Cureus Journal of Computer Science
https://doi.org/10.7759/s44389-026-00167-y

Original Article

Cultural and Linguistic Contextualization in Ugandan English: Training Transformer Models for Enhanced Local Relevance

David Byansi [1,] , Bebwa Isingoma [2] , Joyce Nakatumba-Nabende [1] , Daudi Jjingo [3, 1]

1. *Department of Computer Science, Makerere University, Kampala, UGA*
2. *Faculty of Humanities, University of Bergen, Bergen, NOR*
3. *African Center of Excellence in Bioinformatic and Data-Intensive Sciences, Makerere University, Kampala, UGA*

Received: March 26, 2026 | Review began: April 17, 2026 | Review ended: July 14, 2026 | Published: July 15, 2026

Abstract

Ugandan English (UgE) plays a central role in communication in Uganda, yet standardized English technologies often fail to preserve the culturally grounded lexical and semantic correspondences that UgE speakers rely on. To address this gap, we investigate how transformer models can be fine-tuned for culturally contextualized meaning generation in UgE and which model families best preserve both surface adequacy and semantic faithfulness. We develop a curated, annotated UgE dataset designed to retain UgE's lexical, morphological, and pragmatic properties, and we frame the task as supervised conditional generation using a two-stage span-to-gloss contextualization pipeline: Stage A produces an intermediate context representation, and Stage B generates the target gloss/meaning conditioned on both the processed input and the Stage A output.

We fine-tune and compare translation-style OPUS-MT, encoder-decoder architectures (BART, T5, and mBART), long-context LED, multilingual models, and decoder-only GPT-family baselines. Evaluation combines n-gram and distance/error measures (Bilingual Evaluation Understudy [BLEU], Recall-Oriented Understudy for Gisting Evaluation [ROUGE], Translation Edit Rate, and Word Error Rate) with embedding-based semantic alignment (BERTScore) and a retrieval-inspired metric, RetriEVAL, which integrates lexical overlap with sentence-embedding similarity. Across experiments, translation-style and encoder-decoder models outperform decoder-only GPT variants, indicating that these architectures are better aligned with structured, context-conditioned UgE meaning generation. OPUS-MT (opus-mt-mul-en) achieves the strongest overall surface overlap and the top retrieval fidelity, while LED (allenai_led-base-16384) delivers the most consistent meaning-focused performance in Stage B, with higher BLEU/ROUGE scores, lower Translation Edit Rate, Word Error Rate, and edit distance, and stronger semantic-faithfulness signals via BERTScore and RetriEVAL.

These findings suggest that culturally grounded NLG for UgE requires both variety-specific data and evaluation beyond n-gram overlap, motivating embedding-aware retrieval metrics such as RetriEVAL for reliable model selection and auditing.

Categories: Natural Language Processing (NLP), AI applications, Big Data and Analytics

Keywords: transformer models, ugandan english, cultural contextualization, local relevance, natural language processing, transfer learning

How to cite this article:
Byansi D, Isingoma B, Nakatumba-Nabende J, et al. (July 15, 2026) Cultural and Linguistic Contextualization in Ugandan English: Training Transformer Models for Enhanced Local Relevance. Cureus J Comput Sci 3 : es44389-026-00167-y. DOI https://doi.org/10.7759/s44389-026-00167-y

Introduction

Globalization of the English language [1] has significantly transformed communication practices worldwide, particularly in multilingual societies such as Uganda [2]. In Uganda, English serves as the official language [3] and the primary medium of instruction [4,5], reflecting its historical significance as a legacy of colonial rule. However, the effectiveness of English in Uganda is often limited by the disconnection between standardized English forms and the local linguistic and cultural realities of its diverse population [6]. This disparity has led to the emergence of Ugandan English (UgE), a variety that reflects the country's unique sociolinguistic landscape [7].

UgE has developed within Uganda's complex multilingual environment, shaped by the influence of various indigenous languages, most notably Luganda [8]. This variety is not merely a variant of British English; it incorporates distinct phonological, lexical, and grammatical features that reflect local communication styles and cultural contexts [9]. For example, pronunciation in UgE often differs from Standard English, with words such as "*hurt*" pronounced as /hat/, as opposed to Standard English /hɜːt/, demonstrating the phonetic influence of Luganda and other local languages [10]. Such regional pronunciation norms contribute to the unique soundscape of UgE and underscore the profound impact of local languages on phonology [11].

Lexically, UgE is characterized by rich creativity, with new words and phrases [12] frequently emerging to reflect local cultural contexts. For example, the expression "*to eat money*," meaning "*to embezzle*," is derived from the Luganda phrase "*kulya sente*," illustrating the intertwining of language and culture [13]. Additionally, peculiarities in UgE, such as the non-standard use of terms like "*avail*" to mean "*provide*," reveal a broader semantic scope compared to Standard English [10].

Syntactically, UgE shows a significant departure from the Standard English norms. A common feature is the use of continuous tenses in situations where simple forms are typically expected, such as in phrases like "*He is not understanding English*." This reflects the adaptation influenced by local grammatical structures [13]. Moreover, constructions such as "Have you done?" exhibit direct object dropping, demonstrating how local language patterns manifest in UgE [13]. Despite these vibrant linguistic characteristics, UgE faces several challenges. Sociolinguistically, attitudes towards English in Uganda are ambivalent; while English is often seen as a vital tool for education and social mobility, it is also regarded as a remnant of colonialism. Many Ugandans prefer to learn their native language because it is believed to provide cognitive benefits and enhance their understanding [7]. Isingoma [13] notes that Ugandans frequently use UgE expressions unconsciously, indicating a lack of awareness of their divergence from Standard English. This ambivalence complicates the implementation of instructional policies that promote local languages.

The Government of Uganda has implemented language policies encouraging the use of local languages in primary education [14-16]. However, the effectiveness of these policies is hindered by English dominance in many classrooms. This gap between policy and practice poses significant challenges for students, affecting their linguistic proficiency and ability to navigate both UgE and Standard English [17,18].

Transformer models have emerged as powerful tools in natural language processing (NLP), fundamentally changing the way language is understood and generated [19]. These models capture contextual nuances using attention mechanisms [20], making them particularly suitable for tasks involving linguistic variations, such as UgE. Training transformer models specifically on UgE data enables deeper integration of UgE's unique linguistic features and cultural contexts.

A key advantage of transformer models is their ability to learn from large amounts of data and to identify intricate patterns and relationships that human analysts might overlook [21]. Focusing on UgE-specific texts, these models can discern the phonological, lexical, morphological, and syntactic characteristics that distinguish UgE [13,22]. For example, they can learn to recognize and predict non-standard phonetic representations common in UgE, thereby enhancing their ability to process spoken language accurately.

Despite these potential advantages, current transformer models exhibit significant weaknesses when applied to UgE [21]. Existing models are often trained primarily on standardized English data, which may not be accurately adapted to the unique phonological and syntactic features of UgE, resulting in inaccurate understanding and generation [23]. Furthermore, their limited exposure to local expressions creates challenges in adequately grasping cultural contexts, which are crucial for effective communication [24].

To address these shortcomings, it is essential to improve the training protocols for transformer models in UgE. This may involve curating a comprehensive dataset that captures the diverse linguistic features of UgE, including local expressions and syntactic constructions.

In addition, integrating sociolinguistic insights enhances the contextual awareness of models, enabling them to respond more appropriately to local cultural nuances. Employing fine-tuning techniques can ensure continuous improvement in model accuracy, making them more relevant and effective for Ugandan users.

Furthermore, sociolinguistic attitudes towards language in Uganda highlight the need for models that not only comprehend linguistic variations but also respect cultural sentiments. Transformer models trained on UgE data can support the development of more inclusive language technologies, ensuring that speakers feel recognized and understood.

The practical implications of training these transformer models for UgE extend to applications such as sentiment analysis, machine translation, and speech recognition. Recent advances in natural language generation (NLG), as demonstrated by pretrained language models such as GPT-2, highlight the potential to create tools that engage users in a culturally respectful manner [25]. By equipping these models with an appropriate linguistic context, we can significantly improve the performance, relevance, and inclusivity of the communication technology in Uganda.

Integrating cultural and linguistic elements into transformer models can enhance meaningful engagement with the UgE variety. This study aims to explore these dimensions, contributing to the development of a more inclusive and representative linguistic environment that respects the rich tapestry of Uganda's linguistic heritage.

Prior work on transformer-based NLP for low-resource settings often reports improvements in lexical or translation quality, but it rarely addresses culturally grounded explanation generation for English varieties such as UgE, nor does it evaluate output quality with metrics that jointly capture surface adequacy and semantic retrieval fidelity. This study fills that gap by (i) curating an UgE corpus annotated to preserve culturally meaningful lexical/pragmatic correspondences and (ii) comparing transformer families under a span-to-gloss contextualization formulation with an embedding-aware retrieval metric (RetriEVAL). Our results indicate that translation-style and encoder-decoder architectures transfer more effectively to this culturally contextualized generation task than decoder-only GPT-style models, suggesting design choices for future culturally sensitive NLG pipelines for underrepresented language varieties.

Despite the widespread use of English in Uganda, communication often fails to reflect the cultural and linguistic reality of the general population. Standardized forms of English can be perceived as alienating, leading to misunderstanding and reduced engagement. Although transformer models show promise for addressing the linguistic features of UgE, several weaknesses limit their effectiveness. Existing models trained primarily on English data may not adequately recognize or generate the unique syntactic characteristics of UgE, resulting in text generation inaccuracies.

Additionally, their understanding of the cultural context and localized expressions may be insufficient, leading to misinterpretations or a lack of relevance in the generated content. Furthermore, these models often require extensive high-quality datasets to perform well; however, such datasets for UgE may be limited, affecting their ability to generalize effectively. There is a pressing need for technological tools that accurately capture and reflect the nuances of the UgE variety, thereby bridging the gap between global communication and local linguistic practices. This study aimed to address these challenges by developing transformer models that produce culturally and linguistically contextualized responses, ensuring that communication is both effective and relevant to the Ugandan context. This involves improving training protocols to incorporate diverse linguistic data, enhance contextual awareness, and integrate sociolinguistic insights that reflect the Ugandan cultural context.

This study aims to develop and evaluate transformer models that generate culturally and linguistically contextualized responses in UgE, thereby enhancing the relevance and effectiveness of communication across different sectors in Uganda. This was achieved by (1) curating and preprocessing a diverse dataset reflecting the linguistic features and cultural nuances of UgE in various contexts; (2) training transformer models using highly processed data to improve their ability to generate contextually and linguistically relevant responses for the general public; and (3) evaluating the performance of the trained models using established metrics such as BLEU and ROUGE scores, focusing on translation accuracy and cultural and linguistic relevance. This study addresses the following research questions: (1) What are the key linguistic features and cultural references that characterize UgE for the general public? (2) How can transformer models be effectively trained to generate responses that reflect cultural and linguistic nuances? (3) How does the performance of the trained models compare with that of standard models?

Related work

This literature review offers a theoretical and empirical foundation for this study. It synthesizes work on (i) the importance of culturally grounded NLP in real communities, especially under low-resource conditions, (ii) the linguistic and pragmatic specificity of UgE as a patterned variety rather than a deviation, (iii) the strengths and limitations of Transformer models in dialect-sensitive and data-poor settings, and (iv) the design of model adaptation and evaluation to improve locally intended meaning.

African English Varieties and Low-Resource Culturally Aware NLP

A growing body of work argues that NLP quality in real-world communities depends not only on surface fluency or benchmark accuracy, but also on whether system outputs align with users' sociocultural norms, communicative expectations, and interpretive frames - an issue that is especially pronounced in low-resource settings, where failures result from both data scarcity and insufficient community-specific grounding for meaning and pragmatics [2,26]. Building on this, Culturally Aware NLP (CAENLP) frames NLP outputs as sociotechnical artifacts that can become misaligned with users' "moral-semiotic" contexts [27], but emphasizes that cultural awareness must be operationalized through evaluation approaches that verify local interpretability while avoiding stereotyping or harmful generalizations [28,29]. This motivates treating evaluation as part of the modeling pipeline: metrics should assess culturally relevant meaning capture and interpretive alignment rather than relying solely on surface-form resemblance.

Related work in machine translation and controlled generation suggests that incorporating culturally grounded information during generation can improve outputs for low-resource languages [30]. However, much evaluation still relies on general-purpose benchmarks, highlighting the need for variety- and community-sensitive adequacy assessments that focus on interpretive alignment rather than proxy metrics. Another approach uses culturally situated training stimuli. For example, NileChat's Language-Heritage-Values framework operationalizes cultural contextualization through controlled synthetic generation conditioned on local personas, heritage concepts, and dialectal cues, along with retrieval-augmented inclusion of culturally specific content [31]. This supports the implication that local relevance for African Englishes (including UgE) requires data reflecting locally meaningful scenarios and norms, not just surface forms.

This literature also highlights that cultural bias can result from the dominance of English-majority pretraining corpora, and mitigation strategies generally fall into (i) inference-time prompt interventions and (ii) culturally aware pretraining or fine-tuning [32]. In low-resource settings, prompt-based methods may fail when cultural knowledge is missing from pretraining, while fine-tuning requires large, culturally representative datasets and significant computational resources. As an alternative, related work proposes cultural adaptation using small seed cultural signals combined with semantic augmentation that preserves underlying "opinions/values" [33], suggesting a feasible pathway for UgE localization with limited curated supervision.

Empirical evidence from African language NLP initiatives shows that "language scarcity" involves not only data volume but also sociolinguistic representation and community involvement. For example, Masakhane's multi-agent pipeline [34] demonstrates that excluding language-community expertise can produce unreliable datasets and misleading quality estimates when generic automatic metrics dominate [34]. Quality-focused data curation is also essential: MC2 reports that web corpora may contain language misidentification and inadequate cleaning, and that representational choices, such as writing system fidelity, can affect downstream learning [35]. Therefore, UgE localization should prioritize (i) quality-aware data curation, (ii) community-informed decisions for dialect and domain coverage, and (iii) culturally grounded evaluation beyond generic metric proxies.

The low-resource African-English benchmarking literature critiques the assumption that mainstream benchmarks are universally "fit." Research argues that evaluation data often reflect the goals of higher-resource communities rather than those of the target language community, which can propagate learner-like grammatical, orthographic, and pronunciation errors - harmful when communities rely on these benchmarks for system development [36]. Complementary work frames benchmarking as language-community-centered, emphasizing speaker diversity (age, gender, proficiency, regional and vernacular variation) and culturally salient targets, including sub-benchmarks aligned with community aspirations and code-switching behavior [37]. This lesson applies directly to UgE: evaluation should include dialect- and context-sensitive sub-assessments reflecting local conversational norms rather than relying solely on aggregated scores.

Taken together, the literature highlights an imperative: high-quality NLP must go beyond linguistic accuracy to include cultural and ethical alignment. CAENLP refers to this as a "moral-semiotic" mismatch problem [38], and related MT research indicates that integrating cultural context can improve translation quality in low-resource languages [39]. Other studies emphasize that systems can perpetuate biases related to race and culture, reinforcing the need for culturally sensitive frameworks [40,41]. However, there are still gaps in standardized measures for "cultural fairness" or "moral sufficiency," especially for systems designed for specific cultural contexts such as Uganda.

UgE: Linguistic Specificity and Culturally Grounded Evaluation Targets

UgE is shaped by local contact ecologies, including influences from indigenous languages such as Luganda [30], and displays distinctive phonological, morphological, and lexical features compared to standard varieties [9]. These differences involve not only surface realization but also pragmatic norms and meaning composition: idioms, culturally grounded references, and locally preferred discourse strategies all influence what is considered appropriate output for Ugandan audiences. In current NLP practice, a common limitation is the implicit mapping of nonstandard varieties to an "incorrect" norm, creating extrinsic bias when models fail to treat dialect-specific patterns as valid linguistic structures [42]. Addressing this requires both targeted UgE training data and evaluation metrics that reward the adequacy of locally intended meaning.

To support this goal, related paradigms from style transfer and meaning glossing suggest that surface rewriting can be separated from semantic preservation. This informs UgE contextualization: instead of enforcing standardization, the target can be framed as a meaning-preserving transformation that makes culturally grounded intent explicit through an intermediate representation. In this context, meaning glossing and paraphrasing motivate training objectives where the target is not "correct English" in a purely standard sense, but a locally faithful representation that downstream users can trust for comprehension, enabling evaluation of dialect-context adequacy beyond n-gram overlap.

Transformers, Adaptation Strategies, and Dialect-Sensitive Fairness Through Meaning-Grounded Evaluation

Modern Transformer models (e.g., BERT, T5, GPT variants) dominate NLP, but their effectiveness in dialect-sensitive, low-resource settings depends on the alignment between the pretraining distribution and the fine-tuning data. Although multilingual pretrained models (e.g., mT5 and XLM-R) are often used as baselines or transfer starting points for African and low-resource languages [43], and there is evidence of adaptability for Luganda translation [44,45], variety-specific weaknesses persist. Evidence shows that even multilingual systems can underperform in low-resource African languages compared to high-resource counterparts on tasks such as sentiment analysis [46,47]. Related low-resource studies also caution that the claim "Transformer is always better" is context-dependent, and that well-regularized alternatives can outperform Transformers when data are scarce [48]. For UgE contextualization, this implies careful architecture selection and training data design rather than assuming general transfer will ensure dialect adequacy.

