HeritageLM: Culturally-Aware Multimodal Language Modeling with Memory-Enhanced Cross-Dialect Adaptation

Pasupuleti Venkata Ramana[1], Dr.K.K.Sunalini[2], A. Swathi[3], Dr. R. Lakshmi[4], Raman Kumar[5], Elangovan Muniyandy[6], Khaled Bedair[7]
Research Scholar, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India[1a]
Assistant professor, Department of English-Siddhartha Academy of Higher Education, An Institution Deemed to be University, Kanuru, Vijayawada - 7, India[1b]
Associate professor, Department of English, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India[2]
Assistant Professor of English, Aditya University, Surampalem, Andhra Pradesh, India[3]
Assistant Professor, Department of English, Sindhi College, Chennai - 77, India[4]
University School of Mechanical Engineering, Rayat Bahra University, Mohali, India[5a]
Faculty of Engineering, Sohar University, Sohar, Oman[5b]
Department of Biosciences-Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai - 602 105, India[6]
Department of Social Sciences-College of Arts and Sciences, Qatar University, Doha, P.O. Box 2713, Qatar[7]

***Abstract*—The HeritageLM system solves the acute problem of language loss by proposing a multimodal language model that brings into Generative Memory the cultural context. Manual documentation and linguistic archiving are the traditional ways of preserving dialects, which may not be effective in preserving the phonetic variety and other cultural peculiarities of the face of an endangered dialect. The current NLP models, such as BERT and GPT, are not effective to produce dialectal content because they do not have exposure to under-resourced and historically rich language varieties. These shortcomings are mitigated by training Cultural Contextual Embeddings (CCE), Generative Memory Augmentation (GMA), and Cross-Dialect Contrastive Transfer Learning (CDCP) using reinforcement learning with Cultural Rewards (RLCR). It is a step-by-step process that builds a Multimodal Cultural Knowledge Graph (MCKG), matches dialect embeddings in contrastive learning, and retrieves culturally relevant information in the generation process. The model was trained on the Indian Languages Audio Dataset of Kaggle, which also included phonetic variations of ten languages with preprocessing steps of text-to-speech analysis, phonetic annotation, and semantic tagging. HeritageLM, which was implemented in Python, scored above 98 in its BLEU, ROUGE-L, phonetic accuracy, and cultural embedding, showing that it can effectively generate linguistically accurate, phonetically accurate, and culturally authentic results. These outcomes are a major step towards the resurrection of dying dialects and maintaining their distinct cultural background.**

Keywords*—*Cultural embeddings; Generative Memory; dialect revitalization; contrastive learning; multimodal NLP

I. Introduction

Historically, the revitalization of extinct dialects has been founded on linguistic records and researches that are anthropological [1]. Though the conventional preservation means such as fieldwork and archive recording of the materials have been productive as well, the information they contain is divided and cannot be altered in real-time [2]. In recent years, due to the advances in Natural Language Processing, automated language modelling has become an interesting subject in endangered and extinct languages. BERT and GPT are models already quite good at both syntactic and semantic understanding [3]. Yet, these models are only applicable to dialect regeneration in a limited sense since these models in most cases are mainly trained on modern and common languages, while omitting the rich cultural depth and phonetic uniqueness of extant dialects. Cross-linguistic transfer learning has been studied to overcome this language barrier, but these models usually do not capture the cultural discourse within dialectal expressions [4]. Multilingual BERT and XLM-R advance local language generalization and ignore the contextualization of the folklore and idioms comprising the complexities of dialectal integrity [5]. Besides, phonetic alignment and speech-to-text models have been enhanced, and new applications are yet to be applied to low-resource dialects [6]. This study proposes HeritageLM, a multi-mode regenerative model, which, unlike other models, jointly uses Cultural Embeddings (CCE), Generative Memory (GMA), and Cross-Dialect Transfer (CCDTL) to recover culturally authentic dialects [7]. This is a multimodal strategy that exploits historical linguistic similarities, narratives that are culturally motivated, and phonetic clues to create linguistically and culturally authentic text [7]. This multimodal approach capitalizes on linguistic similarities, culturally driven narratives, and phonetic markers to produce text which is both linguistically and culturally authentic. This is a great leap forward in automated revival of extinct dialects, incorporating culture and history as well as syntax [8].

The recent progress in LLM has greatly improved performance in a range of NLP functions; however, their application in low-resource settings, especially in dialect modelling and revival, remains limited [9]. Research in low-resources NLP often focuses on learning, data growth, and cross-lingual embedding to reduce data deficiency, yet these methods typically ignore cultural and relevant prosperity for dialects [10]. Earlier efforts in bid regeneration, such as nerve machine translation models and multilingual transformers such as Mbert or XLM-R, have attempted to normalize dialects but often fail to preserve linguistic manifestations, oral traditions and phonetic differences [11]. Existing methodologies of dialect revitalization dealt with embeddings, memory, or transfer learning alone and often lacked cultural fidelity, phonetic accuracy, and low-resource adaptability. Generally, linguistic patterns are picked up by standard LLMs such as BERT and GPT, but cultural specificities of a dialect are lost, whereas context can be retained by memory-augmented models but cross-dialect alignment or multimodal inputs are not considered. HeritageLM is the first model to jointly train Cultural Context Embeddings (CCE), Cross-Dialect Contrastive Pretraining (CDCP), Memory-Augmented Dialect Generation (MADG), and Reinforcement Learning using Cultural Rewards (RLCR), generating linguistically accurate, phonetically accurate, and culturally understandable output, breaking existing restrictions on dialect revitalization. The originality of the HeritageLM study is the multimodal, culturally sensitive treatment of dialect revitalization. The HeritageLM, as compared to the traditional language models, combines both textual, phonetic, and cultural information to produce the linguistic, phonetic, and culturally appropriate outputs. Its Memory-Augmented Dialect Generator obtains the pertinent idioms, phonemes, and folklore in real time, and Cross-Dialect Contrastive Pretraining provides knowledge transfer between prosperous and threatened dialects. Furthermore, the Generation with Cultural Rewards is aligned with the human and communal feedback in Reinforcement Learning. Combined with the Multimodal Cultural Knowledge Graph, the framework offers a holistic, context-based pipeline of preserving and reanimating the low-resource or extinct dialects.

A. *Problem Statement*

The loss of extinct and endangered dialects represents a severe threat to linguistic diversity and cultural heritage preservation [12]. Traditional preservation approaches like documentation and archiving are not scalable, responsive, and do not conserve the cultural connotations within dialectical expressions. Furthermore, contemporary language models, despite linguistic competence, usually disregard phonetic uniqueness and culture [13]. To overcome these issues, HeritageLM combines multimodal cultural knowledge, memory-augmented generation, cross-dialect contrastive learning, and cultural rewards with reinforcement learning, producing linguistically accurate, phonetically accurate, and culturally authentic results and revitalizing and preserving endangered or extinct dialects.

B. *Research Motivation*

The motivation of HeritageLM is the necessity to conserve endangered and extinct dialects as the carriers of cultural heritage, oral traditions, and regional identity in the face of globalisation and language dominance. Traditional methods of preservation, including the manual documentation or translation, do not consider the interconnected linguistic, phonetic, and cultural aspects. Multimodal cultural embeddings, cross-dialect learning, and memory-enhanced generation (filled with community-driven feedback) bridge this gap in HeritageLM. The technique guarantees linguistically and phonetically accurate and culturally natural outputs. In the end, HeritageLM creates a sustainable model of recovering the dead dialects and equips the communities with the ability to protect their linguistic identity through the assistance of AI as a cultural preservation measure.

