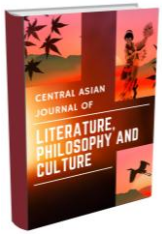

CENTRAL ASIAN JOURNAL OF LITERATURE, PHILOSOPHY, AND CULTURE

https://cajlpc.casjournal.org/index.php/CAJLPC

Volume: 07 Issue: 03 | July 2026 ISSN: 2660-6828

Article

Methods for Punctuation Restoration in Automatic Speech Recognition Systems

Sobirova Zarnigor Ganijon kizi[*1]

1. Computational Linguistics and Digital Technologies, Tashkent State University of Uzbek Language and Literature

*Correspondance: sobirovazarnigor1996@gmail.com

Citation: Ganijon kizi, S. Z. Methods for Punctuation Restoration in Automatic Speech Recognition Systems. Central Asian Journal of Literature, Philosophy, and Culture 2026, 7(3), 98-102.

Received: 20th Mar 2026
Revised: 05th Apr 2026
Accepted: 20th Apr 2026
Published: 17th May 2026

Abstract: Punctuation restoration constitutes one of the most significant Natural Language Processing (NLP) tasks within Automatic Speech Recognition (ASR) systems. Automatic Speech Recognition refers to a computational technology that enables machines to recognize, process, and transcribe human speech into textual form. ASR systems serve as a fundamental component in numerous NLP applications, including intelligent voice assistants, automated call-center systems, speech analytics, and machine translation technologies. With the exponential growth of digital content and spoken communication platforms, ASR technologies have become an indispensable element of modern intelligent information systems and services.

Keywords: Punctuation Restoration, Automatic Speech Recognition, ASR, Speech Recognition, Speech Transcripts, Uzbpunct Dataset

1. Introduction

In the contemporary digital era, the accuracy, clarity, and linguistic quality of written communication have become critically important. Whether in informal online interactions or professional documentation, the effectiveness of communication largely depends on the precision and coherence of language usage. Traditional spelling and grammar checking systems have long been utilized as practical tools for identifying linguistic errors [1, 2]; however, such systems are generally limited in terms of contextual understanding and semantic interpretation. These limitations have accelerated the development of more advanced approaches based on Natural Language Processing (NLP) technologies. This study investigates the development of modern punctuation and grammar restoration approaches utilizing NLP methods and emphasizes their advantages over conventional rule-based systems in improving user interaction and textual readability in digital communication environments [3, 4].

2. Materials and Methods

The increasing importance of digital communication has intensified the demand for precise and comprehensible textual information. From online messaging platforms to formal written correspondence, the successful transmission of information is directly associated with the linguistic correctness of textual content. Although conventional spelling and grammar correction approaches have proven effective for detecting surface-level errors, they remain insufficient in handling contextual ambiguity and adaptive

language processing. Consequently, NLP-based approaches have emerged as a promising research direction capable of addressing complex language-processing tasks more efficiently and accurately.Natural Language Processing is an interdisciplinary scientific field that integrates linguistics, computer science, artificial intelligence, and machine learning methodologies in order to enable computational systems to analyze, interpret, and generate human language. By leveraging NLP techniques, modern text correction systems can provide context-aware grammatical and orthographic analysis, thereby improving the precision of error detection and correction mechanisms. NLP-based systems analyze lexical, syntactic, semantic, and contextual information to generate more reliable and linguistically appropriate correction suggestions. In recent years, voice-based intelligent assistants such as Google Assistant, Amazon Alexa, Siri, and Cortana have become increasingly widespread. The core technological foundation of these systems is Automatic Speech Recognition (ASR), which represents one of the most important tasks in NLP. ASR technologies transform audio signals into textual representations; therefore, such approaches are commonly referred to as Speech-to-Text technologies. Modern intelligent systems not only convert speech into text but also perform semantic interpretation, contextual analysis, and intent recognition. Consequently, these systems are capable of generating appropriate responses and executing user commands automatically.

3. Results and Discussion

Human speech represents one of the primary means of communication in both personal and professional domains, and speech-to-text functionality possesses extensive practical applicability. Such NLP applications can be employed for customer support systems, transcription of business negotiations and meetings, conversational agents, voice-driven chatbots, and automated documentation systems. Audio information consists of acoustic signals, environmental noise, and speech components [5]. Human speech represents a particularly complex form of audio information because it encodes natural language structures and semantic meaning. For computational processing, audio signals are digitized and transformed into spectrogram representations. However, spoken language processing remains substantially more challenging than general audio classification tasks due to the inherent linguistic structure embedded within speech signals. Audio classification tasks generally aim to determine the corresponding category of an audio signal from a predefined set of classes [6, 7].