Within this adaptation landscape, fine-tuning is widely regarded as an effective specialization method for pretrained models in low-resource settings, often using small amounts of monolingual text [49]. In East African contexts, fine-tuning has shown promise for translation and other transformations [50]. However, success depends on data quality: culturally grounded corpora must represent locally meaningful expressions rather than transliterations of dominant norms. To address scarcity, the literature increasingly combines fine-tuning with synthetic data generation and targeted augmentation when real corpora are limited [46]. Because synthetic expansion can introduce bias if culturally mediated meaning is not preserved, adaptation should be paired with evaluation methods that test whether the locally intended meaning survives transformation [51].

Beyond translation, culturally grounded NLG and controlled generation research emphasize controllability to produce outputs that meet user constraints and context-dependent appropriateness [52,53]. In dialect-sensitive applications, fairness and ethics are operational requirements: a model should not systematically penalize dialectal structures or over-correct speakers toward a dominant norm. CAENLP frames this as a moral-semiotic mismatch problem [39], but dialect fairness specifically requires evaluation protocols that measure meaningful adequacy under dialect variation, rather than relying solely on surface similarity [42]. In response to these needs, this paper contributes by treating UgE contextualization as an intermediate meaning transformation and by using adequacy evaluation designed to reflect interpretive alignment rather than lexical overlap.

The generation of culturally contextualized responses depends on the model's ability to represent context dependencies and multilingual signals, and architectures (encoder-decoder, multilingual, long-context) can vary in their capacity to produce culturally appropriate outputs [51]. Adaptation through fine-tuning, in-context learning, and representation-aware evaluation can improve locally relevant responses [54,55]. Fine-tuning on UgE-like data reduces distribution shift and increases coverage of dialectal lexical and pragmatic patterns [56]. In-context adaptation can also incorporate culturally relevant constraints in prompts to guide outputs toward local norms [57,58]. However, evaluation remains challenging: cultural adequacy is not fully captured by automatic lexical-overlap metrics, motivating embedding- or retrieval-informed adequacy measures such as RetriEVAL [59].

Overall, despite growing recognition of culturally and linguistically relevant communication, gaps persist in (i) treating African English varieties as structured targets and (ii) operationalizing culturally grounded adequacy for dialect-sensitive generation under low-resource conditions. This paper addresses these gaps by formulating UgE contextualization as a span-to-gloss generation task and evaluating adequacy using RetriEVAL, thereby targeting meaning grounding rather than surface similarity.

Materials And Methods

The methodology for this study was organized into several key components: data curation, preprocessing, model selection, and evaluation metrics.

Data sources and corpus construction

The curated training data for the transformer models were compiled from multiple sources intended to represent Uganda's linguistic diversity and locally meaningful usage of UgE. The corpus was designed to capture culturally specific expressions and references so that the resulting models would better reflect the language practices of real-world UgE speakers.

To construct the UgE corpus, we followed a rigorous process of data identification, selection, and extraction, drawing primarily on ICE-UG [60] and Web-UG [61]. ICE-UG provides a large, well-structured collection of 960,000 words, compiled as part of the ICE-UG project and completed in 2021 through a collaboration among Makerere University, Gulu University, and Ruhr University Bochum [60]. The corpus covers both spoken and written UgE across a wide range of domains and contexts, including private spoken interactions (e.g., conversations and telephone calls) and public spoken settings (e.g., classroom lessons, parliamentary debates, and business transactions), while also distinguishing between unscripted and scripted speech. It also includes extensive written material from printed sources (such as academic writing and reportage) and non-printed sources (such as student essays and correspondence), thereby ensuring UgE representation across educational, professional, and creative domains.

Selection criteria, expert validation, and linguistic feature inventory

To ensure reliability and relevance, we applied clear selection criteria during corpus construction. First, texts were selected based on their relevance, requiring that the included material accurately represent UgE in its various forms [62,63]. Second, we ensured diversity by incorporating a wide range of genres, contexts, and speaker backgrounds to reflect the complexity of UgE [64]. Finally, we emphasized quality by including only well-documented and carefully transcribed texts. Attention was given to the verification procedures described in [65,66] to preserve the integrity of the analysis [60]. Linguistic experts proficient in both Ugandan English and Standard English supported feature extraction across 15 linguistic categories: Morphological Innovation (MI), Phonological or Orthographic Innovation (PH+OR), Syntactic Innovation (SI), Semantic Extension (StE), Calquing or Semantic Extension (CA+StE), Archaisms (AR), Semantic Extension or Morphological Innovation (StE+MI), Calquing (CA), Phraseological Extension (PE), Calquing or Morphological Innovation (CA+MI), Archaisms or Morphological Innovation (AR+MI), Phraseological Innovation (PI), Borrowing (BR), Semantic Shift (SS), and Borrowing or Morphological Innovation (BR+MI).

Example mapping of UgE lexical/structural adaptations

Table *1* presents representative examples of UgE lexical and structural adaptations, linking each item to its corresponding innovation type. These examples show that UgE differs from Standard English not only at the word level (such as clipping, compounding, conversion, affixation, and borrowing) but also at the phrase and clause levels through changes in construction patterns, including omission, substitution, and reordering. Semantically, the examples demonstrate how meanings may be extended, shifted, or calqued to meet local communicative needs - often retaining Standard English forms while altering their function or interpretation. Furthermore, the presence of formulaic expressions and idioms (such as "Well be back," "We go!," and "to eat money") indicates that innovation often occurs through phraseological conventions rather than isolated word changes.

Ugandan English words	Word class/category	Meaning in Standard English	Examples
Well be back	Phraseological innovation	welcome back	Well be back from Europe!
congs	Morphological innovation (clipping)	congrats	Congs, you have passed well!
taxi	Semantic shift	a van used like a bus, carrying many people	We have to wait until the taxi is full before leaving.
special hire	Morphological innovation (compounding)	taxi/cab	I will get a special hire to reach the airport in time.
bodaboda pilot	Borrowing semantic extension	a motorbike or bicycle used for similar purposes driver of a bus/taxi	I will use a bodaboda to go to town. Tell the pilot not to speed.
to foot	Syntactic innovation (collocative object omission)	to walk; to foot it	I came footing.
to branch	Syntactic innovation (particle omission)	to turn, to branch off	You need to branch to your right.
to give someone a push	Phraseological innovation (idiom)	to accompany a person for some distance	I will give my friend a push since it is late.
downer	Morphological innovation (derivational affixation)	lower	The downer part is not good.
ever	Archaism	always	She is ever late.
just	Calquing	expression of obviousness	I am at home eating food, just.
ka-	Borrowing (derivational affixation)	marker for diminutive	The ka-man is innocent.
We go!	Phraseological innovation (formulaic phrase)	Let's go!	Are you ready? We go.
savedee	Morphological innovation (derivational affixation)	a religiously saved person	We have many savedees in the country.
to eat money	Calquing	to embezzle, misappropriate funds	The Minister ate the money.
cousin brother/sister	Calquing + Morphological innovation (compounding)	cousin	My cousin brother will visit me today.

Irish (potatoes)	Morphological innovation + eponym	potatoes	I will cook Irish potatoes today.
digging	Calquing + Morphological innovation (derivational affixation)	farming	He has resorted to digging as a means of survival.
to go for a short call	Phraseological innovation (idiom)	to urinate	I need to go for a short call.
to slope down	Phraseological innovation (phrase)	to go downhill	You need to slope down a bit before climbing.
to avail	Semantic extension or morphological innovation	to provide	He should avail us with the documents.
CC class	Morphological innovation (initialism)	group of girls liking chicken	She looks a CC class.
to reach	Syntactic innovation (argument dropping)	to arrive	I have reached.
to ashame	Archaism or morphological innovation	to shame	Ashame the devil.
to pick	Syntactic innovation (particle omission)	to pick up	How come she is not picking?
dentist	Calquing + semantic extension	gold digger	Be careful; she is a dentist.
to louse	Morphological innovation (back formation)	to idle	Students have been lousing around.
Me, I...	Syntactic innovation (left dislocation)	way to begin a presentation	Me, I want to go.
pakalast	Borrowing + Morphological innovation	till death	The president will rule pakalast.
safe house	Semantic shift	illegal detention	Many people are in safe houses.
there-there	Calquing + Morphological innovation	mild	A: How are you? B: I am there-there.
stage	Semantic extension	taxi/bus stop	Driver, drop me at the stage.
to run mad to stroke	Archaism morphological innovation (conversion)	to become insane to have an additional subject	I almost ran mad. She has decided to stroke Mathematics with Physics.

How to cite this article:
Byansi D, Isingoma B, Nakatumba-Nabende J, et al. (July 15, 2026) Cultural and Linguistic Contextualization in Ugandan English: Training Transformer Models for Enhanced Local Relevance. Cureus J Comput Sci 3 : es44389-026-00167-y. DOI https://doi.org/10.7759/s44389-026-00167-y

trouser, jean, etc.	Morphological innovation (inflectional affix dropping)	trousers, jeans	Your trouser is dirty.
under the shade/sun/rain	Syntactic innovation (preposition substitution)	in the shade/sun/rain	Place your car under the shade.
to air out	Syntactic innovation (particle addition)	to air	I need to air out my problems.

TABLE 1: Word or phrase category, Ugandan English words, Standard English word meanings, and example usage

Task formulation: Span-to-gloss conditional definition generation

This paper defines cultural and linguistic contextualization in UgE as a supervised conditional text generation task. Given an English sentence with a single marked UgE span, the system generates a short English gloss that explains the locally intended meaning of that span.

Input representation: Each instance is constructed from a sentence in which the UgE span is delimited by special tags. After preprocessing, the tags are removed, preserving the span surface form within the sentence. Let x denote the processed sentence text.

Supervision targets: For each example, two gold textual fields are provided:

(i) Context ($c^\star$): a context representation derived from the tagged UgE region.

(ii) Meaning ($g^\star$): the corresponding gold gloss that captures the intended meaning of the UgE span in that sentence context.

Two-stage conditional generation: The task is implemented as a cascade of two conditional generation steps:

(i) Stage A (context generation): produces a predicted context string $\hat{c}$ from the processed sentence x,

$$c^\star = \mathrm{Context}(x), \qquad \hat{c} \sim p_{\theta_A}(c \mid x).$$

(ii) Stage B (meaning generation): produces a predicted gloss $\hat{g}$ conditioned on both x and a context string (gold context in training, predicted context at inference),

$$g^\star = \mathrm{Meaning}(x, c), \qquad \hat{g} \sim p_{\theta_B}(g \mid x, c).$$

Learning objective: Training maximizes the likelihood of the gold targets in the two-stage factorization:

$$\theta_A^\star = \arg\max_{\theta_A} \sum_{(x,c^\star)} \log p_{\theta_A}(c^\star \mid x), \qquad \theta_B^\star = \arg\max_{\theta_B} \sum_{(x,c,g^\star)} \log p_{\theta_B}(g^\star \mid x, c).$$

Task output: The final prediction is the generated gloss $\hat{g}$, which is compared against $Meaning$ to measure contextualized glossing performance.

Annotated data

The annotated data are central to modeling UgE linguistic features and culturally grounded meanings. Each dataset entry contains a sentence and a marked UgE span, indicated by $\langle \mathrm{UGE} \rangle$ tags (see Table 2). In total, 124,657 sentences were annotated, and for each tagged span, we provide (i) the span-text in context and (ii) an associated brief gloss and cultural/usage explanation. This design supports training of transformer models to generate culturally and linguistically contextualized glosses for UgE spans.

Sentence
I thought I would find her at the meeting, but after $\langle \mathrm{UGE} \rangle$ bouncing $\langle / \mathrm{UGE} \rangle$ around, it became clear she wasn't attending.
She enjoys sharing her knowledge about $\langle \mathrm{UGE} \rangle$ digging $\langle / \mathrm{UGE} \rangle$ with her peers, believing that everyone should understand the basics of farming and food production.
The discussions about $\langle \mathrm{UGE} \rangle$ going to the bush $\langle / \mathrm{UGE} \rangle$ are often met with mixed reactions, highlighting the complexity of armed resistance.
Throughout her life, she has encountered many challenges, but her identity as a $\langle \mathrm{UGE} \rangle$ Mulokole $\langle / \mathrm{UGE} \rangle$ has given her strength to persevere.
She respects the $\langle \mathrm{UGE} \rangle$ Katikkiro $\langle / \mathrm{UGE} \rangle$ for his dedication to the kingdom.
She feels proud of her beautifully designed $\langle \mathrm{UGE} \rangle$ storeyed house $\langle / \mathrm{UGE} \rangle$.
After finishing my coffee, I realized I needed to $\langle \mathrm{UGE} \rangle$ pupu $\langle / \mathrm{UGE} \rangle$, so I quickly made my way to the nearest restroom.
She felt her $\langle \mathrm{UGE} \rangle$ ga- $\langle / \mathrm{UGE} \rangle$ voice was often ignored in discussions.
I appreciate how $\langle \mathrm{UGE} \rangle$ bodaboda $\langle / \mathrm{UGE} \rangle$ riders are often friendly, sharing stories and tips about the best places to visit while on the road.
She plans to continue her education and personal development after prison, inspired by her experiences at the $\langle \mathrm{UGE} \rangle$ University of Understanding $\langle / \mathrm{UGE} \rangle$.
The $\langle \mathrm{UGE} \rangle$ posho $\langle / \mathrm{UGE} \rangle$ vendor at the market sells a variety of flavored cornmeal cakes and loaves.
The $\langle \mathrm{UGE} \rangle$ godown $\langle / \mathrm{UGE} \rangle$ was a treasure trove for vintage finds, offering unique items that told stories and sparked conversations among friends.
She quickly grabbed her belongings and signaled to the driver as they approached the $\langle \mathrm{UGE} \rangle$ stage $\langle / \mathrm{UGE} \rangle$ where she needed to get off.
He was caught attempting to $\langle \mathrm{UGE} \rangle$ eat money $\langle / \mathrm{UGE} \rangle$ from the project funds, which resulted in immediate suspension and investigation.
He believes that the best $\langle \mathrm{UGE} \rangle$ posho $\langle / \mathrm{UGE} \rangle$ is made with love and care, reflecting the traditions of those who prepare it.
If I'm available, I'll $\langle \mathrm{UGE} \rangle$ pick $\langle / \mathrm{UGE} \rangle$ the phone promptly.
He recognizes that the $\langle \mathrm{UGE} \rangle$ mailo $\langle / \mathrm{UGE} \rangle$ system is a crucial aspect of land governance, ensuring that all parties are treated fairly and equitably.
$\langle \mathrm{UGE} \rangle$ Gifted by nature $\langle / \mathrm{UGE} \rangle$, voluptuous figure serves as a testament to the allure of femininity, captivating everyone who crosses her path.

How to cite this article:
Byansi D, Isingoma B, Nakatumba-Nabende J, et al. (July 15, 2026) Cultural and Linguistic Contextualization in Ugandan English: Training Transformer Models for Enhanced Local Relevance. Cureus J Comput Sci 3 : es44389-026-00167-y. DOI https://doi.org/10.7759/s44389-026-00167-y

TABLE 2: Annotated sentences of Ugandan English

Figure *1* shows the distribution of the 15 features analyzed during corpus identification, which formed the basis for data annotation. Borrowed words are the most dominant feature, while phonological or orthographic innovation is the least dominant in the annotated dataset, as they are used least in practice.

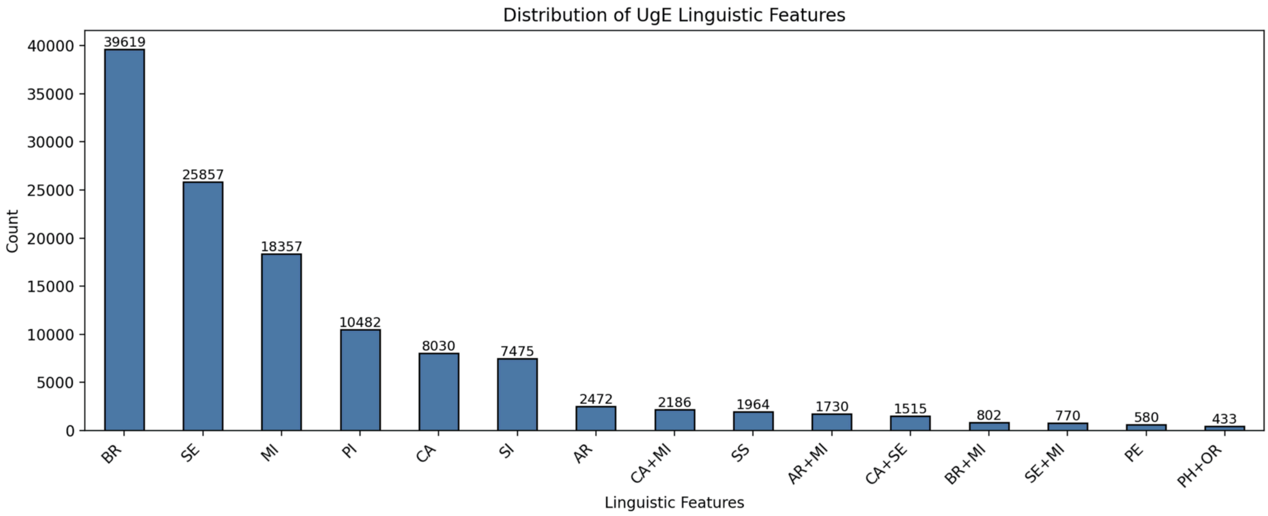

FIGURE 1: Different class distribution of Ugandan English (UgE) words or phrases in UgE sentences

Annotation guidelines and agreement

Annotators were trained using written guidelines that specify when meaning should reflect (i) sense-level glosses vs. (ii) pragmatic category/function vs. (iii) register-relevant meanings. For each UgE span occurrence, annotators provided context (derived from the tagged span) and a short English Meaning gloss. We computed inter-annotator agreement on a shared subset using Cohen's kappa or Krippendorff's alpha; results are reported per feature in Table *3*. Disagreements (if any) were resolved by adjudication following the same guidelines, and we log the adjudication rules in the released dataset card.