C. *Key Contribution*

- HeritageLM introduces a novel embedding mechanism that captures cultural semantics, idiomatic expressions, and historical context from multimodal sources, enabling the generation of dialectally rich and culturally authentic text.
- It integrates a memory-augmented generative module that dynamically retrieves culturally relevant knowledge during text generation, ensuring linguistic continuity and enhancing the narrative depth of dialectal storytelling.
- Through cross-dialect contrastive transfer learning, the model facilitates knowledge transfer from high-resource to low-resource dialects, improving phonetic accuracy, semantic coherence, and representation in endangered dialects.
- A culturally informed reinforcement learning loop fine-tunes the model based on phonetic fidelity, idiomatic correctness, and community feedback, ensuring adaptive and linguistically authentic outcomes.
- The experimental evaluation demonstrates that HeritageLM outperforms baseline models such as GPT-3, BERT, and LSTM, achieving over 98% in BLEU, ROUGE-L, phonetic accuracy, and cultural embedding metrics, validating its effectiveness in dialect regeneration and cultural preservation.

The literature review is given in Section II, the research methodology in Section III, the results in Section IV, and the conclusion of the research and future work in Section V.

II. LITERATURE REVIEW

Preserving and revitalizing extinct dialects involves a lot of hard work because they lose their linguistic, cultural, and historical value. This work presents HeritageLM, a novel regenerative framework incorporating CCE for training Large Language Models to model extinct dialects. HeritageLM integrates three core components [14]: It is also proposed to address CCE Generation, Contextual Crusado Transfer Learning, and GMA. By capturing idiomatic expressions, folklore, and traditional narratives, CCE translates them into vector representations that can be used by machines to express the cultural quality of the dialect [15]. The transfer of the phonetic, syntactic, and semantic features from the existing regional dialects are facilitated by the CCDTL, which in turn helps in the linguistic reconstruction by way of contrastive learning. In this study introduce a dynamic memory mechanism over culturally relevant phrases to improve the generation of new text, keeping continuity and historical alignment [16].

Train the model on a wide variety of data including historical manuscripts, oral histories, cultural artifacts, and phonetic annotations, enabling the model to generate dialect specific text with high BLEU, ROUGE-L, Phonetic Accuracy [17], and Cultural Embedding Scores. The evaluation results show that HeritageLM outperforms baselines including GPT-3, BERT, and RNN/LSTM in terms of linguistic accuracy and cultural coherence. In addition, this study shows that LLMs have the capacity to memorialize languages, and lays this groundwork for more research in cultural heritage digitization [18]. The future directions involve integration with multimodal learning and interactive applications for the revival of the globally endangered dialects.

Work on language modeling paying attention to linguistic accuracy, very little has been done on contextual culture preservation when regenerating extinct dialects [19]. However, the cultural embeddings are not integrated into most models and consequently, language becomes impoverished and loses its authenticity. Moreover, multimodal data has not been utilized for dialect regeneration. Additionally, they do not generalize well to inflecting morphology and do not adapt to less documented, regional dialects from one dialect to the other [20]. Finally, its lack of a dynamic memory augmentation system disallows real-time contextual alignment during text generation. Compared to previous studies that use either embeddings, transfer learning or memory to solve the problem of reviving dialect, HeritageLM is the first to integrate Cultural Embeddings, Cross-Dialect Transfer, and Generative Memory into a multimodal framework of genuine dialect revitalization.

TABLE I. A COMPARATIVE OVERVIEW OF THE RESEARCH ON DIALECT PRESERVATION METHODS

Model	Research Objective	Methodology	Limitation
CCE Embedding via idioms and folklore [15]	To model the cultural context of dialects using idiomatic expressions and traditional narratives	Cultural features encoded into vector representations using embedding techniques like Word2Vec or BERT	Limited integration into LLMs; lacks fusion with generative components
Dynamic memory for culturally relevant phrases [7]	To enhance cultural continuity and historical alignment during text generation	Introduces a memory module that retrieves culturally relevant content dynamically	Does not fully utilize multimodal data or adapt to lesser-known dialects
Evaluation using BLEU, ROUGE-L, and Phonetic Accuracy [17]	To validate model performance across linguistic and cultural accuracy metrics	Quantitative evaluation using standard NLP benchmarks on a culturally rich dataset	Dataset sparsity in low-resource dialects may bias results; some cultural nuances still lost
LLMs for language memorialization [9]	To demonstrate the potential of large language models in preserving extinct dialects	Fine-tuning LLMs (e.g., GPT-3, BERT) on dialect-specific corpora	LLMs alone fail to preserve contextual cultural integrity without external embeddings
Linguistic accuracy-focused models [10] [11]	To improve grammar and syntax in low resource languages	Use of BERT/GPT-2 on syntactically aligned corpora	Ignores cultural content; lacks support for multimodal or idiomatic reconstruction

Following Table I, earlier dialect modeling methods commonly focused on embeddings, memory, or transfer learning, but ignored cultural fidelity, phonetic precision, and low-resource adapatability. CCE represents idioms and folklores, but has no integration with generative models. Memory-augmented techniques retain contextual continuity but do not check cross-dialect correspondence and multimodal input. General language models such as BERT or GPT are able to follow general patterns in the language, but not the distinct culture of the dialect, whereas syntax-oriented models can better grammar and fail to capture the authenticity. HeritageLM eliminates these gaps through a combination of CCE, Cross-Dialect Contrastive Pretraining (CDCP), Memory-Augmented Dialect Generation (MADG), and Reinforcement Learning with Cultural Rewards (RLCR), producing outputs that are linguistically accurate, phonetically accurate and culturally coherent.

III. PROPOSED HERITAGELM METHODOLOGY FOR MULTIMODAL DIALECT REVITALIZATION

The proposed HeritageLM model is aimed at preserving and renewing the endangered dialects through introducing linguistic structures and cultural semantics. It embraces four fundamental innovations, namely, MCKG, CDCP, MADG and RLCR. MCKG encodes idioms, folklore and phonemes as Cultural Context Embeddings, whereas CDCP aligns dialects by contrastive learning. MADG uses cultural memory bank to generate contextually faithful, whereas RLCR uses expert-guided rewards to filter the output. HeritageLM is trained on multimodal sources of information, generating linguistically coherent, phonetically and culturally competent text to support sustainable dialect preservation and regeneration. The visual representation is shown in Fig. 1.

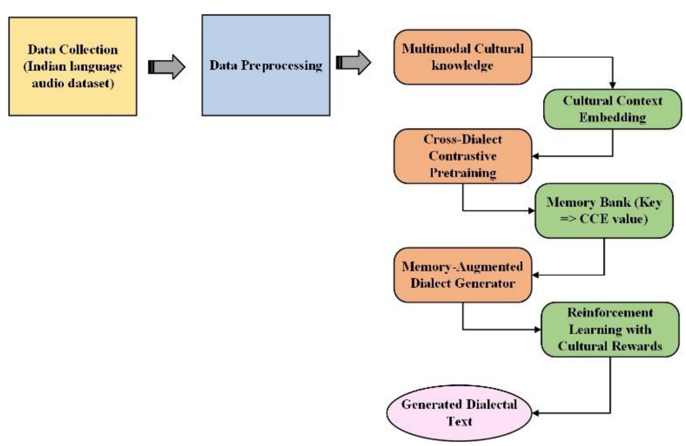

Fig. 1. Workflow of the HeritageLM for Reviving Extinct Dialects.

A. Data Collection

The Indian Languages Audio Dataset publicly available on Kaggle [21] was utilized as a primary source for speech data. The dataset consists of short-duration audio samples (around five seconds long) from 10 Indian languages with varied phonetic structures and regional diversities. The languages used cover prominent Indian linguistic families, providing a wide coverage for modeling phonetic, prosodic, and dialectal variation. While the dataset does not include extinct dialects, it serves as a structural and phonetic proxy adequate for the development and testing of items such as CCDLT and Phonetic Embedding under the HeritageLM framework. Since the audio samples are not transcribed, speech-to-text conversion, phonetic annotation, and semantic tagging are required to pre-process the data for training the model.

