligi hamda fikr-mulohaza mexanizmlari bilan bog‘liq axloqiy masalalar ham tadqiqot doirasida o‘rganib chiqildi. Tadqiqot natijalari ASR texnologiyalarining til ta’limini tubdan o‘zgartirish salohiyatiga ega ekanligini tasdiqladi, ya’ni mazkur texnologiya o‘quvchilarni real vaqtda baholash va individuallashtirilgan fikr-mulohazalar (feedback) taqdim etish imkonini berib, til o‘rganuvchilarning har qanday toifa guruhlari uchun qo‘llanilishi mumkin deb topildi.

Корпусный анализ дискурса: технологии автоматического распознавания речи (ASR) и сбор корпус устной речи

Ключевые слова:
автоматическое распознавание речи (ASR),
транскрипция речи,
корпус,
анализ устного дискурса,
усвоение второго языка.

АННОТАЦИЯ

Данное исследование посвящено интеграции технологий автоматического распознавания речи (ASR) в процессы сбора и анализа корпусов устной речи обучающихся, с особым вниманием к контексту изучения английского языка как второго языка (L2). Используя смешанный метод исследования, автор оценивает эффективность ASR-систем — в частности, моделей Whisper и BERT — в обеспечении точной транскрипции и содействии усвоению иностранного языка. Количественные данные подтверждают высокую точность транскрипций, тогда как качественные результаты свидетельствуют о том, что обратная связь, реализованная с помощью ASR, способствует повышению вовлечённости учащихся, улучшению произношения и развитию навыков устной речи. Также рассматриваются технологические ограничения и этические аспекты, связанные с конфиденциальностью данных и способами предоставления обратной связи. В целом, результаты подчёркивают значительный потенциал технологий ASR в трансформации языкового образования за счёт масштабируемой оценки в реальном времени и персонализированной поддержки обучения, при этом акцентируется необходимость их дальнейшего совершенствования и справедливого внедрения в условиях разнообразия обучающихся.

INTRODUCTION

Automatic Speech Recognition (ASR) Technologies for spoken learner corpus collection are vital tools that significantly impact language acquisition, especially in the context of second language (L2) learning. ASR technologies facilitate the real-time transcription of spoken language, allowing educators and researchers to collect and analyze learner corpora effectively. This capability enhances the assessment of spoken proficiency and offers personalized feedback, thus contributing to improved pronunciation and language skills among learners [1][2].

The evolution of ASR systems has resulted in various approaches tailored for educational settings, including speaker-dependent, speaker-independent, and speaker-adaptive models. Each type presents unique advantages and challenges, particularly in diverse classroom environments where multiple learners interact. Notably, ASR technologies are integrated into applications such as automatic reading tutors and conversational agents, which have been shown to engage learners and foster interactive language practice [3][4]. These tools not only improve language skills but also enhance the accessibility of language learning resources.

Despite their benefits, ASR systems face notable challenges, including accuracy discrepancies across different languages and dialects, as well as ethical concerns surrounding privacy and data usage. Recent studies have highlighted issues such as the potential bias against speakers from diverse linguistic backgrounds, which could impact the fairness of language assessments [5][6]. Furthermore, the technology's performance can be compromised by environmental factors like background noise and the individual characteristics of learners' speech patterns, raising questions about its reliability in varied settings [7].

ASR technologies hold immense potential for revolutionizing language learning through enhanced feedback and engagement, but ongoing research is necessary to address the current limitations and ethical concerns associated with their use. The future of ASR in educational contexts hinges on advancements that improve accuracy, reduce bias, and broaden accessibility, thereby maximizing the benefits for language learners worldwide [5][8].

LITERATURE REVIEW

Automatic Speech Recognition (ASR) technology has evolved significantly since its inception in the mid-20th century, marking a transformative journey in how spoken language is processed and understood by machines. The early attempts at ASR faced challenges with variations in accents, dialects, and speech nuances, which often resulted in inaccuracies in transcription. Over the decades, advancements in acoustic modeling and machine learning have enhanced the robustness and accuracy of ASR systems, allowing for more effective communication between humans and technology [1][9].

In the context of language learning, ASR plays a crucial role in developing spoken learner corpora, which are essential for evaluating and improving speaking proficiency among learners of a second language (L2). These corpora can be categorized into two main types: those focusing on native-speaking children, used primarily for developing virtual tutors in non-language subjects, and those aimed at young language learners to support language acquisition [10][11]. ASR systems provide real-time feedback on learners' pronunciation and fluency, thereby offering personalized practice exercises that are vital for language development [12].