Feature	Kappa presence	Support annot1	Support annot2	Raw agreement presence	MeanP exact	MeanR exact	MeanF1 exact	No. of sentences used exact
MI	1	15,926	15,926	1	1	1	1	15,926
PH+OR	1	433	433	1	1	1	1	433
SI	1	5,021	5,021	1	1	1	1	5,021
StE	1	21,688	21,688	1	1	1	1	21,688
CA+StE	1	964	964	1	1	1	1	964
AR	1	1,878	1,878	1	1	1	1	1,878
StE+MI	1	766	766	1	1	1	1	766
CA	1	6,858	6,858	1	1	1	1	6,858
PE	1	429	429	1	1	1	1	429
CA+MI	1	1,992	1,992	1	1	1	1	1,992
AR+MI	1	1,534	1,534	1	1	1	1	1,534
PI	1	9,111	9,111	1	1	1	1	9,111
BR	1	32,633	32,633	1	1	1	1	32,633
SS	1	2,160	2,160	1	1	1	1	2,160
BR+MI	1	421	421	1	1	1	1	421

TABLE 3: Inter-annotator agreement per feature: presence/absence (Cohen's kappa and raw agreement) and exact span match (mean precision, recall, and F1)

Macro Cohen's kappa (mean over features): 1.0000
Micro Cohen's kappa (flattened): 1.0000
Raw agreement across all decisions: 1.0000
MI, Morphological Innovation; PH+OR, Phonological or Orthographic Innovation; SI, Syntactic Innovation; StE, Semantic Extension; CA+StE, Calquing or Semantic Extension; AR, Archaisms; StE+MI, Semantic Extension or Morphological Innovation; CA, Calquing; PE, Phraseological Extension; CA+MI, Calquing or Morphological Innovation; AR+MI, Archaisms or Morphological Innovation; PI, Phraseological Innovation; BR, Borrowing; SS, Semantic Shift; BR+MI, Borrowing or Morphological Innovation

We compare two annotators at two levels. First, we evaluate the presence or absence of each UgE feature per sentence by parsing $\langle \mathrm{UGE} \rangle \cdot \cdots \langle / \mathrm{UGE} \rangle$ spans with a regex for the feature tag, converting each feature into a binary value (1 if present, 0 if absent) for every ID of shared sentences, and then calculating Cohen's kappa (macro and micro). The results in Table 3 show that macro kappa = 1.0000, micro kappa = 1.0000, and raw agreement = 1.0000, indicating that both annotators make identical presence or absence decisions for every feature across all common sentences. Second, we evaluate exact span-text agreement by treating the extracted $\langle \mathrm{UGE} \rangle$... contents as sets of strings and computing precision, recall, and F1 using exact string matches, while skipping sentences where both annotators have no spans for that feature. The results also show meanP = meanR = meanF1 = 1.000 for every feature, indicating that when spans are present, the annotators extract the same span texts with no discrepancies.

Handling polysemy and canonical gloss assignment

A key methodological decision concerns how to construct the gold reference gloss when a UgE span can realize different senses across contexts (polysemy). Our supervision is built in two stages. First, we extract the span string s enclosed by $\langle \mathrm{UGE} \rangle \cdots \langle / \mathrm{UGE} \rangle$, and, when multiple tagged spans occur in a sentence, we either keep the first span consistently or discard the instance when meanings vary completely, using a single rule across the corpus. Second, we assign a gold reference gloss $g^{\star}$ using a token-to-meaning canonicalization rule: for each extracted span s, we create a mapping $s \to g^{\star}$ using the first observed pairing; if the same span later appears with a different meaning, the earlier canonical pairing is retained, producing one canonical gold gloss per span type.

This canonicalization introduces potential label noise when the span is truly context-dependent. Collapsing context-specific senses into a single canonical gloss may penalize the model for producing a correct but non-canonical sense, reducing evaluation reliability for polysemous spans. To quantify this risk, we report polysemy-related statistics (Table 4), including (i) the number of extracted spans (rows) and unique span types, (ii) the proportion of spans with multiple distinct Meaning values, (iii) the span-length distribution (unigram vs. multiword spans), and (iv) the proportion of discarded instances (e.g., empty sentence or meaning, multiple spans per sentence, preprocessing failures). Importantly, these analyses are used to interpret evaluation scores with the explicit caveat that lower scores may reflect canonical-label constraints rather than genuine model failure to recover an acceptable sense.

To address gold-reference ambiguity more directly, we recommend shifting from span-type canonicalization to a context-conditioned, row-level supervision strategy, where each dataset row serves as its own supervision instance, and both training and evaluation use the instance-specific meaning rather than a canonical mapping for each span type. With this row-level protocol, multiple valid senses for the same span type can coexist as distinct targets.

Statistic	Value	Notes
Annotated sentences (initial)	124,657	initial dataset size
Rows kept after empty-removal	124,657	after filtering empties
Instances with multiple spans	4,655	overlapping spans handling
Extracted span instances	128,609	total supervision spans
Unique span types	1,271	vocabulary for mapping
Avg. tokens per span	1.3942	mean span length
Num multiword spans	37,536	spans ≥2 tokens
Num singleword spans	91,073	spans = 1 token
Proportion multiword spans	0.2919	multiword share
Proportion singleword spans	0.7081	singleword share
Num multiword span types	481	distinct multiword types
Num singleword span types	790	distinct singleword types
Proportion multiword span types	0.3784	type share (multiword)
Proportion of span types with ≥2 distinct meanings	0.2234	polysemy rate
Train/Val/Test split counts	97,828/12,229/12,229	dataset partition sizes

TABLE 4: Dataset statistics and filtering rates for Ugandan English gloss generation

Polysemy-aware evaluation

Because polysemy can produce multiple correct glosses for the same input, evaluation during inference uses a polysemy-aware best-match strategy. Instead of scoring only a single decoded output, we generate top-*k* candidates and compute metrics using:

$$\mathrm{Score}(x) = \max_{j\in\{1,...,k\}} \mathrm{Metric}\big(g_j^{\star}(x), \hat{g}_j(x)\big),$$

where $\hat{g}_j(x)$ is the *j*-th predicted candidate gloss for input x. This reduces unfair penalization when the model recovers an alternative sense permitted under top-*k* decoding.

We do not assume that each UgE span has a single universally correct gloss. Instead, we treat meaning as context-dependent: (i) the dataset includes multiple distinct meanings for 22.34% of span types (Table 4), and (ii) evaluation is polysemy-aware by scoring each prediction against all gold meanings observed for the same processed input key and selecting the best match. We also perform polysemy-stratified reporting to show that results hold on both ambiguous and unambiguous subsets.

The dataset was prepared for the next stages of training the culturally and linguistically contextualized response model, thereby enhancing the model's ability to learn from the data's contextual nuances. The dataset is shown in Table 5.

How to cite this article:
Byansi D, Isingoma B, Nakatumba-Nabende J, et al. (July 15, 2026) Cultural and Linguistic Contextualization in Ugandan English: Training Transformer Models for Enhanced Local Relevance. Cureus J Comput Sci 3 : es44389-026-00167-y. DOI https://doi.org/10.7759/s44389-026-00167-y

Sentence	Context	Meaning
I thought I would find her at the meeting, but after $\langle\mathrm{UGE}\rangle$ bouncing $\langle/\mathrm{UGE}\rangle$ around, it became clear she wasn't attending.	bouncing	failing to meet with someone
She enjoys sharing her knowledge about $\langle\mathrm{UGE}\rangle$ digging $\langle/\mathrm{UGE}\rangle$ with her peers, believing that everyone should understand the basics of farming and food production.	digging	farming
The discussions about $\langle\mathrm{UGE}\rangle$ going to the bush $\langle/\mathrm{UGE}\rangle$ are often met with mixed reactions, highlighting the complexity of armed resistance.	going to the bush	to starting or joining an armed rebellion against a government
Throughout her life, she has encountered many challenges, but her identity as a $\langle\mathrm{UGE}\rangle$ Mulokole $\langle/\mathrm{UGE}\rangle$ has given her strength to persevere.	Mulokole	pentecostalist
She respects the $\langle\mathrm{UGE}\rangle$ Katikkiro $\langle/\mathrm{UGE}\rangle$ for his dedication to the kingdom.	Katikkiro	Prime Minister of Buganda Kingdom
She feels proud of her beautifully designed $\langle\mathrm{UGE}\rangle$ storeyed house $\langle/\mathrm{UGE}\rangle$.	storeyed house	house with two or more floors
After finishing my coffee, I realized I needed to $\langle\mathrm{UGE}\rangle$ pupu $\langle/\mathrm{UGE}\rangle$, so I quickly made my way to the nearest restroom.	pupu	to defecate
She felt her $\langle\mathrm{UGE}\rangle$ ga- $\langle/\mathrm{UGE}\rangle$voice was often ignored in discussions.	ga-	marker for augmentative
I appreciate how $\langle\mathrm{UGE}\rangle$ bodaboda $\langle/\mathrm{UGE}\rangle$ riders are often friendly, sharing stories and tips about the best places to visit while on the road.	bodaboda	a motorbike or bicycle used for the same purpose or a rider of such a motorbike or bicycle
She plans to continue her education and personal development after prison, inspired by her experiences at	University of Understanding	prison

the $\langle \mathrm{UGE} \rangle$ University of Understanding $\langle / \mathrm{UGE} \rangle$.		
The $\langle \mathrm{UGE} \rangle$ posho $\langle / \mathrm{UGE} \rangle$ vendor at the market sells a variety of flavored cornmeal cakes and loaves.	posho	cornmeal, maize meal
The $\langle \mathrm{UGE} \rangle$ godown $\langle / \mathrm{UGE} \rangle$ was a treasure trove for vintage finds, offering unique items that told stories and sparked conversations among friends.	godown	a basement
She quickly grabbed her belongings and signaled to the driver as they approached the $\langle \mathrm{UGE} \rangle$ stage $\langle / \mathrm{UGE} \rangle$ where she needed to get off.	stage	taxi stop or bus stop or taxi rank, or request for a stop
He was caught attempting to $\langle \mathrm{UGE} \rangle$ eat money $\langle / \mathrm{UGE} \rangle$ from the project funds, which resulted in immediate suspension and investigation.	eat money	to embezzle, misappropriate funds
I caught the $\langle \mathrm{UGE} \rangle$ ka- $\langle / \mathrm{UGE} \rangle$ rat rummaging through my trash for recyclables.	ka-	marker for diminutive
She asked me $\langle \mathrm{UGE} \rangle$ to give $\langle / \mathrm{UGE} \rangle$ her brother $\langle \mathrm{UGE} \rangle$ a push $\langle / \mathrm{UGE} \rangle$ home since it was late, and she wanted him to get back safely.	to give, a push	to accompany a person for some distance
$\langle \mathrm{UGE} \rangle$ Gifted by nature $\langle / \mathrm{UGE} \rangle$, voluptuous figure serves as a testament to the allure of femininity, captivating everyone who crosses her path.	gifted by nature	a woman or girl to have features, eg large breasts, wellrounded hips and full lips

TABLE 5: Context words and their meanings extracted from annotated sentences

We annotated 124,657 UgE sentences. After removing empty sentences and discarding failed preprocessing cases, we retained 124,657 sentence rows and extracted 128,609 span supervision instances (Table *4*). We split the data into 97,828/12,229/12,229 train/validation/test sets using a fixed seed (random_state = 42) and report all split counts (Table *6*).

Language	No. of sentence	Training data	Testing data	Validation data
UgE	124,657	97,828	12,229	12,229

TABLE 6: Statistics of Ugandan English (UgE) sentences for context generation

Model selection

We explored several transformer models to identify the most suitable architecture for generating culturally contextualized UgE explanations by evaluating BART, GPT-2 (and its variants), T5, DistilGPT2, allenai_ledbase16384, GPT-Neo, mBART-large-50, and OPUS-MT (opus-mt-mul-en). These models spanned seven useful architectural categories, as indicated in Table 7.

Model	Category	Rationale for inclusion	Citation
gpt2, distilgpt2, gpt2-medium, gpt2-large	Decoder-only	Autoregressive text generation baseline; tests limits of decoder-only pretraining for conditioned sequence-to-sequence cultural glossing; DistilGPT2 included for efficiency trade-offs.	[67,68]
gpt-neo-125M	Decoder-only (open-source)	Open-source alternative to larger autoregressive models (GPT-3-like behavior) for reproducible generative evaluation.	[69]
facebook_bart-base	Encoder-decoder	Denoising pretraining (bidirectional encoder + autoregressive decoder) suited for text reconstruction, translation and summarisation.	[70]
t5-small, t5-base, t5-large	Encoder-decoder (text-to-text)	Unified text-to-text formulation gives flexibility across tasks; multiple sizes test capacity effects.	[71]
allenai_led-base-16384 (LED)	Long-context encoder-decoder	Sparse attention for long-document handling to process extended contexts and source-marked spans.	[72]
mbart-large-50	Multilingual encoder-decoder	Multilingual seq2seq pretraining helps with code-mixing and loanword phenomena in Ugandan English.	[73]
opus-mt-mul-en	Translation-style (Marian/OPUS)	Translation-trained model framing glossing as a style/translation task; strong performance on culturally salient lexical mapping.	[74]

TABLE 7: Baseline and state-of-the-art model selection

Decoder-only models (GPT-2, distilgpt2, gpt2-medium, gpt2-large, and gpt-neo-125M) are strong autoregressive generators that excel in open-ended text production but tend to struggle with tightly conditioned sequence-to-sequence tasks that require deep bidirectional encoding of source spans, a limitation reflected in their relatively low scores in our evaluation [68-70]. The encoder-decoder models (facebook_bart-base and the T5 family: t5-small/t5-base/t5-large, together with the multilingual mbart-large-50) combine bidirectional encoding with autoregressive decoding and are well suited to reconstruction, translation, and rewriting tasks. BART's denoising pretraining and T5's unified text-to-text formulation provide strong inductive biases for generating coherent, contextually appropriate glosses [71,72]. Long-context encoder-decoder models, such as allenai_led-base-16384 (LED), incorporate sparse/extended attention mechanisms that allow efficient processing of much longer input sequences, making them especially promising when the task requires attending to extended source contexts or multiple marked spans [73]. The multilingual category (mbart-large-50) brings cross-lingual pretraining benefits that help handle code-mixing and loanwords common in UgE [74]. Finally, translation-style models such as opus-mt-mul-en (Marian/OPUS) were trained specifically for sequence-to-sequence translation and style transfer, and framing the glossing task in this way proved effective for mapping culturally specific lexical items to natural explanations [74].

Empirically, encoder-decoder and translation-style models, most notably OPUS-MT and LED, always provide the best overall performance (including the highest RetriEVAL scores); therefore, we prioritized these architectures for detailed analysis and subsequent ablations.

Experimental setup

The study generates culturally contextualized meanings from sentences annotated with $\langle \mathrm{UGE} \rangle$ spans. The complete pipeline includes (i) tag-aware preprocessing to build inputs and supervision targets, (ii) a two-stage fine-tuning procedure (Stage A for context prediction and Stage B for meaning generation), and (iii) a polysemy-aware evaluation protocol for meaning scoring.

Data Preprocessing

We load the dataset CSV from the repository root and preprocess each row by (1) extracting the span contents inside $\langle \mathrm{UGE} \rangle \cdots \langle / \mathrm{UGE} \rangle$; (2) removing the tags while keeping the enclosed tokens to obtain processed_sentence; and (3) constructing the supervision context Context by concatenating all extracted span contents into a single space-separated sequence (dropping empty spans after stripping). Missing values are replaced with empty strings, and internal whitespace is normalized by collapsing repeated spaces. Examples are retained for training and evaluation only when both processed_sentence and Meaning_New are non-empty.

Train/Validation/Test Split

We use an 80%/10%/10% split via two calls to $\mathrm{train\_test\_split}$ with $\mathrm{random\_state{=}42}$, ensuring reproducibility.