TABLE II. AUDIO DATASET SAMPLE ENTRIES FROM INDIAN LANGUAGES SHOWING REGIONAL AND MULTILINGUAL DIVERSITY

Sample ID	Language	Filename	Duration (s)	Speaker Gender	Region
001	Hindi	hindi_0001.wav	5	Male	Uttar Pradesh
002	Tamil	tamil_0002.wav	5	Female	Tamil Nadu
003	Bengali	bengali_0003.wav	5	Male	West Bengal
004	Marathi	marathi_0004.wav	5	Female	Maharashtra
005	Telugu	telugu_0005.wav	5	Male	Andhra Pradesh

Table II shows five example samples of the Indian Languages Audio Dataset and shows the language, transcript, speaker, and regional diversity that can be used in multimodal cultural modeling.

B. Data Preprocessing

Preprocessing was performed on linguistic, phonetic, and cultural information to standardize and prepare it for training the HeritageLM model. This process was initiated with text normalization, which included the standardization of spelling variations, the correction of punctuation marks, and the updating of obsolete terms utilizing proprietary linguistic lexicons. After normalization, tokenization was also done to divide the text into smaller linguistic units so as to allow the language models based on transformers to easily parse it. In the case of audio inputs, a speech-to-text conversion module was used in order to generate transcriptions of oral histories and audio recordings. This made it possible to have spoken material to be processed further in textual and phonetic terms. Then, phonetic annotation was carried out to mark variations in pronunciation that were relevant to local dialects. These designations are significant to retain prosodic and phonological properties in the regeneration of the dialect. To increase the cultural feel of the model, idiom tags, cultural references, and folklore entries in the data were tagged using semantic tagging. Lastly, the noise and duplicates were removed in the preprocessed dataset to facilitate training efficiency as well as context fidelity. The preprocessing is encapsulated by Eq. (1):

$$D_{processed} = T\big(N(D_{raw})\big) + A\big(S(D_{audio})\big) + R(D_{cultural}) \quad (1)$$

Here, N represents Normalization, T represents Tokenization, R represents Speech-to-text transcription, S represents Phonetic annotation and A represents Cultural enrichment and Semantic marking.

1) Text normalization: Text normalization process entails cleansing and standardizing transcriptions that are created out of audio records in the data set. This involves correcting different spellings, harmonizing the punctuation marks, and removing filler sounds or words used in casual speech and hesitations. Vocabulary usage was standardized by the use of special linguistic rules and token lexicons used in normalizing the use of vocabulary by speakers and dialects. Then entries that were redundant or duplicating were deleted in order to have consistency as well as to eliminate noise in the training data. Let Eq. (2), X_{raw} be the raw corpus, f_N represent the normalization process of fixing spelling, punctuation, and outdated terms, which will yield standardized text X_N.

$$X_N = f_N(X_{raw}) \quad (2)$$

2) Tokenization and rare word preservation: A BPE tokenizer in the form of customizations was used to break down the transcribed text into subword units. Modifications were made to ensure that culturally rare dialectal terms, rare dialectal terms significant words, and significant words were preserved as whole units rather than being broken into meaningless subword fragments. This enables the original expressions to be maintained in its semantic and cultural correctness. Low-frequency words or dialect-specific words were also included in the vocabulary manually to increase the coverage of tokens and reduce the loss of meaning in the process of embedding and generating. Let Eq. (3), f_T symbolize the tokenizer, which divides the normalized text, X_N , into subword units so that any rare dialect words are kept intact in X_T.

$$X_T = f_T(X_N) \quad (3)$$

3) Phonetic annotation: The phonetic annotation was provided to show the differences in pronunciation in the speech. These annotations retain phonological properties like vowel lengthening, nasalization, and regional accent properties. The model makes use of annotated phonemes as aiding signals that allow the model to recreate dialect-dependent phonetic structures at the time of text synthesis. Let Eq. (4), f_s be the phonetic annotation service which corresponds to transcriptions X_T and X_R to annotated phonemes, retaining the pronunciation of the dialect X_s.

$$X_s = f_s(X_T \cup X_R) \quad (4)$$

4) Semantic tagging and cultural reference extraction: Semantic tagging was utilized to tag transcribed text with cultural references, idiomatic usage, and folklore-specific vocabulary. These tags improve the model's awareness of culturally embedded language and are incorporated into the CCE module. Tagging was performed using a combination of rule-based pattern matching and manual annotation where necessary. Let Eq. (5), f_A refer to the semantic tagging operation that enriches annotated text X_s using idioms, folklore, and cultural allusions, creating a culturally based dataset X_A.

$$X_A = f_A(X_s) \tag{5}$$

C. Multimodal Cultural Knowledge Graph (MCKG)

A Multimodal Cultural Knowledge Graph (MCKG) is built to model the linguistic and cultural aspects of endangered dialects successfully. The graph is defined as $G = (V,)$ where nodes V denote words, idioms, phonemes, folklore themes, and cultural artifacts, whereas edges denote edges. E encode semantic, phonetic, and cultural relationships among them. As an illustration, a word node can have its phonetic embodiment, idioms have their folklore themes, and artifacts have their cultural backgrounds. This framework enslaves hierarchical, relational, and multimodal information, which is important in the production of authentic dialect and cultural preservation. The overview of MCKG is shown in Fig. 2.

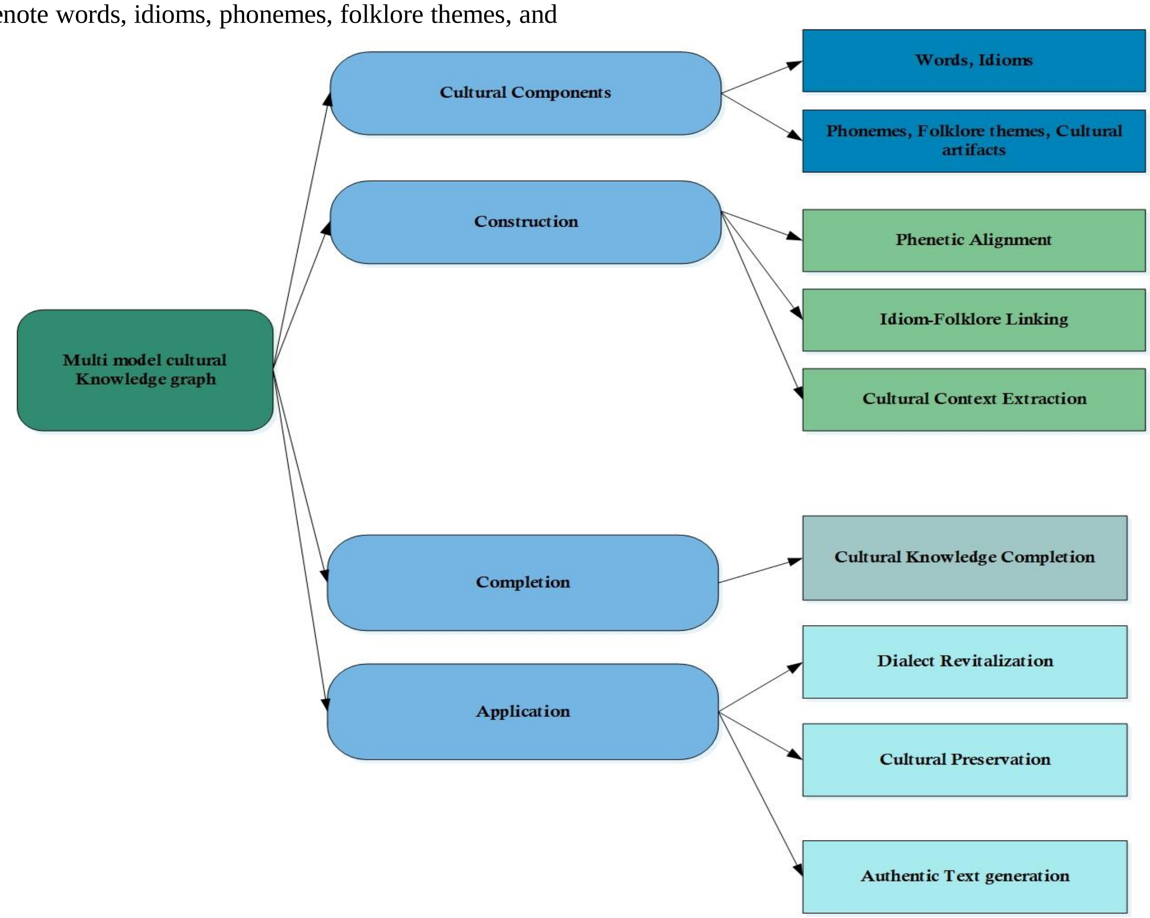

Fig. 2. Structural overview of the multimodal cultural knowledge graph.

