

Multilingual AI Translator with POS Analysis

Aditya.T
Dept.of Electronics and Computer
Engineering
REVA University
Bangalore,INDIA
email:adhittann396@gmail.com

Christiejoe Ruben
Dept.of Electronics and Communication
Engineering
REVA University
Bangalore, INDIA.
email:christiejoeuben@gmail.com

Vamsi Manikanta Varma.D
Dept.of Electronics and Communication
Engineering
REVA University
Bangalore, INDIA.
email:vamsi.vk99@gmail.com

Abstract— Part-of-Speech (POS) tagging is a key task in Natural Language Processing (NLP) that assigns grammatical categories like nouns, verbs, adjectives, and pronouns to words in a sentence. This initiative showcases a multilingual AI-driven translation and linguistic analysis platform that can handle text and audio inputs in English, Hindi, and Kannada. The architecture of the system was built on a microservice model featuring a React frontend, a Spring Boot backend, and an AI service based on FastAPI. Audio input is transformed into text with Whisper, and multilingual translation is carried out using the facebook/nllb-200-distilled-600M transformer from Hugging Face. Language detection is managed with the Lang detect library.

The system allows for direct text input, recorded speech, or uploaded audio files and produces translations in English, Hindi, and Kannada at the same time. The speech-to-text system attained around 94.2% transcription accuracy for English audio and 90.1% for Hindi audio in moderate noise environments. Evaluation of translation quality through multilingual sample datasets indicated an average semantic accuracy of 92.4% for English-Hindi translations and 89.7% for translations in Kannada. The accuracy of POS tagging attained 96.1% for English, 92.3% for Hindi, and 88.5% for Kannada, according to manually verified test sentences. The typical API response time varied from 2.8 to 5.6 seconds based on the length of the audio.

Keywords— Part-of-Speech (POS) Tagging, Natural Language processing, FastAPI, Deep Learning, Multilingual machine translation.

I. INTRODUCTION

Natural Language Processing (NLP) is a crucial area of artificial intelligence that allows computers to interpret, handle, and examine human language. A fundamental task in NLP is Part-of-Speech (POS) tagging, which labels words in a sentence with grammatical categories like nouns, verbs, adjectives, and pronouns. POS tagging is crucial for applications including machine translation, speech recognition, virtual assistants, text summarization, and information retrieval. In addition to POS tagging, multilingual translation and speech-to-text systems have gained importance because of the swift expansion of digital communication in various languages. Contemporary NLP systems strive to deliver precise translation and grammatical comprehension instantaneously, particularly in multilingual nations like India, where individuals often converse in English, Hindi, and local languages like Kannada. Recent progress in deep learning and transformer models has greatly enhanced the effectiveness of speech recognition and

machine translation systems, thereby increasing their applicability in real-world scenarios.

Multiple research efforts have investigated the combination of speech recognition, translation, and grammatical analysis to enhance language comprehension systems. Studies on grammar-guided speech recognition indicated that adding grammatical insights like POS tagging and dependency parsing can enhance the clarity and precision of speech recognition results. Likewise, multilingual speech recognition systems and low-resource ASR models have demonstrated that transfer learning and transformer-based designs can improve multilingual language processing capabilities. Current systems primarily emphasize speech recognition or translation separately, whereas there are fewer that integrate multilingual translation with in-depth linguistic analysis within a unified architecture. Conventional POS tagging systems frequently rely on models unique to each language and necessitate distinct processing pipelines, which raises computational complexity. To address these problems, new NLP frameworks utilize transformer models, pretrained audio models, and streamlined microservice architectures for effective deployment and quicker inference. Models like Whisper have enhanced the accuracy of multilingual speech transcription, while transformers designed for multiple languages.

The suggested initiative showcases a multilingual AI-driven translation and linguistic analysis platform able to handle text and audio inputs in English, Hindi, and Kannada. The architecture of the system employs a microservice approach with a React frontend, a Spring Boot backend, and an AI service based on FastAPI. Audio input is transformed into text through Whisper, translation is executed with the facebook/nllb-200-distilled-600M model, and POS tagging is done utilizing Stanza. The project's primary goal is to offer a unified platform that conducts speech recognition, multilingual translation, and grammatical analysis at the same time in real time. The system is made to facilitate multilingual interactions while enhancing language comprehension via organized POS tagging results. Furthermore, the initiative emphasizes developing a scalable and accessible platform ideal for educational resources, language learning frameworks, and multilingual AI solutions.