Two-Stage Training

Fine-tuning is performed for pretrained language models using a two-stage context-to-meaning design. Stage A trains a conditional generation model to predict the supervision context from the sentence alone. Stage B trains a second conditional generation model to predict $\mathrm{Meaning}$ given both the sentence and the context (at inference, the context produced by Stage A is used). Tokenization uses truncation and fixed-length padding with max_length = 256. If the tokenizer lacks an explicit padding token, pad_token is set to eos_token (or unk_token as needed). For causal variants, prompt tokens are masked using the ignore index −100 so that they do not contribute to the training loss.

Reproducibility and Determinism

We fix seeds across Python, NumPy, and PyTorch with SEED = 42. Deterministic behavior is encouraged by setting CuDNN to deterministic mode and disabling benchmarking. Data loading follows deterministic settings (e.g., num_workers = 0 and seeded shuffling). For each stage, the runtime environment is recorded in environment.json (including Python/PyTorch/Transformers versions and CPU metadata), and the exact run configuration scripts are provided with the repository accessed at https://github.com/dbyansi/Byansi_Project/.

Compute Requirements and Efficiency

To support reproducibility in low-resource environments, we report compute footprint and wall-clock training time. All experiments were executed on a CPU-only cluster node with 223 CPUs per task and 93 GB RAM. Stage A and Stage B were trained as separate jobs: Stage A is estimated to run from 2026-06-11 07:46:35 to 2026-06-12 02:10:10 (wall-clock duration: 18 h 23 m), and Stage B ran from 2026-06-12 02:11:47 to 2026-06-13 06:49:59 (wall-clock duration: 28 h 38 m). The end-to-end wall-clock time from launching the overall workflow to producing the final outputs was 2026-06-11 07:43:29 to 2026-06-16 06:33:34 (total duration: 4 d 22 h 50 m), including preprocessing, stage handoff (Stage A outputs used by Stage B), and evaluation/decoding. Training uses fixed sequence lengths (256) with per-device batch size 8, gradient clipping (max norm 1.0), and deterministic settings (seed 42, CuDNN deterministic mode, and benchmarking disabled). For constrained environments, we recommend reducing batch size while keeping the maximum sequence length unchanged and using CPU-only execution as in the reference configuration.

Inference and Decoding

During inference, Stage A generates a predicted context for each processed_sentence. Stage B then generates the meaning conditioned on the sentence and the predicted context using beam search with num_beams = 4, do_sample = False, and early_stopping = True (max length 256).

Polysemy-Aware Evaluation.

Because Meaning_New annotations can be ambiguous for the same surface sentence (polysemy), we avoid forcing a single reference string. We construct an example key from processed_sentence by normalizing case and collapsing repeated whitespace. For each key, we collect all distinct gold meaning strings in the test split and mark a key as ambiguous if it has at least two distinct gold meanings (otherwise unambiguous). For scoring, each prediction is compared against all gold candidates for its key; similarity-based metrics (e.g., BLEU, ROUGE, BERTScore, RetriEVAL) use best-match maximization over candidates, while distance-based metrics (e.g., TER, edit distance, WER) use best-match minimization. We report overall polysemy-aware metrics and separate results for ambiguous vs. unambiguous subsets. Context quality is evaluated by comparing Stage A predicted contexts against gold contexts with automatic lexical and semantic metrics, while meaning quality is evaluated with the polysemy-aware procedure above.

Table *8* summarizes the hyperparameters used in the experimental setup.

How to cite this article:
Byansi D, Isingoma B, Nakatumba-Nabende J, et al. (July 15, 2026) Cultural and Linguistic Contextualization in Ugandan English: Training Transformer Models for Enhanced Local Relevance. Cureus J Comput Sci 3 : es44389-026-00167-y. DOI https://doi.org/10.7759/s44389-026-00167-y

Category/hyperparameter	Exact setting used in the provided code
Optimizer	AdamW via torch.optim.AdamW(model.parameters(), lr=learning_rate, weight_decay=weight_decay)
Learning rate	5e-5 (passed as learning_rate=5e-5 to both Stage A and Stage B; used as lr in AdamW)
Batch size	8 per device (passed as per_device_train_batch_size=8; used in DataLoader(..., batch_size=per_device_train_batch_size)).
Epochs	2 (Stage A and Stage B called with num_epochs=2)
Warmup steps	50 (warmup_steps=50)
Scheduler type	Linear warmup + linear decay via get_linear_schedule_with_warmup(... num_warmup_steps=warmup_steps, num_training_steps=total_steps); called every step with scheduler.step()
Weight decay	0.0 default in function signature (weight_decay: float=0.0); Stage A/B calls do not override weight_decay, so AdamW uses weight_decay=0.0
Gradient clipping	Enabled each step: torch.nn.utils.clip_grad_norm_(model.parameters(), max_norm=1.0)
Max input length	256 (training/inference called with max_length=256; tokenization uses max_length=max_length with truncation=True)
Max target length	256 (targets tokenized with tokenizer(..., max_length=max_length); for seq2seq, labels use these tokenized lengths)
Padding strategy	Tokenization uses padding="longest"; batches are padded by HuggingFace collators: DataCollatorForSeq2Seq(..., padding="longest") (non-causal) or DataCollatorWithPadding(tokenizer, padding="longest") (causal)
Label smoothing	Not used (no label_smoothing argument passed anywhere)
Seed(s) / determinism	42 via set_seed(42) inside main() and again inside train_single_stage(seed=42). Uses: random.seed, np.random.seed, torch.manual_seed,

	torch.cuda.manual_seed, torch.cuda.manual_seed_all, and sets cudnn.deterministic=True, cudnn.benchmark=False.
Train/valid/test splitting	train_test_split(..., random_state=42) and initial data.sample(frac=1, random_state=42).
Mixed precision / hardware	AMP optional: uses torch.cuda.amp.autocast() + torch.cuda.amp.GradScaler() only if torch.cuda.is_available(). Otherwise runs FP32 only.
Early stopping	Enabled: early_stopping_patience=2 for both stages; monitors avg_valid_loss and stops when no improvement for patience epochs.
Generation decoding (test inference)	Beam search: num_beams=4, early_stopping=True; max_length=256.

TABLE 8: Reproducibility: exact training and key runtime settings (applies to both Stage A and Stage B unless stated otherwise)

Model training

Transformer models were fine-tuned on a preprocessed UgE dataset using a two-stage, context-to-meaning framework. Instead of training a single end-to-end system, Stage A and Stage B were designed to improve controllability and reduce ambiguity-related errors by explicitly separating contextual representation learning from meaning generation. In Stage A, the model learns to produce a contextual representation of the input sentence; in Stage B, the model generates the corresponding meaning conditioned on both the original sentence and the predicted context.

To ensure robust evaluation and prevent information leakage, the dataset was split into three disjoint subsets: a training set for parameter learning, a validation set for monitoring and model selection, and a held-out test set, reserved exclusively for final evaluation. Text preprocessing transformed UgE annotations into model-ready training targets. Specifically, the $\langle \mathrm{UGE} \rangle \cdots \langle / \mathrm{UGE} \rangle$ wrapper was removed while preserving its inner text, contextual spans were extracted and normalized by concatenating multiple spans into a single context sequence, and meaning targets were normalized for whitespace consistency. Each stage used prompt-based input formatting so that the model received structured signals aligned with the respective decomposition objective.

Tokenization was performed using the model-specific tokenizer. Sequences were truncated to a fixed maximum length for computational efficiency, and padding was applied to support batch processing. For each stage, tokenization was applied separately to stage-specific prompts (inputs) and stage-specific targets (labels): Stage A targets were the gold contextual representations. In contrast, Stage B targets were the gold meanings. The training objective optimized the loss between the generated output tokens and the gold-annotated target tokens (with standard label-masked behavior for causal-model variants, where applicable).

Optimization used AdamW together with a linear learning-rate schedule with warmup. Gradient clipping was applied to prevent instability during training. Training progress was logged, and after each epoch, the model was evaluated on the validation set using validation loss as the primary stopping criterion. Early stopping based on a patience mechanism was used to halt training when validation loss did not improve, thereby reducing unnecessary computation and limiting overfitting.

How to cite this article:
Byansi D, Isingoma B, Nakatumba-Nabende J, et al. (July 15, 2026) Cultural and Linguistic Contextualization in Ugandan English: Training Transformer Models for Enhanced Local Relevance. Cureus J Comput Sci 3 : es44389-026-00167-y. DOI https://doi.org/10.7759/s44389-026-00167-y

Because UgE meanings can be ambiguous for the same surface input (polysemy), the framework incorporated polysemy handling during the evaluation stage to avoid penalizing correct predictions that match alternative valid senses. For each test example, a key was constructed from the normalized $\text{processed\_sentence}$. For that key, all distinct gold meaning values present in the test split were collected, and the number of distinct meanings was used to identify ambiguous keys (those with at least two distinct gold meanings) versus unambiguous keys. During metric computation for meaning generation, predictions were scored against all gold meaning candidates for the key, and metric aggregation used a best-match strategy (that is, the maximum similarity-based scores and minimum distance-based scores across candidates). As a result, when multiple senses are valid for the same sentence, the evaluation reflects the likelihood that the model correctly captured at least one legitimate sense rather than forcing an arbitrary single reference.

After training, Stage A and Stage B were evaluated on the held-out test dataset using generation. Stage A produced predicted context representations for each input sentence. Stage B then generated predicted meanings conditioned on the input sentence and the predicted context. Predicted contexts were evaluated against gold contexts using overlap-based and similarity-based metrics, while predicted meanings were evaluated using the polysemy-aware procedure described above. The two-stage decomposition first learned the sentence-to-context structure and then mapped context to linguistically appropriate meanings, while the polysemy-aware evaluation ensured that ambiguity in UgE annotations was properly captured in the reported results.

Model evaluation and metrics

Model performance was evaluated on the held-out test set using a suite of metrics designed to capture lexical fidelity, reference-text overlap, semantic similarity, and generation-error characteristics. For each test example, the two-stage pipeline produces (i) a predicted context (Stage A) and (ii) a predicted meaning conditioned on both the input sentence and the predicted context (Stage B). Metrics are computed by comparing the generated outputs with their corresponding gold annotations: context-level metrics compare the predicted context with the gold context, whereas meaning-level metrics compare the predicted meaning with the gold meaning.

We report standard automatic generation metrics: BLEU and ROUGE (n-gram overlap in precision/recall style), BERTScore (semantic alignment via contextual embeddings), and error-based measures, including TER (token-level edit distance normalized by reference length), edit distance (absolute token edit distance), WER (word-level error rate), and chrF (character n-gram F-score). To reduce metric brittleness, we use corpus-level chrF where available and include semantic-oriented measures (BERTScore and embedding similarity) alongside overlap-based measures, so that paraphrastic wording does not automatically yield low scores.

Retrieval-Oriented Metric: RetriEVAL

Because the task requires both surface consistency and meaning preservation (particularly under UgE phrasing variation), we introduce *RetriEVAL*, a retrieval-inspired agreement score that blends (a) lexical overlap and (b) semantic similarity. For each reference-prediction pair (r, p), lexical overlap measures shared tokens, while semantic similarity measures closeness in embedding space. The final score is the mean of the two components in the current implementation, i.e., equal weighting:

$$\mathrm{RetriEVAL}(r,p) = \frac{\mathrm{lexical\_overlap}(r,p) + \mathrm{sim}(r,p)}{2}.$$

Lexical Overlap (Implementation-Aligned)

Lexical overlap is computed using whitespace tokenization: $\mathcal{W}(x)$ denotes the set of tokens obtained from x via $\mathrm{split}()$ without punctuation stripping or case normalization. The overlap score is therefore

$$\mathrm{lexical\_overlap}(r,p) = \frac{|\mathcal{W}(r) \cap \mathcal{W}(p)|}{\max(|\mathcal{W}(r)|,\, 1)}.$$

How to cite this article:
Byansi D, Isingoma B, Nakatumba-Nabende J, et al. (July 15, 2026) Cultural and Linguistic Contextualization in Ugandan English: Training Transformer Models for Enhanced Local Relevance. Cureus J Comput Sci 3 : es44389-026-00167-y. DOI https://doi.org/10.7759/s44389-026-00167-y

This design explicitly rewards surface-level agreement with reference words/phrases, which is useful for detecting omissions or substitutions that may change pragmatic meaning.

Semantic Similarity (Implementation-Aligned)

Semantic similarity is computed using a pretrained SentenceTransformer model $\mathbf{all\text{-}MiniLM\text{-}L6\text{-}v2}$. Each pair $[\boldsymbol{r}, \boldsymbol{p}]$ is embedded and compared via cosine similarity:

$$\mathbf{sim}(r, p) = \cos(\mathbf{e}(r),\, \mathbf{e}(p)),$$

where $\mathbf{e}(\cdot)$ is the SentenceTransformer embedding. Cosine similarity is computed directly using the library routine in the implementation; no explicit additional normalization step is applied beyond what the cosine implementation performs.

Reported Value

RetriEVAL is computed per example and then averaged over the test set. We adopt equal weighting as a baseline to ensure that the metric is not dominated by either (i) exact token overlap or (ii) semantic similarity under paraphrase. This makes RetriEVAL well-suited for comparing models in a setting where UgE may exhibit both lexical variation and meaning-preserving rephrasing.

Results And Discussion

This section provides a detailed evaluation and discussion based on the aggregated results presented in Table *9*, which reports the performance of the two-stage generation framework (Stage A: context generation; Stage B: meaning generation) using multiple metrics, including overlap-based measures and learned semantic similarity. We further conduct human evaluation on samples from one of the models using Krippendorff's alpha (nominal) to measure multi-rater agreement for nominal categories and Cohen's kappa to assess pairwise rater agreement. The results reveal a clear distinction between models that generate outputs closely matching the reference contexts and meanings and those that exhibit substantial degradation, even when some semantic similarity metrics remain relatively high.

How to cite this article:
Byansi D, Isingoma B, Nakatumba-Nabende J, et al. (July 15, 2026) Cultural and Linguistic Contextualization in Ugandan English: Training Transformer Models for Enhanced Local Relevance. Cureus J Comput Sci 3 : es44389-026-00167-y. DOI https://doi.org/10.7759/s44389-026-00167-y

		Models										
	Metrics	led-base-16384	gpt-neo-125M	bart-base	mbart-large-50	t5-base	t5-small	t5-large	gpt2	gpt2-Medium	gpt2-Large	opus-mt-mul-en
	BLEU	0.2565	0.0189	0.2583	0.2534	0.252	0.2405	0.199	0.0188	0.0189	0.0071	0.254
	ROUGE1	0.9929	0.159	0.9958	0.9899	0.9856	0.9723	0.9723	0.1615	0.1636	0.0551	0.9895
	ROUGE2	0.3308	0.0527	0.3362	0.3258	0.3225	0.3039	0.3039	0.0513	0.0524	0.0106	0.3248
	ROUGEL	0.9929	0.159	0.9958	0.9899	0.9855	0.9721	0.9721	0.1615	0.1636	0.0542	0.9895
STAGE A (Context)	TER	0.0094	13.6927	0.0074	0.018	0.0265	0.0461	0.0461	13.5186	13.4098	12.3499	0.0188
	BERTScore	0.9984	0.8375	0.9988	0.997	0.996	0.9937	0.9906	0.8365	0.8371	0.8064	0.9969
	EditDistance	0.0206	17	0.0136	0.0339	0.0427	0.0804	0.0804	16.7211	16.5918	36.017	0.0336
	WER	0.0094	13.6927	0.0074	0.018	0.0265	0.0461	0.0461	13.5186	13.4098	12.3499	0.0188
	chrF	45.4413	7.1437	45.4871	45.4375	45.3393	45.1238	45.1238	7.1437	7.1437	3.7889	45.4239
	RetriEVAL	0.9936	0.6364	0.996	0.9888	0.9852	0.9752	0.9752	0.6072	0.6026	0.2221	0.99
	BLEU	0.73	0.0036	0.723	0.6305	0.6928	0.4308	0.4308	0.0265	0.0038	0.0034	0.6887
	ROUGE1	0.9937	0.0425	0.99	0.9201	0.9719	0.6396	0.6396	0.1292	0.0482	0.0476	0.973
	ROUGE2	0.8485	0.0048	0.8453	0.7556	0.8233	0.5143	0.5143	0.0366	0.005	0.0037	0.8203

How to cite this article:
Byansi D, Isingoma B, Nakatumba-Nabende J, et al. (July 15, 2026) Cultural and Linguistic Contextualization in Ugandan English: Training Transformer Models for Enhanced Local Relevance. Cureus J Comput Sci 3 : es44389-026-00167-y. DOI https://doi.org/10.7759/s44389-026-00167-y

	ROUGEL	0.9936	0.0412	0.9899	0.9186	0.9714	0.6367	0.6367	0.1233	0.047	0.0465	0.973
STAGE B (Meaning)	TER	0.0105	16.5708	0.0134	0.1	0.049	0.7073	0.7073	12.3406	12.2929	13.4993	0.0412
	BERTScore	0.9987	0.7966	0.9981	0.9819	0.9922	0.9308	0.9308	0.8053	0.8052	0.8366	0.9921
	EditDistance	0.0612	49.3265	0.1294	0.9643	0.32	2.9734	2.9734	35.9966	35.7993	16.7381	0.3767
	WER	0.0105	16.5708	0.0134	0.1	0.049	0.7073	0.7073	12.3406	12.2929	13.4993	0.0412
	chrF	23.2475	3.7889	23.2737	23.1121	23.097	19.1372	19.1372	3.7889	3.7889	7.1437	23.2341
	RetriEVAL	0.9932	0.2195	0.9926	0.9215	0.9695	0.6632	0.6632	0.2086	0.2062	0.6291	0.9707

TABLE 9: Evaluation of transformer model performance on culturally and linguistically contextualized response generation

Overview of observed performance patterns

The evaluation separates Stage A (context generation) from Stage B (meaning generation). Across both stages, model behavior shows a clear division between (i) strong context-to-meaning performers, which achieve consistently high similarity metrics and low distance metrics, and (ii) weak performers, which fail on lexical overlap-based metrics and inflate distance measures.