Moreover, the incorporation of ASR in educational settings not only enhances accessibility through automated captions for instructional content but also aids in the efficient assessment of students' speaking abilities. However, despite the impressive progress in ASR technologies, challenges such as the influence of learners' backgrounds, including prior language exposure and educational experiences, must be addressed to ensure a comprehensive understanding of speaking proficiency [3] [13-][14]. Future studies could benefit from including measures to evaluate these factors, potentially through pre-assessment surveys or interviews [3].

As ASR continues to advance, its applications extend beyond language learning into various domains, including healthcare, customer support, and accessibility tools, underscoring its growing significance in our daily lives [9][14]. The ongoing research aims to tackle issues related to privacy, bias, and the computational efficiency of ASR systems, which are crucial for their widespread adoption and effectiveness [9][15].

Automatic Speech Recognition (ASR) technologies have emerged as essential tools for collecting spoken learner corpora, enabling educators and researchers to analyze language acquisition in various contexts. These systems facilitate the assessment of learner language by capturing spoken interactions in real-time and providing transcriptions for further analysis [2].

Types of ASR Systems

ASR systems can be classified into three main types: speaker-dependent, speaker-independent, and speaker-adaptive. Speaker-dependent ASR requires training on an individual's voice, making it less suitable for diverse classroom settings where multiple learners are involved. In contrast, speaker-independent ASR operates without prior training, leveraging extensive speech databases to recognize various speakers' voices, which is particularly beneficial in educational environments [3]. Speaker-adaptive ASR combines aspects of both, using customized databases to improve recognition accuracy based on user input.

Impact on Language Acquisition

ASR technologies are employed in various educational applications aimed at enhancing learner engagement and language development. For example, automatic reading tutors utilize ASR to provide feedback on pronunciation errors by comparing learners' speech with reference transcripts [13]. These systems are generally designed for quiet environments, such as libraries, ensuring that only one learner reads at a time, which facilitates focused feedback.

Conversational tutors represent another innovative use of ASR, combining speech recognition with language generation to create interactive learning experiences. Tools like My Science Tutor (MyST) support small-group learning by engaging learners in real-time dialogues, which enhances their spoken language skills [13][4].

Recent studies indicate that ASR technology significantly enhances learner engagement during language learning activities. By providing instant feedback and interactive practice opportunities, ASR systems contribute to improved pronunciation, vocabulary acquisition, and overall language proficiency [4]. The accessibility of free ASR tools has also democratized language learning, allowing for quick deployment in various educational contexts [3].

Despite their advantages, ASR systems face challenges, particularly in terms of accuracy across different languages and dialects. While modern ASR engines have made

considerable advancements, achieving 100% accuracy remains elusive due to the complexities of human speech [5]. Additionally, privacy concerns regarding data use and storage present significant ethical considerations for educational institutions implementing ASR technologies [6].

METHODOLOGY

This study employed an explanatory sequential mixed methods design, which involved the systematic collection and analysis of both quantitative and qualitative data to comprehensively address the effectiveness of Automatic Speech Recognition (ASR) technologies in collecting spoken learner corpora. The methodology was divided into distinct phases to provide a holistic view of the impact of ASR technologies on language learning outcomes, particularly in the context of English L2 learners [3].

Participants

The research focused on 20 intermediate-level L2 learners from Uzbekistan, differentiating it from previous studies by its emphasis on integrating ASR technology with peer correction techniques during pronunciation and speaking instruction. An experienced IELTS teacher collaborated closely with the researcher to implement the intervention and offer feedback to participants [3]. This structured approach not only ensured consistency across groups but also enriched the data collected through both objective measurements and subjective learner perceptions [3].

Data Collection

The initial quantitative phase assessed the effectiveness of various ASR systems by measuring their precision, recall, and F1 score in detecting keywords in learner speech. A specific focus was placed on evaluating the classification accuracy of a BERT-based model in predicting students' oral assessment grades from ASR transcripts. The results indicated a notable 2% absolute drop in classification accuracy for African American English (AAE) speaking students compared to their non-AAE counterparts [16]. ASR systems, particularly Whisper, achieved a high scoring accuracy of 95.6% in comparison to human-labeled transcripts, which scored at 96.3% [16].

