Interspeech 2025
17-21 August 2025, Rotterdam, The Netherlands

ProBiEM: Acoustic and Lexical Correlates of Prosodic Prominence in English-Malayalam Bilingual Speech

Anindita Mondal[1], *Rahul Biju*[2], *Anil Kumar Vuppala*[1], *Reni K Cherian*[2], *Chiranjeevi Yarra*[1]

[1]Language Technologies Research Center, IIIT Hyderabad, India
[2]CSE, Saintgits College of Engineering, Kottayam, India

anindita.mondal@research.iiit.ac.in, rb.csb2125@saintgits.org, anil.vuppala@iiit.ac.in, reni.cherian@saintgits.org, chiranjeevi.yarra@iiit.ac.in

Abstract

Despite advancements in Speech-to-Speech Translation, maintaining expressiveness between the source and target speech, particularly between English and Indian languages, remains challenging. This study investigates prosodic similarities and variations between English and Malayalam, a language spoken in southern India. A set of 22 prompts from the IViE corpus, covering five categories—simple sentences, WH-questions, questions without morphosyntactic markers, inversion questions, and coordinations—was selected. These prompts, originally spoken by UK speakers, were translated into Malayalam and both language prompts were recorded by bilingual Malayalam speakers while preserving expressiveness. Word-level prominence was manually annotated, and comparisons were made across Indian English, Malayalam, and UK English. The analysis reveals that prominence is retained at key points in Indian English, whereas in Malayalam, it is low due to question specific diacritics and agglutinative nature.

Index Terms: Speech-to-Speech Translation, Prominence, Bilingual prosody, Expressiveness, English-Malayalam

1. Introduction

Significant progress has been made in speech-to-speech translation and audio dubbing, with most research focusing on generating clear and natural speech in the target language. However, relatively little effort has been dedicated to preserving the expressivity of the source language in the translated speech, especially for Indic languages. Several studies such as [1–3] have specifically focused on capturing emphasis in English-to-Japanese translation. Additionally, works like [4–6] have investigated methods for transferring prosodic characteristics across Spanish, English, Italian, French, German and Catalan. The studies by [7–11] made significant contributions to bilingual speech prosody analysis from different perspectives. An exploration of how Chinese-English bilinguals adjust their speech rhythm in quiet versus noisy environments and the role of prosodic features in distinguishing sarcasm and sincerity in bilingual English-Chinese speakers was reported in [7, 8]. The study [9] provided insight into the acquisition of fundamental frequency (F0) in Cantonese-English bilingual children, shedding light on how early bilingual exposure influences pitch patterns in both languages. The focus of [10] was on emotional prosody recognition in bilingual societies, comparing spectral, prosodic, and combined features to enhance speech recognition models whereas [11] examine multilingual speech emotion recognition, comparing English and German speakers' emotional prosody to identify cross-linguistic patterns.

Although bilingual speech prosody has been analyzed across various languages and scenarios, some studies have examined prosodic features in Malayalam speech. Emotional speech prosody in Malayalam and Hindi was compared by [12]. Rhythm Formant Analysis (RFA) was used by [13] to classify Malayalam and other Indic languages. The rhythmic structure of Malayalam poetry was analyzed using computational models by [14]. Studies like [15, 16] demonstrate the complexity of Malayalam morphology, especially in handling verbal and nominal inflections and employ character-level analysis to model morphological patterns. Dictionary Based, Rule Based and Statistical approaches for building Malayalam Machine Translation Systems are depicted in [17, 18]. Despite these, there has been no significant research on the bilingual prosody of English and Malayalam speech.

In this study, prosodic prominence analysis is conducted on English and Malayalam utterances spoken by 132 bilingual speakers of both languages. The analysis is carried out in two stages: (1) to examine the usage of prominence and (2) to analyze word-level prominence variations in conveying the same expressivity. For this study, five different sets of English sentence stimuli, totalling 22 sentences with varying expressivity, were selected from the IViE corpus [19]. To ensure expressivity was preserved, these stimuli were manually translated into Malayalam, and aligned English-Malayalam word pairs along with Parts-of-Speech tags were manually identified. Both English and Malayalam stimuli were used in the recording process, involving 132 bilingual speakers. A total of 5808 utterances in English and Malayalam were manually annotated to indicate prominence at the word and sub-word levels. The prominence variations in these utterances were compared with 792 utterances from the IViE corpus, spoken by 36 UK speakers for the same 22 sentences. We study the usage and preservation of prominence in Indian English, Malayalam, and UK English by analysing the speech at the word level between all three language pairs. From the study, it is observed that speakers retain prominence at critical junctures while speaking English, whereas, in Malayalam, the correlation is lower due to its distinct agglutinative nature.