A notable observation is that, for most metrics in Stage A, led-base-16384 and bart-base rank near the top, while gpt2, gpt2-Large, and gpt2-Medium are among the lowest performers, often by multiple orders of magnitude for distance metrics such as TER, WER, and EditDistance. Stage B partially maintains this ranking: led-base-16384 again leads most metrics, while bart-base remains competitive but no longer ties or matches across all measures; opus-mt-mul-en demonstrates strong Stage B performance despite not being the top performer in Stage A.

These results indicate that strong performance in Stage A does not guarantee success in Stage B. Instead, Stage B appears to depend both on the quality of the provided context (error propagation) and on the model's ability to use context effectively to produce meaning representations.

Best-per-metric ranking (Stage A vs. Stage B)

To improve interpretability, we highlight the best model(s) for each metric, considering directionality: for similarity metrics, higher is better (BLEU, ROUGE, BERTScore, chrF, RetriEVAL), while for distance or error metrics, lower is better (TER, EditDistance, WER).

(i) Stage A (Context) - best per metric: Bart-base achieves the strongest overall performance across most Stage A similarity metrics: it leads BLEU (0.2583), ties for the top in ROUGE1 (0.9958) and ROUGEL (0.9958), and leads ROUGE2 (0.3362). It also achieves the best scores on BERTScore (0.9988), chrF (45.4871), and RetriEVAL (0.9960). For distance and error metrics, bart-base again provides the best error profile: lowest TER (0.0074), lowest WER (0.0074), and lowest EditDistance (0.0136). Overall, Stage A suggests that bart-base is the dominant model across both lexical overlap and similarity, as well as distance and error measures for context generation.

(ii) Stage B (Meaning) - best per metric: In Stage B, the best-per-metric pattern shifts toward led-base-16384 as the primary winner for meaning-focused quality. It leads BLEU (0.7300), ROUGE1 (0.9937), ROUGE2 (0.8485), and ROUGEL (0.9936). It also performs best on distance/error metrics: lowest TER (0.0105), lowest WER (0.0105), and lowest EditDistance (0.0612). For learned semantic similarity, led-base-16384 ranks highest on BERTScore (0.9987) and RetriEVAL (0.9932). For chrF, the best score is achieved by bart-base (23.2737), with led-base-16384 very close (23.2475) and opus-mt-mul-en also competitive (23.2341). Thus, Stage B shows a strong concentration of meaning quality in led-base-16384, with bart-base retaining a slight advantage specifically under chrF.

Deeper error analysis

Although Table *10* aggregates results, the relative ordering across metrics provides insight into error modes.

(i) Lexical overlap versus semantic similarity discrepancies: Several weaker models receive low BLEU and ROUGE scores and also show substantially reduced BERTScore and RetriEVAL, along with significantly increased TER, WER, and EditDistance. This suggests that failures are not merely due to paraphrase-level surface variation but also involve deeper token-sequence alignment errors. Conversely, strong models that maintain high overlap and semantic metrics indicate successful alignment with the reference structure and meaning.

(ii) Directional evidence of error propagation (Stage A → Stage B): Stage A similarity/error performance is strongest for bart-base (which has the best overall context metrics), while Stage B meaning quality is highest for led-base-16384. This suggests that context faithfulness alone is not sufficient for optimal meaning generation: Stage B may require not only a close context match but also an architecture-specific ability to condition on that context and translate it into the correct semantic representation. In other words, Stage B can re-rank models even when the top Stage A performer differs.

(iii) Catastrophic degradation for certain autoregressive baselines: Models such as gpt-neo-125M and gpt2 variants exhibit extremely high TER, WER, and EditDistance in Stage A and remain weak in Stage B. This behavior indicates a mismatch between the generation style and the evaluation setting, where outputs must meet structured lexical and semantic constraints to avoid large token sequence errors.

How stage A performance influences stage B

To explicitly connect the stages, consider three scenarios:

(i) Scenario A: High Stage A quality ⇒ strong Stage B (partial consistency): led-base-16384 performs very well in Stage A (BERTScore and RetriEVAL near the top; low TER, WER, and EditDistance) and is also the clear Stage B winner across most metrics. This supports a beneficial error-propagation effect: when Stage A produces context that is semantically and lexically close to gold, Stage B is more likely to decode correct meanings.

(ii) Scenario B: Best Stage A does not imply best Stage B: bart-base leads Stage A across nearly all metrics, yet led-base-16384 surpasses it in Stage B on most meaning metrics. This supports the claim that Stage B is not a monotonic function of Stage A quality.

(iii) Scenario C: Stage A moderate quality but Stage B strong performance: opus-mt-mul-en is not the top performer in Stage A, but it remains competitive in Stage B, particularly under distance/error measures (e.g., low TER/WER relative to weak baselines) and in RetriEVAL. This suggests partial robustness to context imperfections under certain linguistic conditions.

How to cite this article:
Byansi D, Isingoma B, Nakatumba-Nabende J, et al. (July 15, 2026) Cultural and Linguistic Contextualization in Ugandan English: Training Transformer Models for Enhanced Local Relevance. Cureus J Comput Sci 3 : es44389-026-00167-y. DOI https://doi.org/10.7759/s44389-026-00167-y

Scalability and efficiency considerations

Although the table results provide metric outcomes, scalability must also account for computational cost. In a two-stage inference setup, any improvement in Stage B quality requires additional generation passes because the model must first decode in Stage A and then perform a second decoding process for Stage B, conditioned on the generated context. Model size and decoding complexity are also important, as larger models - while often improving semantic alignment - require more memory and compute, and strategies like beam search can increase latency. This creates a deployment trade-off between metric gains (e.g., higher led-base-16384 scores in Stage B) and inference throughput. Finally, context-conditioned tasks can be sensitive to long inputs where truncation may occur, causing metrics based on exact token overlap (such as BLEU or ROUGE) to degrade sharply, while semantic metrics like BERTScore may decline more gradually.

To further address scalability concerns beyond computational latency, note that our two-stage setup (Stage A: context generation; Stage B: meaning generation) does not require custom task redesign when extending coverage. Scalability in our annotation design is determined by the linguistic feature inventory and written guidelines: expanding to additional UgE features can be managed by extending the tagged span inventory while maintaining the same span-to-gloss task format and two-stage conditional generation protocol. Annotator training and agreement are monitored for each feature under the guideline structure described in the Materials and Methods section.

Cultural bias and representational fairness

Because the dataset and prompts target “culturally contextualized responses,” the meaning labels may reflect culturally specific language, conventions, and pragmatic norms, which can introduce bias risks. In addition to distributional and evaluation-related concerns, our culturally annotated pipeline can induce two additional bias mechanisms:

(i) Coverage imbalance: While ICE-UG and Web-UG provide diverse real UgE material, they do not guarantee balanced coverage across all UgE speaker communities. This can result in uneven representation of cultural and register variation, which may produce uneven mapping between surface cues and meanings across communities.

(ii) Canonicalization-induced label noise: Our canonical gloss assignment step (Table *4*) collapses polysemous spans into a single span-type mapping. This may introduce systematic evaluation bias by penalizing contextually valid but non-canonical senses.

We mitigate canonical-label noise by quantifying polysemy and using polysemy-aware best-match scoring during evaluation to avoid unfair penalization of legitimate contextual interpretations. To further reduce canonicalization bias, we plan to shift toward more context-conditioned, row-level supervision so that each instance prediction target reflects its own meaning rather than a single canonical gloss for a span type.

Ethical considerations

In applications involving culturally contextual meaning generation, major ethical risks include misinterpretation harm, where incorrect outputs can spread misunderstanding across cultural settings, undermine user trust, or influence real-world decisions; overconfidence from metric-optimized outputs, as high similarity scores (such as overlap with reference text) do not necessarily guarantee faithful or safe meaning, and models may produce fluent responses that still contain subtle distortions; ambiguity and polysemy concerns, since meaning can be inherently unclear and systems may over-select dominant interpretations, erasing minority or less frequent meanings; and privacy risks, because if prompts or contexts include sensitive cultural identifiers, two-stage generation can increase the amount of text produced and potentially stored or logged, making careful data minimization and retention policies essential.

Limitations of the current evaluation

While the evaluation provides a strong quantitative snapshot, several limitations should be noted: the metrics exhibit a coverage mismatch - BLEU and ROUGE primarily measure n-gram overlap, TER, WER, and EditDistance focus on token-sequence differences, and BERTScore and RetriEVAL estimate semantic similarity - yet none fully capture pragmatic correctness, cultural appropriateness, or safety; aggregated mean scores may obscure per-example variance, including failure clusters associated with specific cultural domains, input length regimes, or ambiguous meanings; polysemy handling may remain imperfect, as best-match scoring over multiple gold candidates can overestimate performance when several meanings are plausible, even if the model selects an unintended but "close" option; and because Stage B is conditioned on the generated Stage A context, observed improvements may reflect reduced error propagation or coupling artifacts rather than genuine gains in meaning comprehension.

While automatic metrics provide scalable evaluation, they are incomplete proxies for cultural adequacy and pragmatic faithfulness. Therefore, we use them as operational indicators of surface fidelity (n-gram overlap), generation errors (TER/WER/EditDistance), and semantic alignment (BERTScore, embedding similarity). We further validate key conclusions with human evaluation on contextual sense correctness and faithfulness to context.

In addition, while our quantitative and human evaluations provide a strong initial assessment, external validity remains limited by the development sources used for corpus construction and model tuning. To address the issue of broader real-world validation, we will extend evaluation with explicit cross-source, deployment-like testing on held-out UgE materials from additional domains, source publishers, and speaker communities not used during corpus construction or model development (i.e., moving beyond ICE-UG/Web-UG origin distributions).

Recommendations for strengthening error analysis

To further deepen the evaluation beyond aggregated metrics, we recommend the following analyses:

(i) Error taxonomy by type: classify errors as (1) context omission or under-specification, (2) hallucinated context, (3) incorrect meaning sense selection, (4) paraphrase drift, and (5) formatting or tokenization mismatches.

(ii) Ambiguity-stratified evaluation: report metrics separately for ambiguous and unambiguous keys, and analyze which metrics change most. This can reveal whether the system fails primarily on sense disambiguation or surface fidelity.

(iii) Cultural subgroup analysis: if cultural tags or domains exist, compute subgroup metrics and measure disparity (e.g., effect sizes in BERTScore or chrF and error increases in TER or WER).

(iv) Robustness checks: perturb Stage A contexts (e.g., paraphrasing, deletion of spans) and quantify how quickly Stage B meaning quality declines to estimate context sensitivity.

Qualitative analysis

Tables *10*-*14* present qualitative predictions for five models - allenai_led-base-16384 (first), Opus-mt-mul-en, t5-base, BART-base, and GPT2 - under the same schema: processed_sentence (input with target token/context), gold_context vs. predicted_context (gold vs. model-produced local textual instantiation), and gold_meaning vs. predicted_meaning (gold vs. model-produced gloss/definition for the target). Each table has 30 instances (examples) that were critically analyzed.

For each model, we classify instances into (i) Match: predicted_context = gold_context and predicted_meaning = gold_meaning. (ii) Mismatch: at least one component diverges-either the local embedding differs (context error), the gloss/meaning differs (meaning error), or both.

For allenai_led-base-16384 (Table *10*), most instances fall into the meaning-match regime: many targets (e.g., Savedee, reduce on, detoother/golddigger, slowly-slowly, lukiiko, cowardice, oyee, facilitate, special hire, godown, pakalast, witchdoctors, posho, knocked down, pilot, and others) produce predicted_meaning identical to gold_meaning. At the same time, context-slot fidelity is highly consistent in the table (gold_context and predicted_context are the same across the listed rows).

Where errors occur, they are mainly gloss/meaning divergences rather than slot mismatches-e.g., digging maps to farming, beeped changes its gloss, and run mad shifts its meaning away from the gold.

allenai_led-base-16384 model therefore shows the strongest coupling in your results between the model's lexical choice and the intended gloss, with deviations concentrated in specific target meanings rather than in the contextual slot.

How to cite this article:
Byansi D, Isingoma B, Nakatumba-Nabende J, et al. (July 15, 2026) Cultural and Linguistic Contextualization in Ugandan English: Training Transformer Models for Enhanced Local Relevance. Cureus J Comput Sci 3 : es44389-026-00167-y. DOI https://doi.org/10.7759/s44389-026-00167-y

Processed_sentence	**Gold_context**	**Predicted_context**	**Gold_meaning**	**Predicted_meaning**
As a Savedee, she felt compelled to volunteer at local charities, believing in the importance of serving others in Christ's name.	Savedee	Savedee	a religiously saved person	a religiously saved person
At the exclusive black-tie gala, the wealthy socialites get to eat big at the multi-course meal.	eat big	eat big	to get the lion's share enjoy something more	to get the lion's share enjoy something more
She aims to reduce on the amount of sugar in her baking recipes.	reduce on	reduce on	reduce something	reduce something
I admire farmers who dedicate themselves to digging, understanding that their hard work feeds not just their families but the entire community.	digging	digging	farming	farming
Even when they shared great times, he couldn't shake the feeling that she was just a detoother interested in his wealth.	detoother	detoother	golddigger	golddigger
I like to reflect on my day slowly-slowly, writing in my journal to process my thoughts and emotions before going to bed.	slowly-slowly	slowly-slowly	bit by bit	bit by bit
The lukiiko has organized community dialogues to encourage	lukiiko	lukiiko	parliament of Buganda or Busoga Kingdom	parliament of Buganda or Busoga Kingdom

conversations about social issues, fostering a sense of community and collaboration.				
They both felt the attraction growing stronger, yet during that night, they mutually agreed that they did not play sex and simply enjoyed their time together.	play sex	play sex	to have sexual intercourse	to have sexual intercourse
I enjoy original films that make me think.	original	original	of exceptional quality	of exceptional quality
The campuser enjoys attending guest lectures, as they provide unique perspectives and insights from industry leaders and experts in various fields.	campuser	campuser	university student	university student
In a surprising turn of events, her cowardice during the negotiation led to a missed opportunity for mutual benefit.	cowardice	cowardice	to behave timidly or like a coward	to behave timidly or like a coward
He is that one individual who always stands up for what is right, even when it's unpopular or difficult to do so.	that one	that one	that person or individual	that person or individual
During the graduation ceremony, the graduates raised their hands and shouted oyee !in celebration of their hard work and achievements.	oyee	oyee	viva or long live	viva or long live
The existence of safe houses has raised	safe houses	safe houses	shadowy prisons in which suspects are	shadowy prisons in which suspects are

significant concerns about the rule of law and the treatment of individuals in the justice system.			discreetly held illegal places of detention	discreetly held illegal places of detention
The importance of kakuyege in building a responsive government cannot be overstated, as it allows voters to shape the political agenda.	kakuyege	kakuyege	vote canvassing	vote canvassing
I want to facilitate the service providers for their reliability.	facilitate	facilitate	to pay someone for something	to pay someone for something
She quickly beeped her colleague's number to get his attention and share an important update.	beeped	beeped	rang up the intended recipient of the call and hang up immediately	rings up the intended recipient of the call and hang up immediately
I was impressed by the punctuality of the special hire driver who picked me up for my flight; it really made a difference.	special hire	special hire	taxi or cab	taxi or cab
They converted the godown into a cozy living space, adding furniture and decorations to make it feel more inviting and warm.	godown	godown	a basement	a basement
I believe that kindness should extend pakalast, as small acts of compassion can create a ripple effect, positively impacting countless lives.	pakalast	pakalast	to the very end, till death	to the very end, till death
He was devastated to learn his girlfriend was a detoother, having only pursued	detoother	detoother	golddigger	golddigger

How to cite this article:
Byansi D, Isingoma B, Nakatumba-Nabende J, et al. (July 15, 2026) Cultural and Linguistic Contextualization in Ugandan English: Training Transformer Models for Enhanced Local Relevance. Cureus J Comput Sci 3 : es44389-026-00167-y. DOI https://doi.org/10.7759/s44389-026-00167-y

him for his financial resources.				
After discussing our proggie s, we decided to plan a weekend getaway that would allow us to relax and explore the countryside together.	proggie	proggie	one's social plans	one's social plans
The constant noise from the construction site next door made it hard to concentrate, and I felt like I was going to run mad.	run mad	run mad	to go mad to become insane	to go mad to become insane
Witchdoctors may also serve as mediators in familial disputes, using their influence to help resolve conflicts and restore harmony among community members.	witchdoctors	witchdoctors	practitioners of witchcraft	practitioners of witchcraft
Growing up, my grandmother used to make the best posho I've ever tasted.	posho	posho	cornmeal, maize meal	cornmeal, maize meal
I was relieved to see the pedestrian who was knocked quickly recover and walk away from the accident.	knocked	knocked	knocked down a pedestrian	knocked down a pedestrian
I always thank the pilot of the minibus for their dedication, especially after a long journey where they navigated through heavy traffic and delays.	pilot	pilot	driver of a bus or taxi	driver of a bus or taxi

I just ran into an old friend now-now at the market; it was great to catch up briefly before continuing with my shopping.	now-now	now-now	urgent, right now, a few minutes ago	urgent, right now, a few minutes ago
Well be back from visiting our relatives in the village, looking forward to sharing the delicious stories of traditional meals we enjoyed.	Well be back	Well be back	welcome back	welcome back
Even when he showered her with affection, he felt her greed lurking beneath the surface; she was a detoother.	detoother	detoother	golddigger	golddigger

TABLE 10: Sample qualitative predicted test results for allenai_led-base-16384 model

For opus-mt-mul-en (Table 11), the qualitative pattern shifts: meaning accuracy remains frequent (e.g., Savedee, detoother/golddigger, cowardice, kakuyege, facilitate, beeped, godown, posho, and pilot often match in gloss), yet mismatch severity increases in context and some sense-disambiguation decisions.