Following the quantitative analysis, the qualitative phase utilized interviews to delve deeper into learners' perceptions of ASR technology. This phase involved a thematic analysis of the interview data, guided by Grbich's (2012) methodologies. Initial reviews identified broad categories that were further dissected into subthemes, capturing nuanced learner experiences and attitudes towards ASR technology's role in their learning process [3].

Procedure

The intervention protocol was meticulously designed to align with the research objectives, and regular meetings between researchers and the instructor ensured a consistent approach to data interpretation and theme development. The study's design allowed for a rigorous examination of how ASR technologies influence pronunciation and speaking skills, filling a significant gap in existing research by highlighting the perspectives of intermediate L2 learners [3][16].

The methodology incorporated comprehensive testing of ASR systems across various configurations, yielding valuable insights into the strengths and weaknesses of each technology in practical classroom settings. Future studies may expand upon this framework by utilizing larger samples and varying procedures to further validate these findings [3].

RESULTS AND DISCUSSION

Quantitative results demonstrate that the ASR systems exhibit varying levels of precision and recall in detecting keywords, with an F1 score reflecting the balance between these metrics. For instance, the BERT classification model for automatic response scoring achieved its highest accuracy at 95.6% when using Whisper ASR transcripts, in contrast to a 96.3% accuracy with human-labeled transcripts [16]. This suggests that while ASR technologies can effectively transcribe learner speech, the quality of the transcription can significantly affect the assessment outcomes. The integrated approach tested on the UASpeech dataset resulted in an overall word recognition accuracy of 69.4% across multiple speakers, indicating that while ASR systems can achieve reasonable performance, there remains room for improvement, particularly in diverse speaking styles [11][16].

The ASR systems provide a variety of feedback mechanisms, which are crucial for language learners. Explicit corrective feedback is enabled through graphical representations of speech, such as waveforms and spectrograms, allowing learners to understand discrepancies between their pronunciation and that of native speakers [17]. This level of detail can foster a more tailored learning experience, accommodate individual learning styles, and address specific pronunciation challenges [3]. However, there are concerns that some ASR systems may inadvertently limit learners' practice opportunities by relying too heavily on pre-recorded samples for comparison, thus hindering spontaneous speech production [3].

Research indicates that ASR programs providing explicit corrective feedback are particularly beneficial for pronunciation development in language learners. For instance, systems like MyET, which deliver immediate, targeted feedback, have been shown to enhance learner outcomes significantly when compared to systems that offer indirect feedback, such as Speechnotes, which often fail to accurately identify the nature and location of pronunciation errors [16]. This distinction emphasizes the importance of selecting ASR technologies that align with the pedagogical goals of language instruction.

Regarding the effectiveness of ASR in language classrooms, the research highlighted that ASR technology can lead to improved pronunciation among English as a Foreign Language (EFL) learners. ASR systems were found to facilitate instant feedback, allowing learners to recognize and rectify errors more efficiently [3]. Furthermore, the integration of ASR in pronunciation training has been shown to increase student engagement and interaction, contributing positively to their overall learning experience [4][3].

Despite the positive outcomes associated with ASR technologies, several challenges have been identified. For example, some students reported frustrations with ASR systems misinterpreting words, which hindered their learning experience [16]. Additionally, the effectiveness of ASR may be compromised in suboptimal audio conditions, leading to increased inaccuracies in speech recognition [13]. These factors underscore the need for continued research and development in ASR technologies to enhance their reliability and effectiveness in educational contexts.

As ASR technologies continue to evolve, future case studies should explore their application across diverse learner demographics and educational settings. Investigating the long-term effects of ASR on language acquisition and exploring the integration of ASR with other technological tools could provide valuable insights into optimizing language

instruction and learner engagement strategies [14][15]. By addressing current limitations and refining system capabilities, ASR has the potential to transform language learning experiences for students worldwide.

CONCLUSION

This study underscores the significant potential of Automatic Speech Recognition (ASR) technologies to enhance language learning outcomes, particularly within the context of English as a Second Language (L2) instruction. Employing a mixed-methods design, the research demonstrates that ASR systems, specifically, Whisper and BERT-based models, are capable of generating highly accurate transcriptions and delivering real-time, individualized feedback that supports improvements in learners' pronunciation, fluency, and overall oral proficiency. The integration of ASR into pedagogical practice not only facilitates scalable and efficient language assessment but also promotes learner autonomy and engagement.