2. Motivation

The prosodic prominence analysis of the English-Malayalam language pair is critical due to their significant differences in structural and prosodic elements. This is because Malayalam is a Dravidian language that contains agglutinative word forms, which contrasts with English, a Germanic language, in aspects of word order, stress and rhythmic patterns. Malayalam is typically syllable-timed, where each syllable has a similar duration. However, English is stress-timed, where the intervals between stressed syllables are relatively consistent. These differences pose unique challenges in ensuring meaning and naturalness

10.21437/Interspeech.2025-2028

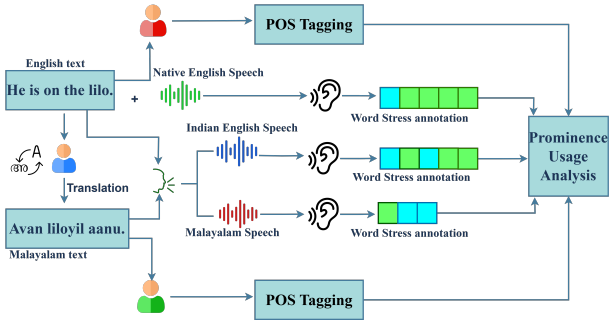

Figure 1: *Steps involved in prominence usage analysis*

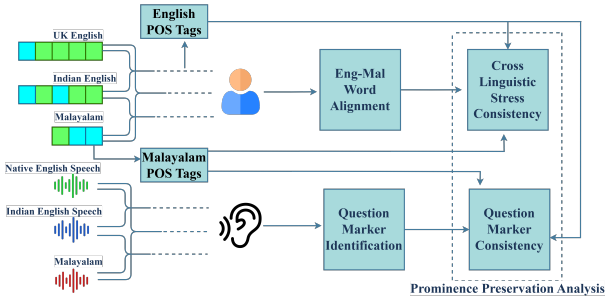

Figure 2: *Steps involved in prominence preservation analysis*

from Speech-to-Speech Translation (S2ST) systems. For effective translation, prosody in Malayalam, which is conveyed through pitch variations for questioning and prominence, must be mapped onto English prosodic features, where stress plays a critical role in distinguishing word meaning. It is to be noted that the prosodic prominence analysis could be helpful in further consideration of English and Indian language pairs.

3. Data Collection

In this study, English and Malayalam sentences were recorded from 132 Malayalam speakers, following the same transcript used in the IViE British Corpus in English (36 UK English speakers) and the corresponding translation in Malayalam. The Malayalam speakers were instructed to read the sentences by preserving expressivity. The English text in the IViE corpus has 22 sentences of the following five sentence types with the acronyms: simple sentences (DEC), WH-questions (WHQ), questions without morphosyntactic markers (DQU), inversion questions (YNO) and coordinations (COO). Thus, the English and the respective translated Malayalam text enable a controlled comparison between the two languages for cross-linguistic studies on prosody, phonetic shifts and prominence variations.

4. Analysis Methodology

A detailed analysis of prominence usage and its preservation between the English and Malayalam language pair, is performed as shown in Figure 1 and 2, respectively. The details of each of the analyses and manual annotations are explained in the two subsections below.

4.1. Prominence Usage Analysis

This analysis examines prominence use across three speech varieties: Indian (Malayalam) English, UK English, and Malayalam as shown in Figure 1. The UK English speech samples are taken from IViE corpus. The Malayalam English and Malayalam speech samples are collected from the same speaker in this study and the word-level prominence markings and parts-of-speech (POS) tags were manually annotated. This analysis focuses on how prosodic prominence is commonly applied to different sentence types and grammatical categories, providing insights into language-specific prominence usage.

4.1.1. Word Prominence and POS annotation

We manually annotated all 5808 speech samples that are recorded from 132 bilingual (Malayalam and English) speakers speaking 22 sentences in both English and Malayalam. Additionally, all speech samples from the IViE corpus were manually reviewed for word prominence markings and POS tagging.