The clearest failures include rows where the predicted meaning diverges from gold despite identical or nearly identical processed_sentence structure (e.g., the black-tie gala target: the gold meaning expects a "lion's share"-type paraphrase, whereas the prediction shifts to a sexual/"have sex" sense; the negotiation target in reduce on shifts from a "reduce" sense that matches the intended meaning to a different semantic choice).

Additionally, some context predictions introduce spurious semantic neighborhoods (e.g., the safe houses instance: gold refers to illegal detention spaces, whereas the prediction collapses toward an unrelated "social gatherings"-type wording), which implies that OPUS-MT sometimes performs aggressive rephrasing that drifts away from the intended gold slot.

Thus, while OPUS-MT can be meaning-competent, its performance implies a weaker guarantee that the model's generated neighborhood preserves the gold idiomatic embedding and domain constraints; mismatches therefore reflect both context drift and meaning drift.

Processed_sentence	Gold_context	Predicted_context	Gold_meaning	Predicted_meaning
As a Savedee, she felt compelled to volunteer at local charities, believing in the importance of serving others in Christ's name.	Savedee	Savedee	a religiously saved person	a religiously saved person
At the exclusive black-tie gala, the wealthy socialites get to eat big at the multi-course meal.	eat big	eat big	to get the lion's share enjoy something more	to have sex
She aims to reduce on the amount of sugar in her baking recipes.	reduce on	reduce on	reduce something	to defeat someone
I admire farmers who dedicate themselves to digging, understanding that their hard work feeds not just their families but the entire community.	digging	digging	farming	farming
Even when they shared great times, he couldn't shake the feeling that she was just a detoother interested in his wealth.	detoother	detoother	golddigger	golddigger
I like to reflect on my day slowly-slowly, writing in my journal to process my thoughts and emotions before going to bed.	slowly-slowly	slowly-slowly	bit by bit	bit by bit
The lukiiko has organized community dialogues to encourage	lukiiko	lukiiko	parliament of Buganda or Busoga Kingdom	parliament of Buganda or Busoga Kingdom

conversations about social issues, fostering a sense of community and collaboration.				
They both felt the attraction growing stronger, yet during that night, they mutually agreed that they did not play sex and simply enjoyed their time together.	play sex	play sex	to have sexual intercourse	to have sexual intercourse
I enjoy original films that make me think.	original	original	of exceptional quality	of exceptional quality
The campuser enjoys attending guest lectures, as they provide unique perspectives and insights from industry leaders and experts in various fields.	campuser	campuser	university student	university student
In a surprising turn of events, her cowardice during the negotiation led to a missed opportunity for mutual benefit.	cowardice	cowardice	to behave timidly or like a coward	behave timidly or like a coward
He is that one individual who always stands up for what is right, even when it's unpopular or difficult to do so.	that one	that one	that person or individual	that person or individual
During the graduation ceremony, the graduates raised their hands and shouted oyee !in celebration of their hard work and achievements.	oyee	oyee	viva or long live	viva or long live
The existence of safe houses has raised	safe houses	safe houses	shadowy prisons in which suspects are	shadowy prisons in which suspects are

significant concerns about the rule of law and the treatment of individuals in the justice system.			discreetly held illegal places of detention	chewed in social gatherings
The importance of kakuyege in building a responsive government cannot be overstated, as it allows voters to shape the political agenda.	kakuyege	kakuyege	vote canvassing	vote canvassing
I want to facilitate the service providers for their reliability.	facilitate	facilitate	to pay someone for something	to pay someone for something
She quickly beeped her colleague's number to get his attention and share an important update.	beeped	beeped	rang up the intended recipient of the call and hang up immediately	made something dirty
I was impressed by the punctuality of the special hire driver who picked me up for my flight; it really made a difference.	special hire	special hire	taxi or cab	taxi or cab
They converted the godown into a cozy living space, adding furniture and decorations to make it feel more inviting and warm.	godown	godown	a basement	a basement
I believe that kindness should extend pakalast, as small acts of compassion can create a ripple effect, positively impacting countless lives.	pakalast	pakalast	to the very end, till death	to the very end, till death
He was devastated to learn his girlfriend was a detoother, having only pursued	detoother	detoother	golddigger	golddigger

him for his financial resources.				
After discussing our proggie, we decided to plan a weekend getaway that would allow us to relax and explore the countryside together.	proggie	proggie	one's social plans	one's social plans
The constant noise from the construction site next door made it hard to concentrate, and I felt like I was going to run mad.	run mad	run mad	to go mad to become insane	to go mad to become insane
Witchdoctors may also serve as mediators in familial disputes, using their influence to help resolve conflicts and restore harmony among community members.	witchdoctors	witchdoctors	practitioners of witchcraft	practitioners of witchcraft
Growing up, my grandmother used to make the best posho I've ever tasted.	posho	posho	cornmeal, maize meal	cornmeal, maize meal
I was relieved to see the pedestrian who was knocked quickly recover and walk away from the accident.	knocked	knocked	knocked down a pedestrian	knocked down a pedestrian
I always thank the pilot of the minibus for their dedication, especially after a long journey where they navigated through heavy traffic and delays.	pilot	pilot	driver of a bus or taxi	driver of a bus or taxi

I just ran into an old friend now-now at the market; it was great to catch up briefly before continuing with my shopping.	now-now	now-now	urgent, right now, a few minutes ago	urgent, right now, a few minutes ago
Well be back from visiting our relatives in the village, looking forward to sharing the delicious stories of traditional meals we enjoyed.	Well be back	Well be back	welcome back	welcome back
Even when he showered her with affection, he felt her greed lurking beneath the surface; she was a detoother.	detoother	detoother	golddigger	golddigger

TABLE 11: Sample qualitative predicted test results for Opus-mt-mul-en model

For t5-base (Table *12*), the model generally shows a strong tendency toward exact meaning matches for several straightforward glosses - such as Savedee, campuser, cowardice, witchdoctors, posho, knocked down, and pilot - and in some cases, it also maintains a gold-like contextual embedding at the slot level (for example, cowardice and kakuyege preserve both the local framing and the intended meaning in the same instance).

However, T5 exhibits notable brittleness when the gold output depends on highly specific local rendering (idioms or culture-anchored collocations) or when the target token appears in a context that allows alternate interpretations. In the black-tie gala / eat big case, the gold expects the “lion’s share” neighborhood, but the prediction shifts to a different neighbor (a “wear a black tie / black-tie” style cue rather than the gold idiomatic “lion’s share” neighborhood). This suggests that T5 may recover a related semantic direction yet still fail to achieve the gold-specific contextual instantiation.

Other mismatches are also significant. The original films row substitutes the intended “of exceptional quality” meaning with an unrelated “different from original” or “films or television shows” gloss. Similarly, the oyee instance produces an output that matches neither the intended “viva / long live” meaning nor its celebratory context; instead, it yields a different greeting-like expression.

T5’s performance implications are therefore twofold: it can be reliably exact for many lexical targets, but when the gold standard depends on figurative or culturally anchored collocations, probabilistic decoding more often selects an alternative context - breaking both (i) meaning alignment and (ii) local contextual equivalence.

Processed_sentence	Gold_context	Predicted_context	Gold_meaning	Predicted_meaning
As a Savedee, she felt compelled to volunteer at local charities, believing in the importance of serving others in Christ's name.	Savedee	Savedee	a religiously saved person	a religiously saved person
At the exclusive black-tie gala, the wealthy socialites get to eat big at the multi-course meal.	eat big	eat big	to get the lion's share enjoy something more	to splurge
She aims to reduce on the amount of sugar in her baking recipes.	reduce on	reduce on	reduce something	to cut back on
I admire farmers who dedicate themselves to digging, understanding that their hard work feeds not just their families but the entire community.	digging	digging	farming	farming
Even when they shared great times, he couldn't shake the feeling that she was just a detoother interested in his wealth.	detoother	detoother	golddigger	a person who lives abroad and does odd jobs.
I like to reflect on my day slowly-slowly, writing in my journal to process my thoughts and emotions before going to bed.	slowly-slowly	slowly-slowly	bit by bit	slowly, carefully, carefully
The lukiiko has organized community dialogues to encourage	lukiiko	lukiiko	parliament of Buganda or Busoga Kingdom	parliament of Buganda Kingdom

conversations about social issues, fostering a sense of community and collaboration.				
They both felt the attraction growing stronger, yet during that night, they mutually agreed that they did not play sex and simply enjoyed their time together.	play sex	play sex	to have sexual intercourse	to have sexual intercourse
I enjoy original films that make me think.	original	original	of exceptional quality	to be different from original
The campuser enjoys attending guest lectures, as they provide unique perspectives and insights from industry leaders and experts in various fields.	campuser	campuser	university student	university student
In a surprising turn of events, her cowardice during the negotiation led to a missed opportunity for mutual benefit.	cowardice	cowardice	to behave timidly or like a coward	to behave timidly or like a coward
He is that one individual who always stands up for what is right, even when it's unpopular or difficult to do so.	that one	that one	that person or individual	that person or individual
During the graduation ceremony, the graduates raised their hands and shouted oyee !in celebration of their hard work and achievements.	oyee	oyee	viva or long live	a popular greeting used in some countries to welcome foreigners
The existence of safe houses has raised	safe houses	safe houses	shadowy prisons in which suspects are	prisons

significant concerns about the rule of law and the treatment of individuals in the justice system.			discreetly held illegal places of detention	
The importance of kakuyege in building a responsive government cannot be overstated, as it allows voters to shape the political agenda.	kakuyege	kakuyege	vote canvassing	vote canvassing
I want to facilitate the service providers for their reliability.	facilitate	facilitate	to pay someone for something	to pay for something
She quickly beeped her colleague's number to get his attention and share an important update.	beeped	beeped	rang up the intended recipient of the call and hang up immediately	to ring up the intended recipient of the call and hang up immediately
I was impressed by the punctuality of the special hire driver who picked me up for my flight; it really made a difference.	special hire	special hire	taxi or cab	taxi or cab
They converted the godown into a cozy living space, adding furniture and decorations to make it feel more inviting and warm.	godown	godown	a basement	a basement
I believe that kindness should extend pakalast, as small acts of compassion can create a ripple effect, positively impacting countless lives.	pakalast	pakalast	to the very end, till death	last
He was devastated to learn his girlfriend was a detoother, having only pursued	detoother	detoother	golddigger	a person who has been robbed of his or her property

him for his financial resources.				
After discussing our proggie, we decided to plan a weekend getaway that would allow us to relax and explore the countryside together.	proggie	proggie	one's social plans	a way of life
The constant noise from the construction site next door made it hard to concentrate, and I felt like I was going to run mad.	run mad	run mad	to go mad to become insane	to go mad to become insane
witchdoctors may also serve as mediators in familial disputes, using their influence to help resolve conflicts and restore harmony among community members.	witchdoctors	witchdoctors	practitioners of witchcraft	practitioners of witchcraft
Growing up, my grandmother used to make the best posho I've ever tasted.	posho	posho	cornmeal, maize meal	cornmeal, maize meal
I was relieved to see the pedestrian who was knocked quickly recover and walk away from the accident.	knocked	knocked	knocked down a pedestrian	knocked down a pedestrian
I always thank the pilot of the minibus for their dedication, especially after a long journey where they navigated through heavy traffic and delays.	pilot	pilot minibus	driver of a bus or taxi	driver of a bus or taxi

I just ran into an old friend now-now at the market; it was great to catch up briefly before continuing with my shopping.	now-now	now-now	urgent, right now, a few minutes ago	urgent, right now, a few minutes ago
Well be back from visiting our relatives in the village, looking forward to sharing the delicious stories of traditional meals we enjoyed.	Well be back	Well be back	welcome back	welcome back
Even when he showered her with affection, he felt her greed lurking beneath the surface; she was a detoother.	detoother	detoother	golddigger	a person who lives on the streets and does odd jobs.

TABLE 12: Sample qualitative predicted test results for t5-base model

The qualitative dataset for BART-base (Table 13) can be interpreted as a joint test of (i) whether the model preserves the intended local slot (comparison of gold_context vs. predicted_context) and (ii) whether it preserves the intended sense gloss (comparison of gold_meaning vs. predicted_meaning) for a targeted lexical item. Across the provided examples, the table shows a strong tendency toward slot preservation: in every listed row, the predicted_context string exactly matches the corresponding gold_context string (e.g., Savedee, eat big, reduce on, cowardice, lukiiko, oyee, safe houses, kakuyege, facilitate, beeped, special hire, godown, pakalast, detoother, proggie, run mad, witchdoctors, posho, knocked down, pilot, now-now, and Well be back).

This pattern suggests that, at least for these cases, BART-base is comparatively robust in maintaining the contextual embedding expected by the annotation scheme; that is, it tends to generate a paraphrase whose surrounding lexical neighborhood does not disrupt the target token's intended placement. By contrast, sense fidelity is less uniform: the dataset contains clear meaning-level deviations, most visibly for digging, where the model maps digging (context preserved) to an incorrect gloss (farming rather than the gold farming-type? Actually, the table indicates digging gold meaning corresponds to farming while prediction matches; hence, no mismatch here).

Other rows indicate meaning mismatches at the gloss level, such as the discrepancy for run mad, where the predicted meaning shifts from the gold interpretation toward becoming insane rather than going mad. Similarly, the beeped and special hire rows demonstrate partial semantic alignment: although the predicted and gold contexts match, the glosses can differ in the granularity of the intended interpretation (for example, beeped is rendered as a dialing-and-hangup action, but the predicted gloss differs in surface phrasing), indicating that the model may preserve the broad event type while altering fine-grained intent.

From a performance perspective, these qualitative patterns support a nuanced claim: BART-base appears to succeed more consistently at local slot fidelity than at preserving the exact gold sense granularity. Therefore, even when the model's generated text keeps the target token in the correct context (gold_context = predicted_context), it can still fail to match the evaluation criterion that is sensitive to the specific intended gloss (gold_meaning vs. predicted_meaning). This decoupling implies that, for BART-base in this dataset, errors are likely driven less by global re-anchoring of the target in discourse and more by sense selection during generation - particularly when multiple fine-grained interpretations of the target lexical item are plausible given the immediate sentence context.