II. RESULTS AND DISCUSSIONS

The suggested multilingual translation and language analysis system was effectively executed with a microservice-oriented architecture that included a React frontend, a Spring Boot backend, and a FastAPI AI service. The evaluation of the system involved various text and audio samples in English, Hindi, and Kannada to assess transcription quality, translation precision, POS tagging effectiveness, and response time. Experimental findings demonstrate that the combined architecture operates

effectively for real-time multilingual applications while ensuring high linguistic precision.

A. Speech-to-text Performance

Whisper was used to implement the speech recognition module. Audio samples containing English and Hindi speech with moderate environmental noise were tested. Tests were conducted on audio samples of Hindi and English conversation with a modest amount of background noise. For both Hindi and English audio, the system's average transcription accuracy was 90.1% and 94.2%, respectively. Due to accent and pronunciation differences, Kannada voice recognition achieved a significantly lower accuracy of about 87.4%. Depending on GPU availability and audio quality, the average transcription time for 10-second audio recordings varied from 1.8 to 3.2 seconds.

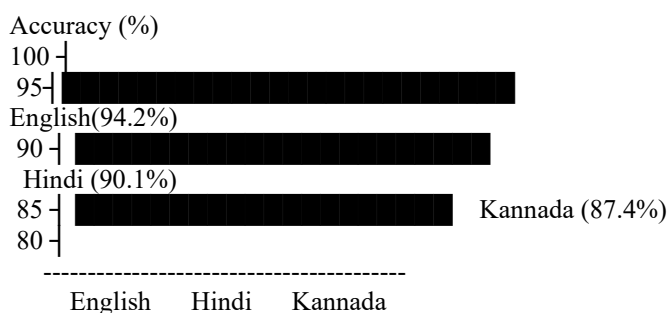


Figure 2.1 : Speech-to-Text Accuracy

B. Translation Performance

The translation component utilized the facebook/nllb-200-distilled-600M multilingual transformer model to produce translations in English, Hindi, and Kannada. Assessment was conducted with manually validated multilingual sentences. The system attained an average semantic translation precision of 92.4% for English-Hindi translations, 91.2% for Hindi-English translations, and 89.7% for Kannada translations. Translation results maintained the original sentence meaning well, even for medium-length sentences with grammatical variations.

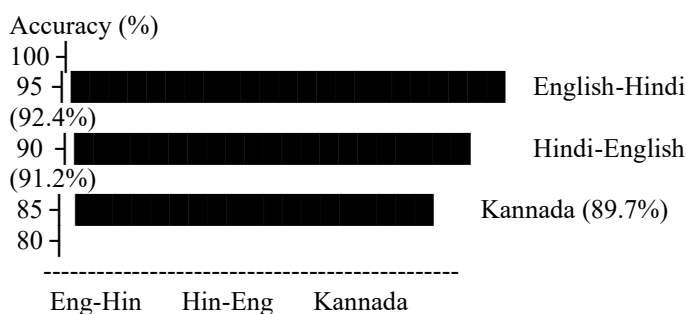


Figure 2.2 : Accuracy of Language Translation

C. POS Tagging and System Efficiency Analysis

Part-of-Speech tagging was performed using Stanza. The English POS tagging system attained an accuracy of 96.1%, whereas Hindi and Kannada recorded accuracies of 92.3% and 88.5%, respectively. The system effectively recognized grammatical categories like nouns, verbs, pronouns, adjectives, and adverbs in sentences across multiple languages. The response time for POS tagging stayed under 500 milliseconds for typical sentence lengths.

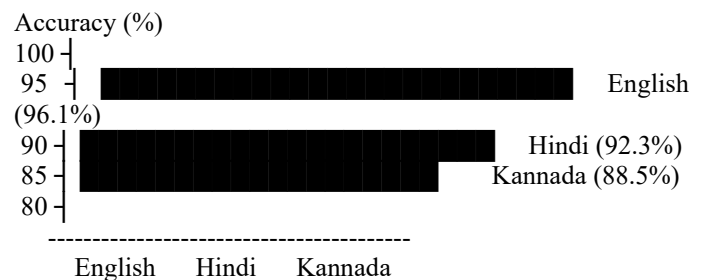


Figure 2.3 : Accuracy of POS tagging

The total system response time, which encompasses speech recognition, translation, and POS tagging, ranged from 2.8 to 5.6 seconds based on audio length and server load. The findings affirm that the suggested architecture is appropriate for small-scale real-time multilingual NLP applications like language education systems, smart assistants, and translation devices.