Word Prominence Annotation: Each audio in all three varieties of speech recordings is listened carefully by an annotator and mark the prominence on each word as 1 or 0; 1 indicates that prominence exists, and 0 for no prominence. The annotator is bilingual in Malayalam and English and proficient in both languages. From these annotations, a subset of annotated prominence markers are cross-checked by another two bilingual Malayalam and English speakers and almost all the prominence markers are consistent.

POS Tagging: POS Tagging assigns grammatical labels, such as Subject (Subj), WH-Word, Adjective (Adj), Adverb (Adv), Verb, Noun Phrase (NP), Noun and Noun Phrase (N & NP), Conjuction (Conj), Verb and Verb Phrase (V & VP), Verb and Linking Verb (V & LV), Nouns (N1,N2,N_{oth}), to each word in the dataset. The same annotator assigns these labels to the English and Malayalam sentences. This process is essential for understanding the differences in syntactic structure, grammatical rules, and word prominence relation in the context of English-Malayalam language pair. For instance, English follows a Subject-Verb-Object (SVO) word order, whereas Malayalam uses a Subject-Object-Verb (SOV) structure.

4.2. Prominence Preservation Analysis

Figure 2 presents the details of this analysis, which is divided into two parts:

1) Cross-Linguistic Stress Consistency: This sub-analysis measures the degree to which speakers retain word-level prominence patterns from one language to another. It evaluates whether prominence is consistently applied to similar word or structure pairs that have the same meaning across Indian English, UK English, and Malayalam. This would provide insights into how prosodic features are transferred or adapted in the considered speech varieties. For this analysis, alignment between the words or structures in all 22 English and Malayalam sentences was manually annotated.

2) Question Marker Consistency: This sub-analysis focuses on interrogative sentences, identifying which specific words contribute to the questioning intonation in each language. In this analysis, we have excluded declarative (DEC) sentences. It examines how prominence is assigned to question markers and whether their role in conveying interrogative meaning is consistent across English and Malayalam. For this purpose, the words and word locations responsible for questioning are manually annotated.

4.2.1. English-Malayalam Word Alignment

Malayalam, an agglutinative language, attaches multiple grammatical markers to root words, while English uses auxiliaries and prepositions. All the 22 English and Malayalam sentence

pairs are manually annotated to align words or multiple words using an annotator of bilingual English and Malayalam speaker. From the manual alignment process, it is found that there are three types of alignment pairs: 1) one English word with one Malayalam word, referred to as ***one-to-one*** aligned words, 2) multiple Malayalam words with one English word, referred to as ***many-to-one***,and 3) one Malayalam word with multiple English words, referred to as ***one-to-many***. The distribution of the three aligned pairs for each sentence type are provided in Table 1. The significant total percentage of many-to-one aligned words indicates Malayalam's agglutinative morphology. As an example of this, English phrases such as "is going" and "has been" become single Malayalam words *"pokunnu"* and *"kondirikkukayanu"*, respectively. Also, the highest percentage of one-to-one aligned words suggests the significant syntactic regularity between English and Malayalam for the considered sentences.

Table 1: *Distribution of sentence types - in One-to-One, Many-to-One and One-to-Many alignments*

Sentence type	One-to-One	Many-to-One	One-to-Many
DEC	27	4	-
WHQ	11	-	1
COO	29	2	-
YNO	14	1	-
DQU	10	1	-
Total (%)	**91.0**	**8.00**	**1.00**

In one-to-one alignments, COO are the most common, reflecting consistent mappings for compound or conjoined structures, followed by DEC, which align with basic sentence structures. WHQ show stable one-to-one mappings, likely due to their distinct lexical counterparts in Malayalam, while YNO, which involve subject-verb inversion in questions, highlight structural similarities between the languages. DQU, including tag and rhetorical questions, align without clear structural markers. In many-to-one alignments, DEC are most frequent, reflecting Malayalam's compact morphology that combines multiple English words into one, while COO decrease in frequency. WHQ are not present, as they prefer one-to-one mappings, while YNO and DQU continue to show complex alignment patterns. Overall, DEC's dominance in many-to-one alignments highlights Malayalam's agglutinative structure and the frequent occurrence of COO in both alignment types shows the language's flexibility in handling different syntactic structures.