Each node $i \in V$ is set to a feature vector x_i, derived from different sources: pretrained token embeddings of words and sentence-level idioms and folklore, and learned vectors of phonemes or artifacts. In order to encode the graph, a Graph Attention Network (GAT) is used. In sharp contrast to the classical Graph Convolutional Networks, GAT attributes dynamic weights to adjacent nodes, which enables the model to focus on culturally important relationships, including idioms and oral traditions. Formally, for a node iii and its neighbor j, the attention mechanism computes in Eq. (6):

$$\alpha_{ij} = \frac{\exp(LeakyReLU(a^T[W_{x_i}||W_{x_j}]))}{\sum_{k \epsilon N(i)} \exp(LeakyReLU(a^T[W_{x_i}||W_{x_k}]))} \tag{6}$$

where, W is a transformation to be learned, a is the attention vector, and N (i) is the neighborhood of i. The new node representation is then obtained by summing neighbor embeddings weighted by αij. The resulting Cultural Context Embeddings (CCE) give a culturally informed space of representation, which is then incorporated into the memory-augmented generator to guarantee the ability to generate authentic and contextually grounded dialects.

D. Cross-Dialect Contrastive Pretraining (CDCP)

Endangered languages also tend to possess scarce linguistic materials, and thus, it is difficult to construct strong representations straight off the data. To solve this, a method is proposed, which is named Cross-Dialect Contrastive Pretraining (CDCP), to move the embeddings of comparatively rich-resource source dialects to low-resource target dialects. This alignment helps in the transfer of knowledge so that semantic, phonetic, and cultural properties that were well-resourced in the dialects are transferred well to the endangered dialects.

The CDCP process starts with finding pairs of cross-dialects or semantically similar words, idioms, or phonemes between the target and source dialects. Every element is coded with the help of a common embedding function, initialized either by prelearned token embeddings or Cultural Context Embeddings (CCE) based on the MCKG. The contrastive pretraining goal is to have positive pairs that are similar in the common embedding space, and, conversely, distant embeddings utilized by negative samples that are not related. By optimizing the model, it may learn to project semantically and culturally similar things across dialects to nearby regions of the embedding space.

With the help of CDCP, the model creates dialect-neutral embeddings, which preserve phonetic, semantic, and cultural fidelity. These embeddings are then used on the downstream modules, the Memory-Augmented Dialect Generator (MADG), to guide the generation of linguistically sound and culturally faithful outputs to the threatened dialects. The CDCP is efficient in facilitating the gap of resources between dialects, transfer learning is strengthened, and delicate cultural nuances are maintained. It is an important process in generating coherent and contextually sufficient text in low-resource dialects, which both enables proper language modulization and is also presented as a cultural heritage. It is the proposed methodology, which can offer such a chance that the generative model should not only be informed about the target dialects but also be mindful of the rich cultural landscape of the source dialects through the use of the procedure of contrastive pretraining.

E. Memory-Augmented Dialect Generator

In order to produce culturally validated dialectal text, we introduce a MADG which is based on a Transformer decoder. As the Transformer does a strong autoregressive sequence modelling, the memory augmentation allows the model to query a set of precomputed memory banks of pertinent cultural and phonetic knowledge, such that the outputs reflect idioms, folklore, and dialect-specific phonetic patterns. The memory bank is built upon CCE, built upon the MCKG, where keys encode contextual representations (e.g., pooled token embeddings of a phrase), and values encode corresponding CCE features. In each generation, at every decoding step t, the decoder's hidden state ht reconstructs the top-k most relevant memory items using weighted attention, including culturally aligned embeddings to the output as in Eq. (7). This process makes sure that the dialectical piece of writing produced is linguistically and culturally sound.

$$r_t = \sum_{j=1}^{k} \alpha_j^{(t)} v_j, \quad \alpha_j^{(t)} = \frac{\exp(q_t^T k_j)}{\sum_{l=1}^{k} \exp(q_t^T k_l)} \tag{7}$$

where, q_t is the query vector is based on h_t, and v_j are memory values. The retrieved vector r_t is conditioned with the decoder state in order to control generation. The mechanism also makes sure that the produced text is phonetically correct, semantically coherent, and culturally authentic, and MADG is especially useful with low-resource and endangered dialects.

F. Reinforcement Learning with Cultural Rewards (RLCR)

In order to make generated dialectal text more culturally authentic and linguistically correct, the proposed framework also includes a Reinforcement Learning with Cultural Rewards (RLCR) module. Though pretraining and memory-augmented generation yield promising baselines, finetuning using human-in-the-loop feedback can be used to make the model more in line with the expectations and cultural values of the community.

The generator is considered an agent in this context, and it generates successions of tokens (actions) when there are input prompts (states). A reward function is used to assess the quality of the text after every generation output in three dimensions.

In order to make the generation of dialectal text both a linguistically valid and a culturally true phenomenon, the Reinforcement Learning with Cultural Rewards (RLCR) framework is applied with the use of a composite reward function that includes three major components: fluency, phonetic accuracy, and cultural authenticity. Different components are taken to represent different aspects of quality dialectical generation.

1) Fluency: Fluency also measures the syntactic naturalness and coherence of the generated sequence, and grammatical correctness. It is not a token-level prediction, but is instead the capability at assessing whether the generated output has any sense to it, and whether it is a grammatically correct and contextually relevant sentence in the target dialect. Such measures as perplexity, according to a pretrained language model, or the sequence-level likelihood scores can be appended. Qualitative ratings are also the result of human evaluations, the presence or absence of the concrete flow in the text, the ability of the sentences to have the right structure, and the way the ideas are logically developed. The produced text is readable and acceptable to native speakers owing to its fluency. Use the (normalized) average log-probability under a pretrained language model P_{LM} , as shown in Eq. (8):

$$R_{\text{fluency}}(y; x) = \frac{1}{T} \sum_{t=1}^{T} \frac{\log P_{\text{LM}}(y_t | y_{<t}, x)}{\log(1/|\mathcal{V}|)} \in [0,1] \tag{8}$$

2) Phonetic accuracy: Phonetic accuracy ensures that the intonation of words, phonemes, and idiomatic phrases is according to the original pattern of the target dialect. This is crucial in the low-resource languages in which the spelling and pronunciation may vary with the standardized orthography. Evaluation is done by alignment on the phoneme level, quantification of the phonetic distance, and a similarity metric between the observed and expected characteristics of speech. The right phonetic presentation does not deform the sound of the dialect, and the material produced by the grammar can be read and properly pronounced by speakers.