Speech transcripts generated by Automatic Speech Recognition systems typically do not contain punctuation marks or capitalization information. The absence of punctuation significantly reduces the readability and comprehensibility of automatically generated speech transcripts. Therefore, punctuation restoration (PR) and capitalization restoration (CR), which are regarded as important NLP tasks, aim to improve the readability and structural integrity of punctuation-free ASR-generated texts [8].

The successful resolution of PR and CR tasks can significantly improve the performance of downstream NLP applications, including Named Entity Recognition (NER, Part-of-Speech (POS) tagging, semantic parsing, and spoken dialogue segmentation. However, the evaluation of punctuation restoration systems based on spoken language transcripts remains a highly challenging problem because punctuation rules themselves may often be ambiguous even in manually written texts, while naturally occurring spoken discourse makes sentence and phrase boundary identification substantially more difficult [9].

Automatic Speech Recognition systems generally produce continuous sequences of words without punctuation marks. Such outputs considerably reduce both human readability and the effectiveness of downstream Natural Language Processing tasks performed on ASR-generated text [10].

K. Makhija and T. N. Chng proposed an architecture for punctuation prediction utilizing the pre-trained BERT model. Their proposed approach substantially improved performance on the IWSLT dataset and achieved an overall F1-score of 81.4% for the joint prediction of periods, commas, and question marks.

K. Xu, L. Xie, and K. Yao introduced a punctuation prediction method based on the Long Short-Term Memory (LSTM) neural network architecture. Their proposed model employed multiple neural layers to capture dependencies between input features and output labels. Furthermore, they empirically investigated the influence of modeling dependencies among output labels. Experimental findings demonstrated that deep bidirectional LSTM architectures achieved state-of-the-art performance in punctuation prediction tasks [11].

X. Wu, S. Zhu, Y. Wu, and K. Yu (2016) implemented punctuation prediction on large-scale language corpora using a multi-view LSTM architecture. Experimental results revealed that LSTM-based approaches significantly outperformed traditional Conditional Random Field (CRF)-based models.

Since the Chinese writing system substantially differs from alphabetic languages, X. Liu, Y. Liu, and X. Song [12] proposed a three-stage LSTM-based framework for punctuation prediction in Chinese corpora. Their method integrated lexical features, pause duration information, and pitch-related acoustic features. The proposed multimodal approach demonstrated superior performance and effectiveness on benchmark datasets through the utilization of lexical, temporal, and prosodic information.

ASR systems can be effectively applied in numerous practical domains. Although these technologies significantly improve the efficiency and usability of existing intelligent systems, the generated outputs generally consist only of unpunctuated word sequences, which may introduce ambiguity in user intent interpretation. A. Silva, B. J. Theobald, and N. Apostoloff [13] applied ASR technologies within consumer electronics systems. By employing a transformer-based punctuation prediction framework, they improved the IWSLT 2012 TED benchmark performance by approximately 8%.

Most contemporary punctuation prediction approaches are primarily trained on clean and manually prepared datasets, which limits their generalization capability in real-world ASR environments containing transcription errors and acoustic noise. To address the discrepancy between clean training data and noisy testing environments, A. Zheng, N. Ye, X. Wang, and X. Song [14] proposed three random (3R) data augmentation strategies: random word deletion (RWD), random word substitution (RWS), and random phoneme edition (RPE). Additionally, they introduced a phoneme-level similarity lexicon for replacing acoustically similar words. Their approach also utilized the large-scale RoBERTa model to capture semantic representations and long-range contextual dependencies for punctuation prediction.

M. Fang, H. Zhao, X. Song, and X. Wang [15] proposed a punctuation prediction approach for Chinese text by combining BLSTM and BERT architectures. In this framework, BERT was utilized as a contextual text encoding layer for semantic representation learning. Compared with previous Recurrent Neural Network (RNN)-based punctuation prediction methods, the proposed model demonstrated superior performance in capturing semantic information and long-distance contextual dependencies in unsegmented Chinese texts. Experimental results obtained on Chinese news datasets indicated that the BERT-BLSTM-based model outperformed baseline approaches by 31.07% in terms of overall micro-F1 score. At present, scientific and practical research dedicated to punctuation restoration and punctuation correctness detection in Uzbek-language texts remains extremely limited.