4.2.2. Question Marker

For this analysis, four distinct sentence types—DQU, YNO, WHQ and COO—were analyzed separately. A manual annotation process was employed to identify the specific words responsible for marking the interrogative nature of each sentence. This method focused on capturing questioning elements based on auditory cues rather than solely on syntactic structures, offering a more nuanced understanding of how questions are formed and perceived.

5. Analysis Results and Discussion

5.1. Prominence Usage

The first analysis examines Indian English, UK English and Malayalam by analyzing the distribution of stressed words across various POS tags. The stressed words for each POS tag were systematically counted in the five type sentences and their proportions were calculated relative to the total number of stressed words within each sentence type and compared across

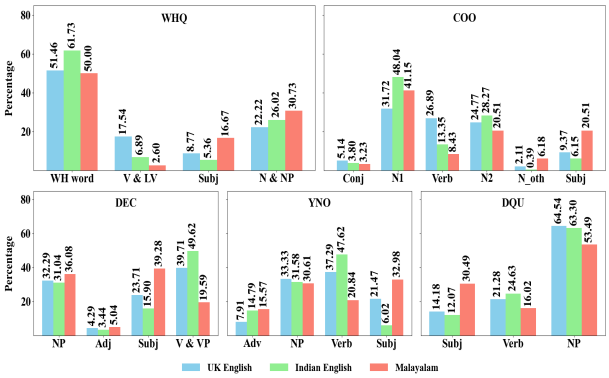

Figure 3: *Stress distribution of POS tags across sentence types*

the three languages. Figure 3 illustrates the distribution of stress words across different sentence types and language varieties, indicating key cross-linguistic differences in prosodic emphasis.

1. In **WHQ**s, WH words receive the highest stress in all three varieties (UK English: 51.46%, Indian English: 61.73%, Malayalam: 50%), indicating their central role in interrogative prosody.
2. **YNO**s show a divergence, with English varieties stressing verbs (UK English: 37.29%, Indian English: 47.62%), while Malayalam stresses subjects more (32.98%).
3. **DQU**s consistently emphasize noun phrases (UK English: 64.34%, Indian English: 63.30%, Malayalam: 53.49%), marking interrogative intent in the absence of explicit markers.
4. **COO** stresses the first noun most in all varieties, decreasing from UK English (31.72%) to Malayalam (41.15%).
5. In **DEC**, English varieties stress verbs (UK English: 39.71%, Indian English: 49.62%), while Malayalam focuses more on subjects (39.28%).

These findings highlight distinct prosodic strategies in UK English, Indian English and Malayalam, with shared patterns such as WH-word emphasis and noun phrase stress, but significant differences in verbal stress, subject emphasis and sentence-level prosody.

5.2. Prominence Preservation

5.2.1. Cross-Linguistic Stress Consistency

The second analysis involved examining the consistency of stress patterns between the following language pairs:

Indian English and Malayalam: In this study, cross-linguistic consistency in stress patterns was evaluated by analyzing word alignments between English and Malayalam. For each sentence type t, there are $N(t)$ alignment pairs. Each alignment pair is checked to determine whether it is matched or unmatched, in each of the sentences. The number of matched pairs, denoted as M, is computed based on the alignment criteria. Consequently, the number of unmatched pairs is given by: $U = N(t) - M$. In the case of a 1-1 alignment, a match was recorded if both the English and Malayalam words exhibited stress simultaneously. For a 2-1 alignment, where two Malayalam words corresponded to a single English word, a match was considered valid if at least one of the two Malayalam words was stressed along with its aligned English word. If neither condition was met, the alignment pair was marked as unmatched. After processing all sentences in a

sentence type, the matching percentage for each sentence type was computed to quantify cross-linguistic stress consistency, given by : Match $\% = \frac{M}{U+M} \times 100$.

UK English and Indian English: To account for differences in the number of speakers between the UK English and Indian English groups, a cross-speaker analysis was conducted to evaluate stress pattern similarities. Each UK English speaker was compared to all 132 Indian English speakers to determine the percentage of stressed word matches across five sentence types.

UK English and Malayalam: In this study, the matching percentages were calculated in the same way as for Indian English and Malayalam, but with each of the 36 UK English speakers compared against all Malayalam speakers. A match is recorded when the stressed word of a UK English speaker is aligned with that of a Malayalam speaker.