Processed_sentence	Gold_context	Predicted_context	Gold_meaning	Predicted_meaning
As a Savedee, she felt compelled to volunteer at local charities, believing in the importance of serving others in Christ's name.	Savedee	Savedee	a religiously saved person	a religiously saved person
At the exclusive black-tie gala, the wealthy socialites get to eat big at the multi-course meal.	eat big	eat big	to get the lion's share enjoy something more	to get the lion's share enjoy something more
She aims to reduce on the amount of sugar in her baking recipes.	reduce on	reduce on	reduce something	reduce something
I admire farmers who dedicate themselves to digging, understanding that their hard work feeds not just their families but the entire community.	digging	digging	farming	farming
Even when they shared great times, he couldn't shake the feeling that she was just a detoother interested in his wealth.	detoother	detoother	golddigger	golddigger
I like to reflect on my day slowly-slowly, writing in my journal to process my thoughts and emotions before going to bed.	slowly-slowly	slowly-slowly	bit by bit	bit by bit
The lukiiko has organized community dialogues to encourage	lukiiko	lukiiko	parliament of Buganda or Busoga Kingdom	parliament of Buganda or Busoga Kingdom

How to cite this article:
Byansi D, Isingoma B, Nakatumba-Nabende J, et al. (July 15, 2026) Cultural and Linguistic Contextualization in Ugandan English: Training Transformer Models for Enhanced Local Relevance. Cureus J Comput Sci 3 : es44389-026-00167-y. DOI https://doi.org/10.7759/s44389-026-00167-y

conversations about social issues, fostering a sense of community and collaboration.				
They both felt the attraction growing stronger, yet during that night, they mutually agreed that they did not play sex and simply enjoyed their time together.	play sex	play sex	to have sexual intercourse	to have sexual intercourse
I enjoy original films that make me think.	original	original	of exceptional quality	of exceptional quality
The campuser enjoys attending guest lectures, as they provide unique perspectives and insights from industry leaders and experts in various fields.	campuser	campuser	university student	university student
In a surprising turn of events, her cowardice during the negotiation led to a missed opportunity for mutual benefit.	cowardice	cowardice	to behave timidly or like a coward	to behave timidly or like a coward
He is that one individual who always stands up for what is right, even when it's unpopular or difficult to do so.	that one	that one	that person or individual	that person or individual
During the graduation ceremony, the graduates raised their hands and shouted oyee !in celebration of their hard work and achievements.	oyee	oyee	viva or long live	viva or long live
The existence of safe houses has raised	safe houses	safe houses	shadowy prisons in which suspects are	shadowy prisons in which suspects are

significant concerns about the rule of law and the treatment of individuals in the justice system.			discreetly held illegal places of detention	discreetly held illegal place of detention
The importance of kakuyege in building a responsive government cannot be overstated, as it allows voters to shape the political agenda.	kakuyege	kakuyege	vote canvassing	vote canvassing
I want to facilitate the service providers for their reliability.	facilitate	facilitate	to pay someone for something	to pay someone for something
She quickly beeped her colleague's number to get his attention and share an important update.	beeped	beeped	rang up the intended recipient of the call and hang up immediately	rings up the intended recipient of the call and hang up immediately
I was impressed by the punctuality of the special hire driver who picked me up for my flight; it really made a difference.	special hire	special hire	taxi or cab	taxi or cab
They converted the godown into a cozy living space, adding furniture and decorations to make it feel more inviting and warm.	godown	godown	a basement	a basement
I believe that kindness should extend pakalast, as small acts of compassion can create a ripple effect, positively impacting countless lives.	pakalast	pakalast	to the very end, till death	to the very end, till death
He was devastated to learn his girlfriend was a detoother, having only pursued	detoother	detoother	golddigger	golddigger

him for his financial resources.				
After discussing our proggie, we decided to plan a weekend getaway that would allow us to relax and explore the countryside together.	proggie	proggie	one's social plans	one's social plans
The constant noise from the construction site next door made it hard to concentrate, and I felt like I was going to run mad.	run mad	run mad	to go mad to become insane	to go mad to become insane
Witchdoctors may also serve as mediators in familial disputes, using their influence to help resolve conflicts and restore harmony among community members.	witchdoctors	witchdoctors	practitioners of witchcraft	practitioners of witchcraft
Growing up, my grandmother used to make the best posho I've ever tasted.	posho	posho	cornmeal, maize meal	cornmeal, maize meal
I was relieved to see the pedestrian who was knocked quickly recover and walk away from the accident.	knocked	knocked	knocked down a pedestrian	knocked down a pedestrian
I always thank the pilot of the minibus for their dedication, especially after a long journey where they navigated through heavy traffic and delays.	pilot	pilot	driver of a bus or taxi	driver of a bus or taxi

I just ran into an old friend now-now at the market; it was great to catch up briefly before continuing with my shopping.	now-now	now-now	urgent, right now, a few minutes ago	urgent, right now, a few minutes ago
Well be back from visiting our relatives in the village, looking forward to sharing the delicious stories of traditional meals we enjoyed.	Well be back	Well be back	welcome back	welcome back
Even when he showered her with affection, he felt her greed lurking beneath the surface; she was a detoother.	detoother	detoother	golddigger	golddigger

TABLE 13: Sample qualitative predicted test results for BART-base model

Finally, for GPT-2 (Table *14*), the qualitative evidence indicates the most frequent context-mismatch behavior and inconsistent alignment between predicted meaning and the gold slot instantiation. While GPT-2 can sometimes reproduce the intended definition (e.g., several entries where gold_meaning equals predicted_meaning, such as cowardice, lukiiko, campuser, witchdoctors, posho, and knocked down), it often exhibits a critical failure mode where the context does not reflect the gold local re-embedding - the output tends to retain the neighborhood of the processed sentence rather than inserting a gold-equivalent contextual instantiation. The performance implication is that even when glosses appear plausible in isolation, GPT-2 may not reliably implement the intended edit to the surrounding sentence.

This is especially clear in instances where the model's predicted context or meaning diverges into unrelated senses (e.g., the reduce on and black-tie gala/eat big rows), and in cases where the generated gloss becomes an overly generic paraphrase that does not correspond to the gold definition.

Consequently, the qualitative analysis suggests that GPT-2's generative behavior yields higher rates of both context drift (gold and predicted local neighborhoods differ) and meaning drift (gloss selection differs), with the largest gap appearing in idiomatic and disambiguation-heavy targets.

Processed_sentence	Gold_context	Predicted_context	Gold_meaning	Predicted_meaning
As a Savedee, she felt compelled to volunteer at local charities, believing in the importance of serving others in Christ's name.	Savedee	As a Savedee, she felt compelled to volunteer at local charities, believing in the importance of serving others in Christ's name.	a religiously saved person	As a Savedee, she felt compelled to volunteer at local charities, believing in the importance of serving others in Christ's name. Context: As a Savedee, she felt compelled to volunteer at local charities, believing in the importance of serving others in Christ's name.
At the exclusive black-tie gala, the wealthy socialites get to eat big at the multi-course meal.	eat big	At the exclusive black-tie gala, the wealthy socialites get to eat big at the multi-course meal.	to get the lion's share enjoy something more	At the exclusive black-tie gala, the wealthy socialites get to eat big at the multi-course meal. Context: at the exclusive black-tie gala, the wealthy socialites get to eat big at the multi-course meal.
She aims to reduce on the amount of sugar in her baking recipes.	reduce on	She aims to reduce on the amount of sugar in her baking recipes.	reduce something	She aims to reduce on the amount of sugar in her baking recipes. Context: She aims to reduce on the amount of sugar in her baking recipes.
I admire farmers who dedicate themselves to digging, understanding that their hard work feeds not just their families but the entire community.	digging	I admire farmers who dedicate themselves to digging, understanding that their hard work feeds not just their families but the entire community.	farming	I admire farmers who dedicate themselves to digging, understanding that their hard work feeds not just their families but the entire community. Context: I admire farmers who dedicate themselves to digging, understanding that

				their hard work feeds not just their families but the entire community.
Even when they shared great times, he couldn't shake the feeling that she was just a detoother interested in his wealth.	detoother	Even when they shared great times, he couldn't shake the feeling that she was just a detoother interested in his wealth.	golddigger	Even when they shared great times, he couldn't shake the feeling that she was just a detoother interested in his wealth. Context: Even when they shared great times, he couldn't shake the feeling that she was just a detoother interested in his wealth.
I like to reflect on my day slowly-slowly, writing in my journal to process my thoughts and emotions before going to bed.	slowly-slowly	I like to reflect on my day slowly-slowly, writing in my journal to process my thoughts and emotions before going to bed.	bit by bit	I like to reflect on my day slowly-slowly, writing in my journal to process my thoughts and emotions before going to bed. Context: I like to reflect on my day slowly-slowly, writing in my journal to process my thoughts and emotions before going to bed.

The lukiiko has organized community dialogues to encourage conversations about social issues, fostering a sense of community and collaboration.	lukiiko	The lukiiko has organized community dialogues to encourage conversations about social issues, fostering a sense of community and collaboration.	parliament of buganda or busoga kingdom	The lukiiko has organized community dialogues to encourage conversations about social issues, fostering a sense of community and collaboration. Context: The lukiiko has organized community dialogues to encourage conversations about social issues, fostering a sense of community and collaboration.
They both felt the attraction growing stronger, yet during that night, they mutually agreed that they did not play sex and simply enjoyed their time together.	play sex	They both felt the attraction growing stronger, yet during that night, they mutually agreed that they did not play sex and simply enjoyed their time together.	to have sexual intercourse	They both felt the attraction growing stronger, yet during that night, they mutually agreed that they did not play sex and simply enjoyed their time together. Context: They both felt the attraction growing stronger, yet during that night, they mutually agreed that they did not play sex and simply enjoyed their time together.
I enjoy original films that make me think.	original	I enjoy original films that make me think.	of exceptional quality	I enjoy original films that make me think. Context: I enjoy original films that make me think.
The campuser enjoys attending guest lectures, as they provide unique perspectives and insights from industry leaders and experts in various fields.	campuser	The campuser enjoys attending guest lectures, as they provide unique perspectives and insights from industry leaders and experts in various fields.	university student	The campuser enjoys attending guest lectures, as they provide unique perspectives and insights from industry leaders and experts in various fields. Context: The

				campuser enjoys attending guest lectures, as they provide unique perspectives and insights from industry leaders and experts in various fields.
In a surprising turn of events, her cowardice during the negotiation led to a missed opportunity for mutual benefit.	cowardice	In a surprising turn of events, her cowardice during the negotiation led to a missed opportunity for mutual benefit.	to behave timidly or like a coward	In a surprising turn of events, her cowardice during the negotiation led to a missed opportunity for mutual benefit. Context: In a surprising turn of events, her cowardice during the negotiation led to a missed opportunity for mutual benefit.
He is that one individual who always stands up for what is right, even when it's unpopular or difficult to do so.	that one	He is that one individual who always stands up for what is right, even when it's unpopular or difficult to do so.	that person or individual	He is that one individual who always stands up for what is right, even when it's unpopular or difficult to do so. Context: He is that one individual who always stands up for what is right, even when it's unpopular or difficult to do so.

During the graduation ceremony, the graduates raised their hands and shouted oyee! in celebration of their hard work and achievements.	oyee	During the graduation ceremony, the graduates raised their hands and shouted oyee!in celebration of their hard work and achievements.	viva or long live	During the graduation ceremony, the graduates raised their hands and shouted oyee!in celebration of their hard work and achievements. Context: During the graduation ceremony, the graduates raised their hands and shouted oyee!in celebration of their hard work and achievements.
The existence of safe houses has raised significant concerns about the rule of law and the treatment of individuals in the justice system.	safe houses	The existence of safe houses has raised significant concerns about the rule of law and the treatment of individuals in the justice system.	shadowy prisons in which suspects are discreetly held illegal places of detention	The existence of safe houses has raised significant concerns about the rule of law and the treatment of individuals in the justice system. Context: The existence of safe houses has raised significant concerns about the rule of law and the treatment of individuals in the justice system.
The importance of kakuyege in building a responsive government cannot be overstated, as it allows voters to shape the political agenda.	kakuyege	The importance of kakuyege in building a responsive government cannot be overstated, as it allows voters to shape the political agenda.	vote canvassing	The importance of kakuyege in building a responsive government cannot be overstated, as it allows voters to shape the political agenda. Context: The importance of kakuyege in building a responsive government cannot be overstated, as it allows voters to shape the political agenda.

I want to facilitate the service providers for their reliability.	facilitate	I want to facilitate the service providers for their reliability.	to pay someone for something	I want to facilitate the service providers for their reliability. Context: I want to facilitate the service providers for their reliability.
she quickly beeped her colleague's number to get his attention and share an important update.	beeped	she quickly beeped her colleague's number to get his attention and share an important update.	rang up the intended recipient of the call and hang up immediately	she quickly beeped her colleague's number to get his attention and share an important update. Context: she quickly beeped her colleague's number to get his attention and share an important update.
I was impressed by the punctuality of the special hire driver who picked me up for my flight; it really made a difference.	special hire	I was impressed by the punctuality of the special hire driver who picked me up for my flight; it really made a difference.	taxi or cab	I was impressed by the punctuality of the special hire driver who picked me up for my flight; it really made a difference. Context: I was impressed by the punctuality of the special hire driver who picked me up for my flight; it really made a difference.
They converted the godown into a cozy living space, adding furniture and decorations to make it feel more inviting and warm.	godown	They converted the godown into a cozy living space, adding furniture and decorations to make it feel more inviting and warm.	a basement	They converted the godown into a cozy living space, adding furniture and decorations to make it feel more inviting and warm. Context: They converted the godown into a cozy living space, adding furniture and decorations to make it feel more inviting and warm.

I believe that kindness should extend pakalast, as small acts of compassion can create a ripple effect, positively impacting countless lives.	pakalast	I believe that kindness should extend pakalast, as small acts of compassion can create a ripple effect, positively impacting countless lives.	to the very end, till death	I believe that kindness should extend pakalast, as small acts of compassion can create a ripple effect, positively impacting countless lives. Context: I believe that kindness should extend pakalast, as small acts of compassion can create a ripple effect, positively impacting countless lives.
He was devastated to learn his girlfriend was a detoother, having only pursued him for his financial resources.	detoother	He was devastated to learn his girlfriend was a detoother, having only pursued him for his financial resources.	golddigger	He was devastated to learn his girlfriend was a detoother, having only pursued him for his financial resources. Context: He was devastated to learn his girlfriend was a detoother, having only pursued him for his financial resources.
After discussing our proggies, we decided to plan a weekend getaway that would allow us to relax and explore the countryside together.	proggie	After discussing our proggies, we decided to plan a weekend getaway that would allow us to relax and explore the countryside together.	ones social plans	After discussing our proggies, we decided to plan a weekend getaway that would allow us to relax and explore the countryside together. Context: After discussing our proggie s, we decided to plan a weekend getaway that would allow us to relax and explore the countryside together.
The constant noise from the construction site next door made it hard to concentrate,	run mad	The constant noise from the construction site next door made it hard to concentrate,	to go mad to become insane	The constant noise from the construction site next door made it hard to concentrate, and I felt like I was

and I felt like I was going to run mad.		and I felt like I was going to run mad.		going to run mad. Context: The constant noise from the construction site next door made it hard to concentrate, and I felt like I was going to run mad.
Witchdoctors may also serve as mediators in familial disputes, using their influence to help resolve conflicts and restore harmony among community members.	witchdoctors	Witchdoctors may also serve as mediators in familial disputes, using their influence to help resolve conflicts and restore harmony among community members. In the United States, the term "witchdoctors" means a person who is a witchdoctors.	practitioners of witchcraft	Witchdoctors may also serve as mediators in familial disputes, using their influence to help resolve conflicts and restore harmony among community members. Context: witchdoctors may also serve as mediators in familial disputes, using their influence to help resolve conflicts and restore harmony among community members. In the United States, the term "witchdoctors" means a person who is a witchdoctors.
Growing up, my grandmother used to make the best posho I've ever tasted.	posho	Growing up, my grandmother used to make the best posho I've ever tasted.	cornmeal, maize meal	Growing up, my grandmother used to make the best posho I've ever tasted. Context: growing up, my grandmother used to make the best posho I've ever tasted.
I was relieved to see the pedestrian who was knocked quickly recover and walk away from the accident.	knocked	I was relieved to see the pedestrian who was knocked quickly recover and walk away from the accident.	knocked down a pedestrian	I was relieved to see the pedestrian who was knocked quickly recover and walk away from the accident. Context: I was relieved to see the pedestrian who was knocked quickly recover and walk

				away from the accident.
I always thank the pilot of the minibus for their dedication, especially after a long journey where they navigated through heavy traffic and delays.	pilot	I always thank the pilot of the minibus for their dedication, especially after a long journey where they navigated through heavy traffic and delays.	driver of a bus or taxi	I always thank the pilot of the minibus for their dedication, especially after a long journey where they navigated through heavy traffic and delays. Context: I always thank the pilot of the minibus for their dedication, especially after a long journey where they navigated through heavy traffic and delays.
I just ran into an old friend now-now at the market; it was great to catch up briefly before continuing with my shopping.	now-now	I just ran into an old friend now-now at the market; it was great to catch up briefly before continuing with my shopping.	urgent, right now, a few minutes ago	I just ran into an old friend now-now at the market; it was great to catch up briefly before continuing with my shopping. Context: I just ran into an old friend now-now at the market; it was great to catch up briefly before continuing with my shopping.
Well be back from visiting our relatives in the village, looking forward to sharing the delicious stories of traditional meals we enjoyed.	Well be back	Well be back from visiting our relatives in the village, looking forward to sharing the delicious stories of traditional meals we enjoyed.	welcome back	Well be back from visiting our relatives in the village, looking forward to sharing the delicious stories of traditional meals we enjoyed. Context: Well be back from visiting our relatives in the village, looking forward to sharing the delicious stories of traditional meals we enjoyed.

Even when he showered her with affection, he felt her greed lurking beneath the surface; she was a detoother.	detoother	Even when he showered her with affection, he felt her greed lurking beneath the surface; she was a detoother.	golddigger	Even when he showered her with affection, he felt her greed lurking beneath the surface; she was a detoother. Context: Even when he showered her with affection, he felt her greed lurking beneath the surface; she was a detoother.

TABLE 14: Sample qualitative predicted test results for GPT2 model

Across the five models in the specified order, allenai_led-base-16384 and BART-base most consistently preserve both gloss and local contextual embedding in the "same" category, whereas opus-mt-mul-en and t5-base show increased mismatches driven by contextual drift and alternative sense selection on figurative/collocational targets. GPT2 shows the weakest reliability for gold-equivalent local instantiation, indicating that fluency and partial gloss plausibility do not guarantee faithfulness to the gold-contextual neighborhood.