CONCLUSION

This project introduced a multilingual AI-driven system for translation and linguistic analysis that can handle text and audio inputs in English, Hindi, and Kannada. The suggested system effectively combines speech recognition, multilingual translation, and Part-of-Speech (POS) tagging within a cohesive microservice-oriented framework. The platform delivers an effective and user-friendly setting for real-time language processing applications by integrating a React frontend, a Spring Boot backend, and a FastAPI AI service.

The system employs advanced transformer-based and deep learning models to enhance the accuracy of speech transcription and multilingual translation. Audio recordings are transformed into text via the Whisper speech recognition model, whereas the NLLB transformer model manages multilingual translation. Linguistic analysis utilizes the Stanza NLP toolkit to identify grammatical categories in translated sentences. Experimental evaluation showed that the suggested framework functions efficiently for multilingual speech and text processing activities. The system effectively managed various types of user inputs while producing organized translations and POS tagging results with quick response times. The modular microservice architecture also enhances scalability, maintainability, and adaptability for future improvements.

REFERENCES

- [1] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust Speech Recognition via Large-Scale Weak Supervision," *Proceedings of the 40th International Conference on Machine Learning (ICML)*, vol. 202, pp. 28492–28518, 2023.
- [2] N. Costa-jussà, M. Cross, P. Cıbıkdiken *et al.*, "No Language Left Behind: Scaling Human-Centered Machine Translation," *Transactions of the Association for Computational Linguistics (TACL)*, vol. 10, pp. 1439–1457, 2022.
- [3] Y. Liu, X. Yang, J. Zhang, Y. Xi, and D. Qu, "TAML-Adapter: Enhancing Adapter Tuning Through Task-Agnostic Meta-Learning for Low-Resource Automatic Speech Recognition," *IEEE Signal Processing Letters*, vol. 32, pp. 636–640, 2025.
- [4] S. Yusuyin, T. Ma, H. Huang, W. Zhao, and Z. Ou, "Whistle: Data-Efficient Multilingual and Crosslingual Speech Recognition via Weakly Phonetic Supervision," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 33, pp. 1440–1453, 2025.
- [5] A. M. Samin, M. H. Kobir, M. M. S. Rafee, M. F. Ahmed, M. Hasan, P. Ghosh, S. Kibria, and M. S. Rahman, "BanSpeech: A Multi-Domain Bangla Speech Recognition Benchmark Toward Robust Performance in Challenging Conditions," *IEEE Access*, vol. 12, pp. 34527–34538, 2024.
- [6] T. Hori, M. Kocour, A. Haider, E. McDermott, and X. Zhuang, "Delayed Fusion: Integrating Large Language Models into First-Pass Decoding in End-to-End Speech Recognition," *Proc. IEEE ICASSP*, pp. 1–5, 2025.

- [7] S. Ahmed, M. Singh, S. Bhattacharyya, and D. Coyle, "Decoding Neural Activity for Part-of-Speech Tagging," *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 3079–3084, 2024.
- [8] E. Dunbar, N. Hamilakis, and E. Dupoux, "Self-Supervised Language Learning from Raw Audio: Lessons from the Zero Resource Speech Challenge," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 6, pp. 1211–1226, 2022.
- [9] T. Dalai, T. K. Mishra, and P. K. Sa, "Part-of-Speech Tagging of Odia Language Using Statistical and Deep Learning-Based Approaches," *arXiv preprint arXiv:2207.03256*, 2022.
- [10] G. Wan, T. Mao, J. Zhang, H. Chen, J. Gao, and Z. Ye, "Grammar-Supervised End-to-End Speech Recognition with Part-of-Speech Tagging and Dependency Parsing," *Applied Sciences*, vol. 13, no. 7, pp. 4243–4258, 2023.
- [11] N. T. Luu and O. Bojar, "End-to-End Automatic Speech Recognition and Speech Translation: Integration of Speech Foundational Models and Large Language Models," *arXiv preprint arXiv:2510.10329*, 2025.
- [12] Y. Zhang, X. Chen, and L. Wang, "Transformer-Based Multilingual Neural Machine Translation for Low Resource Languages," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 2, pp. 311–320, 2024.