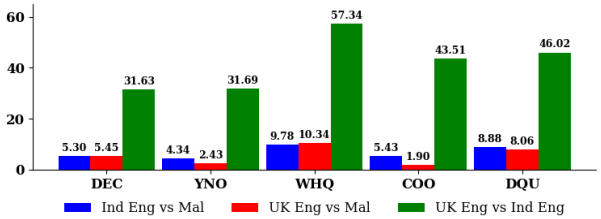

Figure 4: *Comparison of Stress Consistency in Sentence Categories across Different Language Pairs*

A detailed visualization of cross-linguistic prominence consistency patterns across language pairs and sentence types is presented in Figure 4. The key observations are provided below:

1. UK English and Indian English demonstrated the highest overall consistency, with WHQ showing particularly strong matching (57.34%), followed by DQU (46.02%) and COO (43.51%). The relatively high matching percentages between UK English and Indian English suggest substantial prosodic transfer from UK speakers to Indian English speakers.
2. In contrast, both UK English-Malayalam and Indian Enlish-Malayalam language pairs showed significantly lower matching percentages, with WHQ displaying the highest consistency in both Indian English-Malayalam (9.78%) and UK English-Malayalam (10.34%) pairs. However, markedly lower matching percentages between English variants and Malayalam (rarely exceeding 10%) highlight the fundamental prosodic differences between these languages.

5.2.2. Question Marker Consistency

The analysis for question marker was conducted to compare interrogative sentences across UK English, Indian English and Malayalam, focusing on the four sentence types: WHQ, COO, DQU and YNO as shown in Figure 5. The green boxes, dashed boxes and blue ellipses indicate the words responsible for questioning in Indian English, UK English and Malayalam respectively with the corresponding POS tag marked below each word.

WHQ: WH word usage and noun phrase (NP) structures were compared across languages. Examples such as "Where is the manual?" and its Malayalam equivalent *"Laghugrantham evideyaanu?"* demonstrated consistent use of WH words and NPs.

COO: Noun phrase patterns in compound questions were

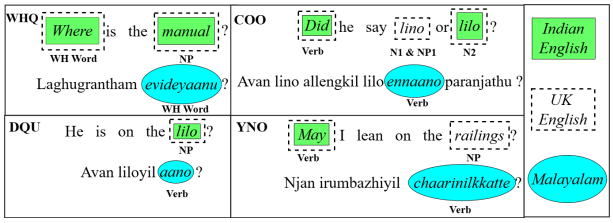

Figure 5: *Comparison of Question Marker Consistency in Interrogative Sentences*

analyzed. Differences were noted: UK English employed multiple noun components, Indian English simplified structures, while Malayalam compressed expressions, as shown in *"Avan lino allenkil lilo ennaano paranjathu?"* where the verb *"ennaano"* functioned as the question marker.

DQU: The use of NP structures in English and verb forms in Malayalam to indicate questions was analyzed. For example, "He is on the lilo?" and *"Avan liloyil aano?"* illustrated Malayalam's verb-centric *("aano")* question marker usage.

YNO: Verb-centric constructions were examined through sentences like "May I lean on the railings?" and *"Njan irumbazhiyil chaarinilkkatte?"*, highlighting the critical role of verbs) in forming interrogatives, with Malayalam employing a verb-driven structure.

Table 2: *Distribution of question markers across the four sentence types*

Sentence Types		Question Markers (POS)		
Type	**Sentence**	**UK English**	**Indian English**	**Malayalam**
WHQ	S1	WH word, NP	WH word, NP	WH word
	S2	WH word, NP	WH word, NP	WH word
	S3	WH word, NP	WH word, NP	WH word
COO	S1	Verb, N2, N1&NP1	Verb, N2	Verb
	S2	Verb, N2, N1 & NP1	Verb, N2, N1 & NP1	Verb
	S3	Verb, N2, N1 & NP1	Verb, N2	Verb
	S4	Verb, N1 & NP1	Verb, N2	Verb
	S5	Verb, N2, N1 & NP1	Verb, N2	Verb
DQU	S1	NP	NP	V & LV
	S2	NP, V & LV	NP	V & LV
	S3	NP	NP	V & LV
YNO	S1	Verb, NP	Verb	Verb
	S2	Verb, Adverb	Verb, Adverb	Verb
	S3	Verb, NP	Verb, NP	Verb

Table 2 shows the distribution of question markers across the four sentence types highlighting cross-linguistic variations in syntactic patterns and the distinct morphological strategies employed by each language. The results revealed Malayalam's agglutinative nature, resulting in verb-dominant question markers, contrasting with the noun-based structures of English. And, Indian English exhibited simplified patterns compared to UK English, reflecting regional linguistic adaptations.