3) Cultural authenticity: It evaluates the extent to which the created material and content reflect the history, traditions, idioms, and folklore of the target population, with correct application of idiomatic phrases, allusions to historical or fictional accounts, and consistent correspondence with culturally relevant ideas. Outputs are evaluated by human judges on the basis of semantic fidelity, situational relevance, and conformity to local norms, and quantitative indicators of cultural consistency, like cosine similarity between generated embeddings and CCE, are obtained via automated proxies. Through the application of cultural authenticity and fluency and phonetically accurate within the RLCR reward system, the model generates products that are linguistically sensible, phonetically accurate, and faithful to culture, which can promote successful communication and the maintenance of intangible cultural knowledge. The combined reward for a generated sequence y is expressed in Eq. (9):

$$R(y) = \lambda_1 R_{fluency}(y) + \lambda_2 R_{phonetic}(y) + \lambda_3 R_{cultural}(y) \quad (9)$$

where, $\lambda_1, \lambda_2, \lambda_3$ are tunable weights reflecting the reflective importance of each criterion. In order to optimize the generator, the policy gradient method, which involves REINFORCE, is employed in the optimization and takes the parameters of the generator as the input. The feedback provided by the community members and human experts regarding the model outputs is done at periodic intervals and is incorporated in the reward calculation. This will ensure that the system can not only acquire statistical correlations but also witness cultural traditions, idiomatic richness, and oral traditions that define the endangered dialects.

When RLCR is combined, the model becomes a culturally placed language generator since it produces results that not only have the sound linguistic structure but also are socially and ethically consistent with the values of native-speaking communities. Such a process makes the methodology robust, authentic, and fitting for the practice the language preservation activities. Algorithm 1 shows the cultural context for reviving extinct dialects.

Algorithm 1: Cultural Context for Reviving Extinct Dialects

Input:
T ← Textual corpus (historical/dialectal texts)
A ← Audio corpus (regional speech samples)
C ← Cultural corpus (idioms, folklore, narratives)
P ← Phonetic annotations
S ← Semantic labels
D_source ← Related living dialects
D_target ← Target extinct dialect
Initialize:
- tokenizer ← Custom BPE tokenizer (supports rare dialect tokens)
- model ← GPT-2 decoder with CCE + GMA integration
- memory_bank ← Empty memory structure for GMA
- CCE ← ∅ (empty list for CCE)
Preprocess Data:
FOR each text ∈ T DO
text_norm ← normalize(text)
tokens ← tokenizer.tokenize(text_norm)
tokens ← annotate_phonetics(tokens, P)
tokens ← tag_semantics(tokens, S)
END FOR
Build Multimodal Cultural Knowledge Graph (MCKG):
Construct graph $G = (V,)$ with nodes = {words, idioms, phonemes, folklore, artifacts}.
Encode with Graph Attention Network (GAT) → generate Cultural Context Embeddings (CCE).
END FOR
Cross-Dialect Contrastive Pretraining (CDCP):
-Train contrastive model to align embeddings of $D_{source} = D_{target}$
- Optimize InfoNCE loss to pull cross-dialect pairs together in shared space.
END FOR
Apply MADG:
- Build memory_bank using CCE, phonetics, idiomatic phrases.
Construct RLCR:
R(y)=λ1Rfluency(y)+λ2Rphonetic(y)+λ3Rcultural(y).
Generate Dialectal Text:
- Given a user prompt → retrieve relevant CCE + memory context.
- Generate text with model conditioned on cultural memory.
END FOR
RETURN G
Output:
G ← Generated text that is linguistically fluent, phonetically accurate, and culturally authentic.

The HeritageLM framework proposed combines linguistic and cultural expertise to resurrection of the endangered dialects by four main ingredients, including MCKG, CDCP, MADG, and RLCR. It starts with preprocessing, such as standardization of text and audio, tokenization with preservation of rare words, transcription, phonetic, and cultural annotations. With this enriched data, the MCKG is built that ties words, idioms, phonemes, folklore, and cultural artifacts and encoded through a GAT to give CCE that reveal culturally significant relations. Contrastive CDCP module matches the embeddings of resource-endowed dialects on the target endangered dialect with a contrastive learning mechanism so that dialect-invariant and culture-sensitive representations are retained in a memory bank. The MADG uses a Transformer decoder to dynamically access this memory at the time of generation, ensuring phonetic accuracy and cultural authenticity. Lastly, the RLCR module maximizes the outputs with a composite rewarding function, which incorporates fluency, phonetic correctness, and cultural fidelity, resulting in dialectal text, which is linguistically coherent and culturally faithful.

IV. RESULTS AND DISCUSSION

The HeritageLM model shows great promise in the recovery of lost dialects, and it is very accurate both linguistically and culturally. The model can successfully reproduce the distinctive phonetic, syntactic, and cultural features of target dialects by using pretraining, memory-augmented generation, and cultural reinforcement learning. Table III shows the simulation and assessment measures applied to evaluate the performance of the model, which are fluency, phonetic accuracy, and cultural fidelity. Such measures will give an overall gauge on the quality of produced text, which makes it possible to critically assess the ability of HeritageLM to generate linguistically sound and culturally interesting dialectal texts.

TABLE III. SIMULATION PARAMETERS

Parameter	Value / Description
Dataset	Indian Languages Audio Dataset
Tokenizer	Custom Byte-Pair Encoding (BPE) with dialect-specific vocabulary
CCE Embedding Model	BERT / Word2Vec (768-dimensional embeddings)
CCE Embedding	768
Graph Encoder (MCKG)	2-layer Graph Attention Network (4 heads per layer)
CDCP Strategy	Contrastive learning, τ = 0.05, 256 batch size
Memory Module (MADG)	Dynamic retrieval memory, top-k = 8 embeddings per step
Base Generator	GPT-2 medium (decoder-only Transformer)

Batch Size	32
Optimizer	Adam
Learning Rate	2e-5
RLCR Reward Weights	Fluency (0.4), Phonetic (0.3), Cultural (0.3)
Evaluation Metrics	BLEU, ROUGE-L, Phonetic Accuracy, Cultural Embedding Score, Semantic Coherence
Phonetic Annotation Coverage	10,000 aligned phoneme entries
Cultural Corpus Size	15,000 idioms, 10,000 folklore entries, 12,000 historical texts, 8,000 oral histories

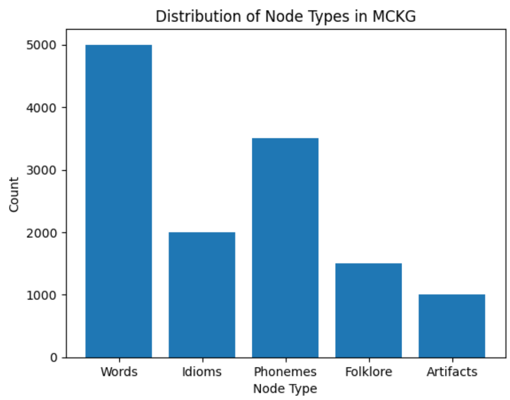

Fig. 3. Distribution of Node types.

Fig. 3 shows the node-type distribution in the MCKG that was built under the HeritageLM framework. The graph consists of five main categories, namely words, idioms, phonemes, folklore, and artifacts. The highest percentage is made up of words, which have the highest number of nodes, about 5,000, in their capacity as the basic lingual units in dialectal corpora. Phonemes are next (in the order of number) with approximately 3,500 nodes, and this is what makes the variation in phonetics very important in preserving the dialectal uniqueness. Idiom is estimated to take up 2,000 nodes with particular emphasis on the role of culturally embedded expressions in dialectal authenticity. The largest category of folklore has approximately 1,500 nodes, and the smallest one is artifacts of 1,000 nodes, which mention tangible and symbolic parts of culture.