4. Conclusion

Traditional spelling and grammar checking approaches have long served as valuable tools for identifying and correcting linguistic errors. In recent years, a substantial body of scientific research has focused on the application of Artificial Intelligence (AI) techniques for the automatic detection and correction of spelling and grammatical mistakes in textual data. In particular, NLP-based correction systems analyze contextual, syntactic, and semantic information in order to identify linguistic inaccuracies and generate more precise correction suggestions. Currently, voice assistants such as Google Assistant, Amazon Alexa, Siri, and Cortana are fundamentally built upon the Natural Language Processing task known as Automatic Speech Recognition (ASR). These NLP applications transform audio-based information into textual representations. However, speech transcripts generated by Automatic Speech Recognition systems generally lack punctuation marks and capitalization information. The absence of punctuation in automatically recognized speech fragments significantly reduces textual readability and negatively affects the comprehension and semantic interpretation of the generated text. Consequently, punctuation restoration has become one of the essential research directions in modern NLP and speech processing systems. The integration of advanced deep learning and transformer-based architectures into ASR pipelines can substantially improve the linguistic quality, readability, and semantic coherence of automatically generated speech transcripts.

REFERENCES

[1] J. Yi, J. Tao, Y. Bai, Z. Tian, and C. Fan, "Adversarial transfer learning for punctuation restoration," arXiv preprint arXiv:2004.00248, 2020.

[2] T. B. Nguyen, Q. M. Nguyen, T. T. H. Nguyen, Q. T. Do, and C. M. Luong, "Improving Vietnamese named entity recognition from speech using word capitalization and punctuation recovery models," arXiv preprint arXiv:2010.00198, 2020.

[3] P. Hlubík, M. Španěl, M. Boháč, and L. Weingartová, "Inserting punctuation to ASR output in a real-time production environment," in *International Conference on Text, Speech, and Dialogue*, Cham, Switzerland: Springer International Publishing, 2020, pp. 418–425.

[4] K. Sirts and K. Peekman, "Evaluating sentence segmentation and word tokenization systems on Estonian web texts," in *Human Language Technologies – The Baltic Perspective*, IOS Press, 2020, pp. 174–181.

[5] X. Wang, "Analysis of sentence boundary of the host's spoken language based on semantic orientation pointwise mutual information algorithm," in *2020 12th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, 2020, pp. 501–506.

[6] K. Makhija, T. N. Ho, and E. S. Chng, "Transfer learning for punctuation prediction," in *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2019, pp. 268–273.

[7] K. Xu, L. Xie, and K. Yao, "Investigating LSTM for punctuation prediction," in *2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, 2016, pp. 1–5.

[8] X. Wu, S. Zhu, Y. Wu, and K. Yu, "Rich punctuations prediction using large-scale deep learning," in *2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, 2016, pp. 1–5.

[9] X. Liu, Y. Liu, and X. Song, "Investigating punctuation prediction in Chinese speech transcriptions," in *2018 International Conference on Asian Language Processing (IALP)*, 2018, pp. 74–78.

[10] A. Silva, B. J. Theobald, and N. Apostoloff, "Multimodal punctuation prediction with contextual dropout," in *ICASSP 2021 – IEEE International Conference on Acoustics, Speech and Signal Processing*, 2021, pp. 3980–3984.

[11] A. Zheng, N. Ye, X. Wang, and X. Song, "3R: Word and phoneme edition-based data augmentation for lexical punctuation prediction," in *2020 16th International Conference on Computational Intelligence and Security (CIS)*, 2020, pp. 1–5.

[12] M. Fang, H. Zhao, X. Song, X. Wang, and S. Huang, "Using bidirectional LSTM with BERT for Chinese punctuation prediction," in *2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP)*, 2019, pp. 1–5.

[13] R. S. Madatovich and S. D. Maxamadiyevna, "The role of the system of education and family education in forming youth's world view," *European Journal of Humanities and Educational Advancements*, vol. 4, no. 4, pp. 128–130.

[14] R. Madatovich, "The role of civic responsibility in educating youth in a healthy spiritual environment in an information society," *Pubmedia Social Sciences and Humanities*, vol. 3, no. 1, p. 6, 2025.

[15] R. S. Madatovich, "The role of preschool education and family education in the raising of a healthy balanced generation," *For Teachers*, vol. 57, no. 4, pp. 520–523, 2024.