6. Conclusion

Our study examined prosodic prominence in bilingual English-Malayalam speech, highlighting differences in stress patterns. Indian English retains prominence similar to UK English, while Malayalam shows lower prominence retention due to its syllable-timed rhythm and agglutinative structure. English relies on prominence for meaning, whereas Malayalam uses pitch variations and verb modifications, making direct prosody transfer difficult. Our findings have important implications for speech translation and dubbing. Preserving expressivity is key to natural speech synthesis. Future work will expand to other Indian languages and develop automated prominence modeling to enhance multilingual speech technology, ensuring more expressive and natural communication.

7. References

[1] Q. T. Do, T. Toda *et al.*, "Preserving word-level emphasis in speech-to-speech translation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 3, pp. 544–556, 2017.

[2] Q. T. Do, S. Sakti, and S. Nakamura, "Sequence-to-sequence models for emphasis speech translation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 10, pp. 1873–1883, 2018.

[3] Q. T. Do, S. Sakti *et al.*, "Transferring Emphasis in Speech Translation Using Hard-Attentional Neural Network Models," in *Proc. Interspeech 2016*.

[4] P. Aguero, J. Adell, and A. Bonafonte, "Prosody generation for speech-to-speech translation," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 1, 2006, pp. I–I.

[5] W.-C. Huang, B. Peloquin *et al.*, "A holistic cascade system, benchmark, and human evaluation protocol for expressive speech-to-speech translation," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

[6] J. Swiatkowski, D. Wang, M. Babianski, P. Lumban Tobing, R. Vipperla, and V. Pollet, "Cross-lingual prosody transfer for expressive machine dubbing," in *Interspeech 2023*, 2023, pp. 4838–4842.

[7] S. Y. Chin, "Decoding sarcasm and sincerity: a comparative analysis of prosodic features in english-chinese bilingual speech using svm," 2024.

[8] B. Li and J. Gong, "Comparison of chinese and english rhythm produced by chinese speakers in quiet and noise," in *2024 International Conference on Asian Language Processing (IALP)*, pp. 109–113.

[9] Y. Zhu, "The acquisition of speaking fundamental frequency (f0) features in cantonese and english by simultaneous bilingual children," *Linguistics Vanguard*, vol. 10, no. 1, pp. 373–384, 2024. [Online]. Available: https://doi.org/10.1515/lingvan-2023-0105

[10] O. Goncharova, "Towards sustainable development in speech recognition: Enhancing emotional understanding in bilingual societies," *Journal of Lifestyle and SDGs Review*, vol. 5, p. e04198, 12 2024.

[11] K. Ghaayathri Devi, K. Likhitha, J. Akshaya, R. Gokul, and G. Jyothish Lal, "Multi-lingual speech emotion recognition: Investigating similarities between english and german languages," in *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, pp. 1–10.

[12] M. M. Mathew and J. S. Bhat, "Aspects of emotional prosody in malayalam and hindi," 2010. [Online]. Available: https://api.semanticscholar.org/CorpusID:73712787

[13] P. Gogoi, S. Kalita, P. Sarmah, and S. R. M. Prasanna, "Exploring rhythm formant analysis for indic language classification," 2024.

[14] J. M P and K. Balakrishnan, "Identification and extraction of features from malayalam poems for analyzing syllable duration patterns," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, Mar. 2023.

[15] J. Jayan, R. R R, and R. Sankaravelayuthan, "Morphological analyser and morphological generator for malayalam tamil machine translation," *International Journal of Computer Applications*, vol. 13, pp. 15–18, 01 2011.

[16] P. B, S. K.P., and M. A. Kumar, "A deep learning approach for malayalam morphological analysis at character level," *Procedia Computer Science*, vol. 132, pp. 47–54, 2018, international Conference on Computational Intelligence and Data Science.

[17] B. T.K, Professor, and J. Abraham, "A study of machine translation methods an analysis of malayalam machine translation systems," 02 2014.

[18] S. S, "Machine translation between malayalam and english," vol. 14, pp. 7–31, 12 2020.

[19] F. Nolan and B. Post, "475the ivie corpus," in *The Oxford Handbook of Corpus Phonology*. Oxford University Press, 2014.