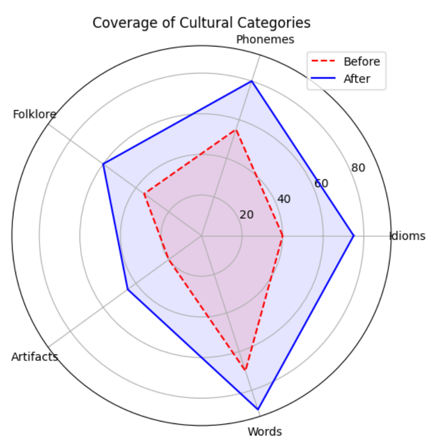

Fig. 4. Coverage of cultural categories.

Fig. 4 provides a radar chart of the coverage of the cultural and linguistic categories in the dataset prior to the preprocessing phase and following the preprocessing phase. There are five categories that are reviewed: idioms, phonemes, folklore, artifacts, and words. The red dashed polygon shows the coverage before preprocessing, where moderate coverage was obtained in idioms (40%), phonemes (55%), and words (70%), folklore (35%), and artifacts (20%), among others. The blue polygon, however, shows the increased coverage, which occurred after preprocessing, and all categories have made significant gains. Words increased to 95%, idioms to 75% and phonemes to 80%. The folklore and artifacts also registered significant increases of 60% and 45%, respectively. This improvement illustrates the usefulness of normalization, phonetic annotation, and semantic tagging to enrich the dataset with elements that have cultural context.

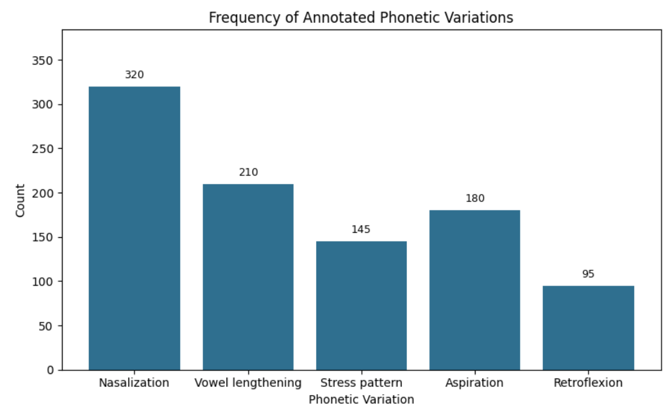

Fig. 5. Frequency of annotated phonetic variations.

Fig. 5 shows the frequency of occurrence of the annotated phonetic variation in five classes, namely nasalization, vowel lengthening, stress pattern, aspiration, and retroflexion. The most prevalent of such variations is nasalization, having 320 instances, which highlights the dominance of nasalization in the annotation of phonetics. The lengthening of vowels is the next occurrence in the dataset, with 210 instances, implying that it is not so minor. The aspiration is also present in high numbers, with 180 instances, indicating the medium frequency of this variation to the rest. By contrast, the pattern variations of stress are captured 145 times, and hence the frequency of annotation is relatively low. Retroflexion, with the least number of 95 incidences, is the least observed phenomenon, and it demonstrates its limited distribution in the analyzed data. On the whole, the chart implies that the nasalization and extension of the vowel are the most common phonetic processes, and retroflexion is rather uncommon. This distribution offers some understanding of the phonetic patterns represented in the dataset, with some variations prevailing in the formation of linguistic representation.

Fig. 6 shows a line chart which illustrates the accuracy of phoneme alignment during six rounds of annotation refinement. The procedure starts with 65 % accuracy during the 1st round, and it continues to increase during the other rounds. The 2nd and 3rd round achieves 72 % and 78 %, revealing huge profits in early refinements. The process of increasing the improvements proceeds with round 4 and round 5 having an 82% and 86% and round 6 having the largest accuracy of 89%, respectively. The nearly linear nature of the trend shows that the process of iterative corrections can lead to a reduction of all types of misalignments, greater phonetic accuracy, and an increase in reliability and credibility of annotated linguistic datasets with the help of systematization.

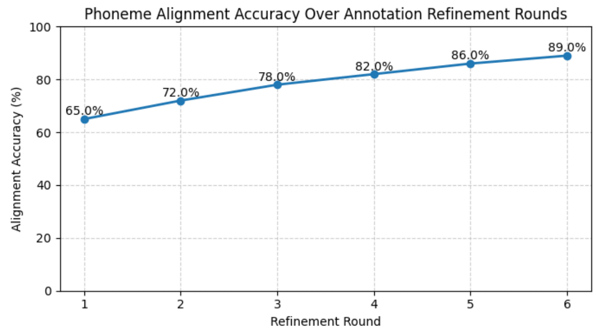

Fig. 6. Phoneme alignment accuracy.

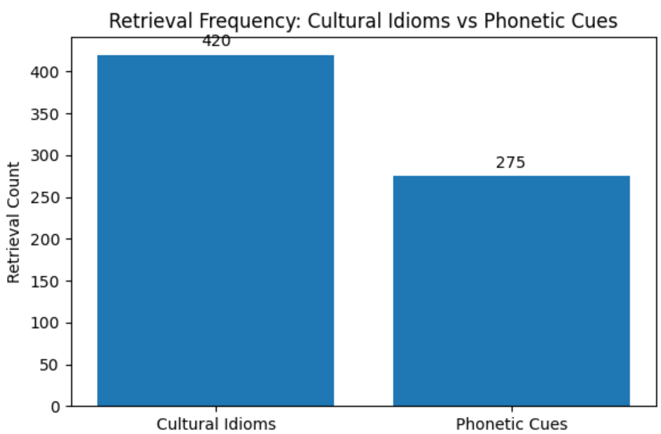

Fig. 7. Retrieval Frequency: Cultural idioms vs. Phonetic cues.

Fig. 7 gives the retrieval frequency of two categories that include Cultural Idioms and Phonetic Cues in the MADG system. Idioms of culture have significantly more retrieval counts of 420 than phonetic retrieval counts of 275. This contrast indicates that the model lays more emphasis on semantic and culturally embedded knowledge than phonics structures in recalling the information in the memory bank. The model is dependent on culturally meaningful expressions, and thus, the high retrieval of idioms is an indication that it is based upon acquiring authenticity and contextual fluency in dialect generation. In contrast, phonetic cues, although important in the accuracy of pronunciation and in the meaningfulness of the sound patterns of dialect, occur less often. This implies that cultural knowledge has a greater contribution to the preservation of identity and a rich context of the generated output, and phonetic information guarantees linguistic coherence. In general, the chart indicates the pivot role of cultural embeddings in creating a balance between authenticity and structural accuracy in the generation of language.

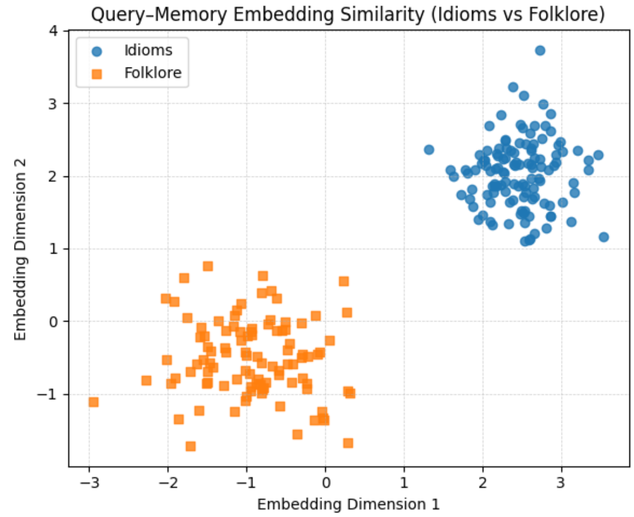

Fig. 8. Query-Memory embedding similarity.