Nevertheless, the findings also reveal persistent challenges, including reduced accuracy in varied linguistic and acoustic conditions, potential algorithmic bias against speakers of certain dialects, and ethical concerns related to data privacy and informed consent. These limitations highlight the need for continued refinement of ASR technologies and the development of equitable, ethically sound implementation strategies in educational contexts. Addressing these dimensions is essential for maximizing the pedagogical benefits of ASR and ensuring its effective and inclusive integration into language learning environments.

REFERENCES:

1. Nakamura, S., Spring, R., & Sakurai, S. (2024). The impact of ASR-based interactive video activities on speaking skills: Japanese EFL learners' perceptions. The Electronic Journal for English as a Second Language, 27(4). https://doi.org/10.55593/ej.27108a5
2. Gladia. (2023, December 19). A review of the best ASR engines and the models powering them in 2024. Gladia Blog. https://www.gladia.io/blog/a-review-of-the-best-asr-engines-and-the-models-powering-them-in-2024
3. Michot, J., Hürlimann, M., Deriu, J., Sauer, L., Mlynchyk, K., & Cieliebak, M. (2024). Error-preserving automatic speech recognition of young English learners' language. Beta archives, https://arxiv.org/html/2406.03235v1
4. Qian, Z., Xiao, K., & Yu, C. (2023). A survey of technologies for automatic dysarthric speech recognition. Journal of Audio, Speech, and Music Processing, 2023(48). https://doi.org/10.1186/s13636-023-00318-2
5. Bhatnagar, N. (2024 Deep dive into ASR systems. Medium. https://medium.com/@captnitinbhatnagar/deep-dive-into-asr-systems-c571a576ff26
6. Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: A mixed methods investigation. Frontiers in Psychology, 14, Article 1210187. https://doi.org/10.3389/fpsyg.2023.1210187
7. Southwell, R., Pugh, S., Perkoff, E. M., Clevenger, C., Bush, J., Lieber, R., Ward, W., Foltz, P., & D'Mello, S. (2022). Challenges and feasibility of automatic speech recognition for modeling student collaborative discourse in classrooms. In Proceedings of the 15th International Conference on Educational Data Mining (EDM 2022) (pp. 342–

353). https://educationaldatamining.org/edm2022/proceedings/2022.EDM-long-papers.26/index.html

8. What is ASR? (2023) How does it work? Our in-depth 2023 guide. SimonTech. https://simontech.com/asr-guide-2023

9. Sciforce. (2021, July 7). Automatic speech recognition (ASR) systems compared: Ultimate guide to Google, Azure, IBM, Amazon, SpeechMatics, Kaldi, and HTK ASR systems. Medium. https://medium.com/@sciforce/automatic-speech-recognition-asr-systems-compared

10. Abe, M., & Kondo, Y. (2019). Constructing a longitudinal learner corpus to track L2 spoken English. Journal of Modern Languages, 29(1), 21–42. https://doi.org/10.22452/jml.vol29no1.2

11. Xiao, Y. (2025). The impact of AI-driven speech recognition on EFL listening comprehension, flow experience, and anxiety: A randomized controlled trial. Humanities and Social Sciences Communications, 12, Article 425. https://doi.org/10.1057/s41599-025-02626-3

12. Torres, G. (2025, May 6). What is automatic speech recognition? An overview of ASR technology. Voice Flow. https://www.voiceflow.com/blog/automatic-speech-recognition

13. Omniscien Technologies. (2025). Automatic speech recognition accuracy and customization. Retrieved [July 20, 2025], from https://omniscien.com/lsev6/features/asr/automatic-speech-recognition-accuracy-and-customization/

14. Johnson, A., Chance, C., Stiemke, K., Veeramani, H., Shankar, N. B., & Alwan, A. (2023). An analysis of large language models for African American English speaking children's oral language assessment. Black Excellence in Engineering, Science & Technology Manuscripts, 1. https://doi.org/10.36227/beestm.vol1.art1

15. Ngo, T. T.-N., Chen, H. H.-J., & Lai, K. K.-W. (2023). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. ReCALL, 35(2), 169–185. Cambridge University Press. https://doi.org/10.1017/S0958344023000070

16. Arriaga, C., Pozo, A., Conde, J., & Alonso, A. (2024, September 9). Evaluation of real-time transcriptions using end-to-end ASR models [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2409.05674