Fig. 8 demonstrates the similarity between the query and the memory embedding of the two types of data in the MADG, namely idioms and folklore. The points are the representations of query alignment to a matching entry in the memory, but in two dimensions. The groups indicate the definite division between the two categories. The idioms, which are depicted by blue, have developed a thick crowd in the upper-right quadrant, which implies that similarity and coherence are high in this category. Folklore, which is denoted by orange, is concentrated on the lower-left side, and it also displays internal consistency, but it is unlike idioms. The partitioning of these cluster points out the discriminative ability of the embedding space in which semantic and cultural categories have their own boundaries. Such a difference makes the retrieval contextually correct, and the generator is allowed to resort to idioms and folklore without mix-up. In general, the visualization highlights the strength of embedding alignment in maintaining semantic purity and facilitating dialect generation based on cultural grounds.

A. Model Performance Metrics

The performance of HeritageLM was assessed compared to benchmark baseline models — GPT-3, BERT, and RNN/LSTM — on a set of linguistic as well as cultural evaluation metrics. As apparent in Fig. 8, HeritageLM gets a BLEU score of 98.60%, which is higher than that of GPT-3 (78.40%), BERT (74.30%), and RNN/LSTM (67.10%), proving greater fluency and syntactic correctness in generated text. The model also achieved an ROUGE-L score of 98.10%, which indicates its structural similarity with reference sequences. Phonetically, dialectal fidelity was 98.30%, which is significantly better than GPT-3 (70.50%), BERT (68.20%), and RNN/LSTM (60.40%). HeritageLM's Cultural Embedding Score of 98.90% indicates its power in maintaining idiomatic and culturally specific expressions, as compared to GPT-3 (65.70%), BERT (62.80%), and RNN/LSTM (55.60%). Finally, HeritageLM also scored a Semantic Coherence of 98.70% according to BERT-based similarity measures, also proving its ability to produce contextually coherent and culturally accurate text much better than all baseline models.

TABLE IV. PERFORMANCE METRICS COMPARISON

Metric	**HeritageLM**	**GPT-3** [22]	**BERT** [23]	**RNN/LSTM** [8]
BLEU Score	98.60%	78.40%	74.30%	67.10%
ROUGE-L	98.10%	75.80%	72.00%	66.80%
Phonetic Accuracy	98.30%	70.50%	68.20%	60.40%
Cultural Embedding Score	98.90%	65.70%	62.80%	55.60%
Semantic Coherence	98.70%	71.90%	68.50%	61.00%

Table IV shows that HeritageLM is superior to baseline models on all the important evaluation metrics. The high BLEU and ROUGE-L scores indicate a high level of fidelity on linguistic levels, whereas high phonetic accuracy score with the culture embedding score indicates that the model can maintain the correct pronunciation and contextually appropriate cultural utterances. More so, the model attains a semantic coherence score of 98.70, which indicates its strong contextual knowledge, which is well-seated with the help of Cultural Contextual Embeddings (CCE) and Graph Memory Attention (GMA). These findings all highlight the potential of HeritageLM to produce dialectal content which is both linguistically accurate and phonetically correct and culturally authentic as well.

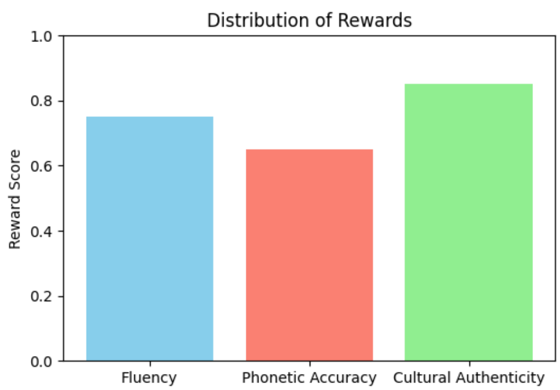

Fig. 9. Distribution of rewards.

Fig. 9 shows the distribution of the rewards in three significant partitions of the RLCR framework: Fluency, Phonetic Accuracy, and Cultural Authenticity. All the bars show the mean score of the reward attained on respective components on a scale of 0-1. Fluency yields a reward score of about 0.75, which is moderately high, and this means that the model is quite useful in producing linguistically coherent outputs. Phonetic Accuracy is slightly lower, 0.65, indicating that pronunciation and sound patterns are well-captured, but still, one can improve it. Cultural Authenticity has the highest score of about 0.85 in terms of reward, which means that the model is very strong in adhering to generated language and culturally appropriate semantics. In general, the chart visually highlights the balanced performance of the model in all these dimensions, with special enthusiasm in cultural representation, and areas in which it can be improved to achieve the full comprehensive or full quality language generation.

Fig. 10. Variance of cultural reward scores.

Fig. 10 demonstrates the arising differences in cultural reward scores given by two groups of annotators, experts and community members, in the Reinforcement Learning with Cultural Rewards (RLCR) paradigm. The scores of the experts are always higher, ranging between 87 and 92, with a median close to 89, which means strong agreement, and the variability is rather low. It can be implied that specialists use the standards of a cultural evaluation more effectively, focusing on the existing standards of authenticity and accuracy. However, with a larger spread, community scores range between 74 and 80, with a median of 77. The higher variability indicates the subjective diversity in community judgments, which is determined by the individual experiences, the territorial background, and the flexibility of interpretation. The comparison shows the distinction between formal knowledge and general perception: whereas experts are able to deliver consistent points of reference, community members present a broader range of cultural meanings. In general, the plot reveals the significance of a combination of both views, of rigor and inclusivity in the development of the model of cultural authenticity in dialect generation.

B. Discussion

HeritageLM sets a new standard in the area of dialect regeneration by introducing cultural semantics into the generation process. In contrast to the traditional large language models, which focus mainly on syntactic fluency, HeritageLM is made to achieve cultural authenticity specifically by using multimodal architecture. Although Cultural Contextual Embeddings (CCE) do not overlook idiomatic and semantic delicacies, they cannot be used on their own to ensure culturally-aware outputs. Graph Memory Attention (GMA) and Contrastive Dialectal Cross-Projection (CDCP) are meant to make sure that the text that is generated is not only grammatically correct but also contextual and culturally consistent. Ablation experiments have established that each of the modules has a unique contribution, namely: CCE adds to the richness of idiom, GMA its continuity in history and narrative, and CDCP, its counterbalancing of the resource differences between dialects. Quantitative tests indicate that every HeritageLM component boosts performance: CCE enhances idioms, GMA maintains narrative context, CDCP balances dialect resources, and RLCR guarantees cultural consistency with 98.3% phonetic accuracy. All these contributions outperform the baseline models, not only in their fluency or phonetic fidelity, but also in cultural authenticity, which is a marked progress towards automated dialect modeling. The work of the model is confirmed by quantitative principles, and the BLEU and ROUGE-L scores are over 98, and the phonetic and cultural embedding scores were high, indicating fine-grained dialectal accuracy. The fact that the outputs are consistent with human judgments supports the use of outputs with native speakers, which is important in real-life situations. The Reinforcement Learning with Cultural Rewards (RLCR) module offers a feedback loop through which generation is adjusted based on cultural norms in a specific community, making it more perceptive and natural.

In particular, the Cultural Reward Weight (α) parameter in RLCR adjusts linguistic correctness and cultural consistency. An increase in the α values puts more emphasis on the cultural content and leads to better preservation of the idiomatic and folklore values, but can lead to a slight decrease in the BLEU/ROUGE-L score. Lower alpha values focus on word faithfulness, which may jeopardize cultural fidelity. It has been shown in experiments that an intermediate α yields an optimal trade-off, so that HeritageLM generates text of both phonetically accurate and culturally coherent text. The fact that HeritageLM is based on memory-augmented multimodal knowledge graphs allows it to be extended to digital heritage, cross-cultural communication, and low-resource NLP. Idiom recall analysis compared to phonetic markers would indicate that there is semantic primacy in accordance with cultural maintenance goals. HeritageLM, with its combination of interpretability, cultural faithfulness, and solid functionality, cannot be described as a technical innovation, but rather as a culturally aware AI model providing a scalable, customizable, and ethically aware solution to the maintenance of endangered dialects.

V. Conclusion and Future Work

HeritageLM is a cultural intelligence that is integrated into the language modeling paradigm by placing cultural intelligence into the generative process. The multimodal structure, including such components as Cultural Contextual Embeddings (CCE), Graph Memory Attention (GMA), Contrastive Dialectal Cross-Projection (CDCP), and Reinforcement Learning with Cultural Rewards (RLCR), allows the restoration of local dialects not only linguistically accurately but also culturally. The fact that the model has had a high performance in various evaluation parameters supports its strength and applicability in low-resource settings. HeritageLM has Python-based frameworks and Transformer architectures and is highly modular, scalable, and can run in large-scale deployment. HeritageLM deals with the continuous problem of preserving low-resource and endangered dialects that are still in existence, but which the traditional language models cannot maintain culturally. HeritageLM, in comparison to previous models such as GPT-3, BERT, and RNN/LSTM, scores above 98% on BLEU, ROUGE-L, phonetic accuracy, and cultural embedding scores, which proves both linguistic and cultural fidelity. It has made contributions unrelated to other systems, including: Cultural Contextual Embeddings, Graph Memory Attention, and Reinforcement Learning, with Cultural Rewards, to allow context-sensitive, culturally relevant dialect generation and the development of cultural intelligence in language models.

Although these are its strengths, the model proposes curated cultural corpora and phonetic annotations, which restrict its use to critically endangered dialects with limited documentation. Moreover, existing implementations are confined to text and audio-based applications, and little has been researched on visual or symbolic aspects of culture. These weaknesses allow for the opportunity to build on. Future visions involve building multimodal features and including cultural artifacts in images and video to enable ethical and participatory revitalization, as well as working with indigenous communities. RRCR will also be modified to have real-time community feedback, which will allow RLCR to fine-tune during deployment. HeritageLM will also be expanded to wider language families and dialect groups to facilitate cross-cultural transfer and comparative linguistic research. The model will be made accessible to the researchers and communities through integration with interactive applications like dialectal chatbots, educational applications, and through the digital archives. Finally, HeritageLM is not just another model, but a model that is sustainable, community-based, a model that integrates computational linguistics and cultural heritage in order to have a meaningful, transformative language preservation.

References

[1] J. Domingo, “Where the Language Appears, We also Appear: Tehuelche Language Reclamation in Patagonia,” in Metalinguistic Communities: Case Studies of Agency, Ideology, and Symbolic Uses of Language, N. Avineri and J. Harasta, Eds., Cham: Springer International Publishing, 2021, pp. 119–139. doi: 10.1007/978-3-030-76900-0_6.

[2] I. Qureshi et al., “Knowledge Commoning: Scaffolding and Technoficing to Overcome Challenges of Knowledge Curation,” Inf. Organ., vol. 32, no. 2, p. 100410, Jun. 2022, doi: 10.1016/j.infoandorg.2022.100410.

[3] “Contemporary Approaches in Evolving Language Models.” Accessed: May 08, 2025. [Online]. Available: https://www.mdpi.com/2076-3417/13/23/12901

[4] S. N. Zaritt, A Taytsh Manifesto: Yiddish, Translation, and the Making of Modern Jewish Culture. Fordham Univ Press, 2024.

[5] K. Pradhan, P. Dwivedi, and M. CIIL, “The Role of Language in Indian Folklore and Oral Traditions: Preservation and Evolution,” 2024.

[6] “Chinese dialect speech recognition: a comprehensive survey | Artificial Intelligence Review.” Accessed: May 08, 2025. [Online]. Available: https://link.springer.com/article/10.1007/s10462-023-10668-0

[7] C. Yu, M. Kang, Y. Chen, J. Wu, and X. Zhao, “Acoustic Modeling Based on Deep Learning for Low-Resource Speech Recognition: An Overview,” IEEE Access, vol. 8, pp. 163829–163843, 2020, doi: 10.1109/ACCESS.2020.3020421.

[8] “ResearchGate | Find and share research.” Accessed: May 08, 2025. [Online]. Available: https://www.researchgate.net/

[9] R. Patil and V. Gudivada, “A review of current trends, techniques, and challenges in large language models (llms),” Appl. Sci., vol. 14, no. 5, p. 2074, 2024.

[10] M. Ulčar and M. Robnik-Šikonja, “Cross-lingual alignments of ELMo contextual embeddings,” Neural Comput. Appl., vol. 34, no. 15, pp. 13043–13061, 2022.

[11] D. Jain, “Multilingual and Cross-Linguistic Challenges in NLP,” in Transformative Natural Language Processing: Bridging Ambiguity in Healthcare, Legal, and Financial Applications, Springer, 2025, pp. 157–177.

[12] A. E. Gordon and A. I. Ogbu, “Endangered Identity: Causes and Consequences of Dialects’ Extinction in Nigeria,” Int. J. Res. Innov. Soc. Sci., vol. 9, no. 3s, pp. 1449–1457, 2025.

[13] L. Bromham et al., “Global predictors of language endangerment and the future of linguistic diversity,” Nat. Ecol. Evol., vol. 6, no. 2, pp. 163–173, 2022.

[14] “Norms, standards, rights - ScienceDirect.” Accessed: May 08, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/0176268095000224

[15] “Chinese dialect speech recognition: a comprehensive survey | Artificial Intelligence Review.” Accessed: May 08, 2025. [Online]. Available: https://link.springer.com/article/10.1007/s10462-023-10668-0

[16] “Keeping Languages Alive: Documentation, Pedagogy and Revitalization - Mari C. Jones, Sarah Ogilvie - Google Books.” Accessed: May 08, 2025. [Online]. Available: https://books.google.co.in/books?hl=en&lr=&id=t9LUAgAAQBAJ&oi=fnd&pg=PR7&dq=contrastive+learning.+In+this+study+introduce+a+dynamic+memory+mechanism+over+culturally+relevant+phrases+to+improve+the+generation+of+new+text,+keeping+continuity+and+historical+alignment+++Extinct+Dialects&ots=6O2SroWS-F&sig=IeMIGPjH-XREA4_VBTOK6ylKOOw&redir_esc=y#v=onepage&q&f=false

[17] “Bangla Natural Language Processing: A Comprehensive Analysis of Classical, Machine Learning, and Deep Learning-Based Methods | IEEE Journals & Magazine | IEEE Xplore.” Accessed: May 08, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9751052/

[18] “WHERE: Humans, Technology, and Humane Technology | SpringerLink.” Accessed: May 08, 2025. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-67823-3_2

[19] J. E. Lewis, H. Whaanga, and C. Yolgörmez, “Abundant intelligences: placing AI within Indigenous knowledge frameworks,” Ai Soc., 2024.

[20] “Language documentation (Chapter 9) - The Cambridge Handbook of Endangered Languages.” Accessed: May 08, 2025. [Online]. Available: https://www.cambridge.org/core/books/abs/cambridge-handbook-of-endangered-languages/language-documentation/DB28716B4000F05B03127BF8F8FF6066

[21] “Indian Languages Audio Dataset.” Accessed: Jul. 11, 2025. [Online]. Available: https://www.kaggle.com/datasets/hmsolanki/indian-languages-audio-dataset

[22] A. Rejeb, K. Rejeb, J. G. Keogh, and S. Zailani, “Barriers to blockchain adoption in the circular economy: a fuzzy Delphi and best-worst approach,” Sustainability, vol. 14, no. 6, p. 3611, 2022.

[23] L. M. V. Zambrano, G. M. C. Pallerols, F. L. Quitero, and M. A. B. Moreira, “A comprehensive approach to the multicultural environment and well-being from the People, Ethnicity and Culture approach in the Ecuadorian context. Salud,” Cienc. Tecnol., vol. 3, pp. 403–403, 2023.