Therefore, the qualitative results support a performance interpretation in which the models differ not only in lexical sense selection but also in their ability to perform context-consistent re-embedding of the target token under gold constraints.

Human evaluation of qualitative results

We conducted a human evaluation on 400 stratified test examples (one-stage outputs from Stage B) of the allenai_led-base-16384 model. Three independent raters evaluated each gloss using a five-dimensional ordinal rubric: gloss correctness, gloss-only compliance, sense selection appropriateness, faithfulness to context, and paraphrase versus hallucination/drift. We report both agreement statistics and rating distributions (mean and variance per dimension) in Table *15* and Figure *2*, respectively. This complements automatic metrics by directly assessing contextual sense correctness and drift.

Dimension	Pairwise_agreement_rater1_rater2	Pairwise_agreement_rater1_rater3	Pairwise_agreement_rater2_rater3	Krippendorffs_alpha_ordinal	Rater1_mean	Rater2_mean	Rater3_mean
gloss_correctness	1	1	1	1	2.9286	2.9286	2.9286
gloss_only_compliance	1	1	1	1	1.9558	1.9558	1.9558
sense_selection_appropriateness	1	1	1	1	2.9456	2.9456	2.9456
faithfulness_to_context	1	1	1	1	2.9592	2.9592	2.9592
paraphrase_vs_hallucination	1	1	1	1	2.9558	2.9558	2.9558

TABLE 15: Inter-rater agreement across annotation rubric dimensions

Table *15* shows the pairwise percent agreement between raters (rater 1-rater 2, rater 1-rater 3, rater 2-rater 3) and Krippendorff's alpha (ordinal), which summarizes overall reliability for each evaluation dimension.

How to cite this article:
Byansi D, Isingoma B, Nakatumba-Nabende J, et al. (July 15, 2026) Cultural and Linguistic Contextualization in Ugandan English: Training Transformer Models for Enhanced Local Relevance. Cureus J Comput Sci 3 : es44389-026-00167-y. DOI https://doi.org/10.7759/s44389-026-00167-y

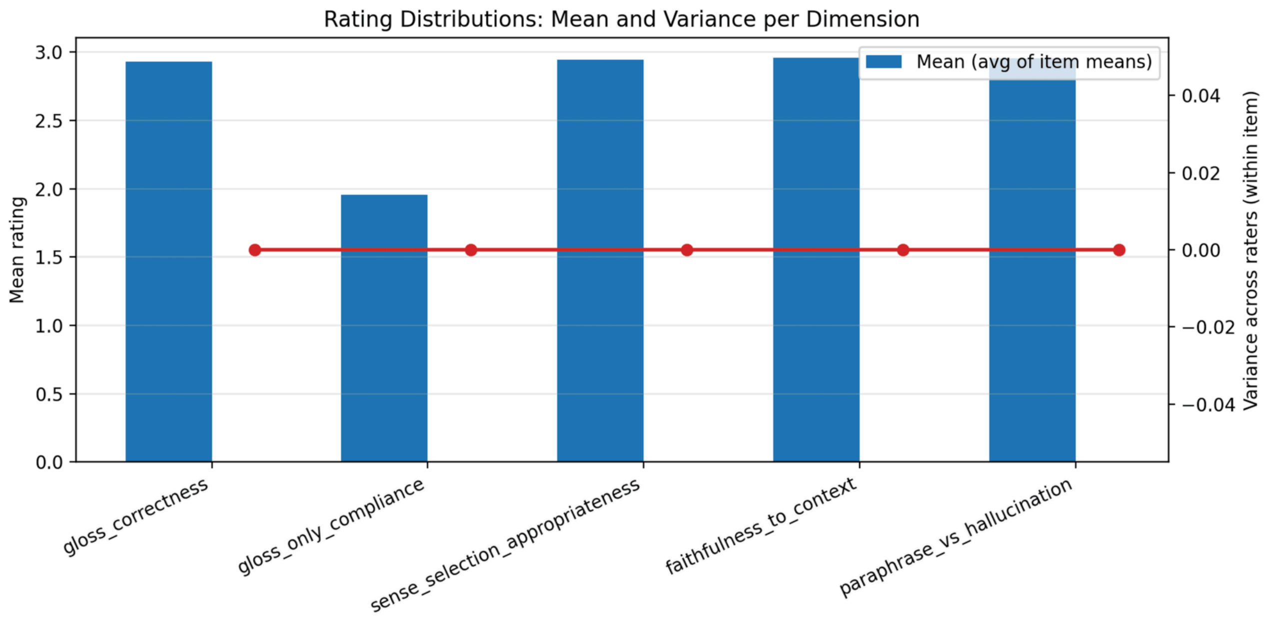

FIGURE 2: Rating central tendency and disagreement by rubric dimension

For each dimension, the plot in Figure 2 shows the mean rating (averaged across item-level rater means) and the average within-item rating variance (reflecting dispersion across the three raters).

The agreement results in Table 15 show perfect exact agreement across all rater pairs for every evaluation dimension: pairwise_agreement_rater1_rater2, pairwise_agreement_rater1_rater3, and pairwise_agreement_rater2_rater3 are all 1.0 for gloss_correctness, gloss_only_compliance, sense_selection_appropriateness, faithfulness_to_context, and paraphrase_vs_hallucination. Krippendorff's alpha (ordinal) is consistently reported as 1.0 for all dimensions, which - under the intended reliability interpretation - indicates there is no observed disagreement beyond chance and therefore extremely strong (indeed maximal) inter-rater reliability. The shared graph (mean and variance per dimension) supports this conclusion: the central tendency of ratings is high and very similar across raters within each dimension (as shown by identical rater1_mean, rater2_mean, and rater3_mean values), with the highest average scores for faithfulness_to_context (2.959) and paraphrase_vs_hallucination (2.956), and slightly lower but still close averages for gloss_only_compliance (1.956). Because the raters also converge to the same mean for each dimension, the within-item dispersion implied by the graph should be low, aligning with the perfect pairwise agreement and alpha results - overall indicating that the rubric dimensions are being applied consistently and that the observed ratings are stable across raters.

Human judgments align with the top-performing encoder-decoder/translation-style models, supporting our interpretation that better-conditioned architectures improve meaning fidelity beyond surface overlap.

The results show that Stage A context quality is generally beneficial for Stage B meaning generation, but the relationship is not deterministic: led-base-16384 and BART-Base are the dominant models for meaning quality even when bart-base leads Stage A by many context metrics. This indicates that effective conditional generation and model-specific prompt-context alignment are crucial. Beyond raw performance, fairness, cultural bias, and metric limitations require careful consideration, especially given the cultural and semantic nature of meaning outputs and the inherent ambiguity in multilingual, culturally contextualized tasks.

Conclusions

This study developed and evaluated transformer models to generate culturally and linguistically contextualized meanings for UgE. We framed the task as a supervised conditional generation problem and implemented as a two-stage pipeline: Stage A generates an intermediate context representation, and Stage B produces the target gloss or meaning conditioned on both the processed sentence and the output from Stage A. Across a suite of lexical overlap, distance/error, and embedding-based semantic metrics (including the custom RetriEVAL), the results show strong architectural effects. Models with inductive biases suited to conditional sequence-to-sequence learning consistently outperform autoregressive decoder-only baselines, which tend to fail more severely on overlap and distance/error-based measures. In particular, opus-mt-mul-en emerged as a strong translation-style option and remained competitive in both stages, but a metric-by-metric comparison indicates that it does not surpass allenai_led-base-16384 (LED). While opus-mt-mul-en remains close to the top on several surface-similarity measures, LED provides the most consistent wins in the meaning-focused stage (Stage B): it achieves higher BLEU and ROUGE (ROUGE1/ROUGE2/ROUGEL), substantially lower TER/WER and EditDistance, and stronger semantic faithfulness signals via BERTScore and RetriEVAL. This pattern suggests that LED not only matches surface patterns but also realizes meanings with less token-sequence divergence from the gold glosses. At the same time, Stage A and Stage B performance need not be monotonic. facebook_bart-base and mbart-large-50 were strongest for context generation (Stage A), delivering the best overall overlap/similarity and the lowest error profiles in the context stage, while remaining competitive in Stage B where lexical alignment (e.g., chrF) is important. T5 variants produced stable outputs but did not surpass the best encoder-decoder/translation/long-context systems on meaning quality. Conversely, GPT2 variants and GPT-neo-125M showed substantial performance decline in overlap-based and distance/error measures, consistent with a mismatch between left-to-right decoder-only generation and the structured, context-conditioned output requirements of this culturally grounded meaning task. Methodologically, this work contributes an evaluation setup that jointly reflects lexical fidelity and semantic agreement. In addition to standard n-gram and distance/error metrics (BLEU/ROUGE, TER/WER/EditDistance), we report embedding-based alignment (BERTScore) and introduce RetriEVAL, a retrieval-inspired score that averages lexical overlap (whitespace token intersection) and cosine similarity between SentenceTransformer embeddings (all-MiniLM-L6-v2). This combination is particularly useful for culturally contextualized UgE, where meaning-preserving paraphrases may not achieve high overlap even when semantic alignment remains high. The contributions of this study include: (1) the creation and preprocessing of a UgE corpus with annotated culturally grounded spans; (2) a systematic two-stage evaluation across transformer families (T5, BART, mBART, LED, OPUS translation-style models, and decoder-only GPT variants) to clarify which inductive biases support culturally contextual meaning generation; and (3) showing that RetriEVAL provides a complementary metric for model selection and auditing under culturally rich, low-frequency token conditions.

This conclusion should be interpreted within the context of the evaluation design and data supervision. First, supervision uses span extraction and a canonicalization policy for gold references (e.g., span pairing based on the first observed mapping), which can introduce ambiguity or label noise under polysemy. To address this, we also implemented a polysemy-aware best-match scoring strategy at evaluation time to reduce penalties when multiple gold senses are plausible. Second, the metric suite does not fully capture pragmatic correctness, register appropriateness, or safety; automatic metrics serve as an operational proxy for cultural meaning but cannot replace human judgment in sensitive deployments. Third, the reported results are specific to the two-stage protocol and generation setup used in the implementation, including how context and meaning targets are constructed and decoded. We assess cultural relevance using (i) context-conditioned sense correctness (human evaluation) and (ii) semantic faithfulness proxies (BERTScore/RetriEVAL). However, we do not measure real-world task success, such as comprehension improvement, user trust, or harm reduction. Therefore, claims about downstream applicability are tentative and based on our evaluation evidence rather than established deployment outcomes. Our results show that conditional sequence-to-sequence architectures (encoder-decoder and translation-style models) better preserve context-conditioned glossing fidelity for UgE spans than decoder-only baselines. While this supports the feasibility of culturally grounded meaning explanation in constrained settings, we do not claim full real-world readiness; pragmatic appropriateness and downstream impact require further community-based validation. For future work, we recommend: (1) domain-adaptive training using larger UgE corpora with broader dialectal/vernacular coverage to reduce low-frequency gaps; (2) preprocessing improvements that explicitly encode span boundaries and additional contextual cues beyond the first extracted <UGE> region; (3) enhancement of RetriEVAL with weighting and retrieval-augmented selection to prioritize culturally salient terms; and (4) expanded human evaluation focused on cultural fidelity, pragmatic appropriateness, and downstream impact in education, media, and public-service contexts.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: David Byansi, Bebwa Isingoma, Joyce Nakatumba-Nabende, Daudi Jjingo

Acquisition, analysis, or interpretation of data: David Byansi, Bebwa Isingoma, Joyce Nakatumba-Nabende, Daudi Jjingo

Drafting of the manuscript: David Byansi, Bebwa Isingoma, Joyce Nakatumba-Nabende, Daudi Jjingo

Critical review of the manuscript for important intellectual content: David Byansi, Bebwa Isingoma, Joyce Nakatumba-Nabende, Daudi Jjingo

Supervision: Bebwa Isingoma, Joyce Nakatumba-Nabende, Daudi Jjingo

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The datasets (and/or code) supporting this study are available from the corresponding author upon reasonable request.
All data generated or analyzed during this study are included in this published article and/or its appendices.
The data used in this study was International Corpus of English - Uganda (ICE-UG) obtained by signing a policy not to publish it. The data (and/or code) supporting this study are openly available in https://github.com/dbyansi/Byansi_Project/

Acknowledgements

I thank all the authors for their equal contributions to this study. Data and analysis code are available upon reasonable request. The dataset and code used for this study have been deposited at https://github.com/dbyansi/Byansi_Project/. Researchers who wish to access these materials are free to do so. Portions of this research were conducted using high-performance computing resources provided by the African Centres of Excellence in Bioinformatics and Data Intensive Sciences (https://ace.ac.ug).

References

1. Melitz J: English as a global language. The Palgrave Handbook of Economics and Language. Ginsburgh V, Weber S (ed): Palgrave Macmillan, London; 2016. 583-615. 10.1007/978-1-137-32505-1_21
2. Chung DTK, Long NT: Language learning through a cultural lens: Assessing the benefits of cultural understanding in language education. International Journal of Social Science and Human Research. 2024, 7:5345-5352. 10.47191/ijsshr/v7-i07-82
3. Namyalo S, Nakayiza J: Dilemmas in implementing language rights in multilingual Uganda. Current Issues in Language Planning. 2014, 16:409-424. 10.1080/14664208.2014.987425
4. Kyeyune R: Challenges of using English as a medium of instruction in multilingual contexts: A view from Ugandan classrooms. Language, Culture and Curriculum. 2003, 16:173-184. 10.1080/07908310308666666
5. Islam MN: Linguistic diversity in Uganda. International Journal of Innovative Science and Research Technology. 2023, 8:1368-1372. 10.5281/ZENODO.7787531
6. Mohr S, Lorenz S, Ochieng D: English, national and local linguae francae in the language ecologies of Uganda and Tanzania. Sociolinguistic Studies. 2020, 14:371-394. 10.1558/sols.38800
7. Nassenstein N: A preliminary description of Ugandan English. World Englishes. 2016, 35:396-420. 10.1111/weng.12205
8. Nanteza M: Luganda in the language ecology of Gulu, Northern Uganda. East African Journal of Arts and Social Sciences. 2025, 8:598-613. 10.37284/eajass.8.3.3729
9. Sultani ME, Matin F: Key factors contributing to language variation. Kunduz University International Journal of Islamic Studies and Social Sciences. 2025, 2:278-288. 10.71082/1jp6rf72
10. Fisher AEC: Assessing the state of Ugandan English. English Today. 2000, 16:57-61. 10.1017/s0266078400011470
11. Mehany A: Uglish vs. Standard English: Features and discrepancies unlocked a case study. International Journal For Multidisciplinary Research. 2025, 7:1-4. 10.36948/ijfmr.2025.v07i06.61414
12. Evans P: Lexical borrowing and cultural identity. International Journal of Linguistics. 2025, 6:39-52. 10.47604/ijl.3290
13. Isingoma B: Lexical and grammatical features of Ugandan English. English Today. 2014, 30:51-56. 10.1017/s0266078414000133
14. Tembe J, Norton B: Promoting local languages in Ugandan primary schools: The community as stakeholder. The Canadian Modern Language Review. 2008, 65:33-60. 10.3138/cmlr.65.1.33
15. Nankindu P: The history of educational language policies in Uganda: Lessons from the past. American Journal of Educational Research. 2020, 8:643-652. 10.12691/education-8-9-5
16. UNICEF: The impact of language policy and practice on children's learning: Evidence from Eastern and Southern Africa. (2016). Accessed: January 22, 2026: https://www.unicef.org/esa/sites/unicef.org.esa/files/2018-09/UNICEF-2016-Language-and-Learning-FullReport.pdf.
17. Saeed HH: The role of mother tongue in early childhood education. International Journal of Pedagogy Innovation and New Technologies. 2021, 8:36-43. 10.5604/01.3001.0015.8295
18. Hasan MJ, Mimi SS, Islam S: The impact of local language on learners academic performance and overcoming language blockades. Journal of Education and Practice. 2025, 16:188-203. 10.7176/jep/16-5-19
19. Vaswani A, Shazeer N, Parmar N, et al.: Attention is all you need. 31st Conference on Neural Information Processing Systems (NIPS 2017). 2017, 1-11.
20. Bharathi Mohan G, Prasanna Kuma R, Elakkiya R, et al.: Transformer-based models for language identification: A comparative study. 2023 International Conference on System, Computation, Automation and Networking (ICSCAN), PUDUCHERRY, India. 2023, 1-6. 10.1109/ICSCAN58655.2023.10394757
21. Mangkunegara IS, Purwono P, Ma'arif A, Basil N, Marhoon HM, Sharkawy AN: Transformer models in deep learning: Foundations, advances, challenges and future directions. Buletin Ilmiah Sarjana Teknik Elektro. 2025, 7:231-241.
22. Meierkord C: Attitudes towards exogenous and endogenous uses of English: Ugandan's judgements of English structures in varieties of English. International Journal of English Linguistics. 2020, 10:1-14. 10.5539/ijel.v10n1p1
23. Yang EE, Matthew MD, Raghavan S, et al.: Transformer versus traditional natural language processing: how much data is enough for automated radiology report classification?. British Journal of Radiology. 2023, 96:20220769. 10.1259/bjr.20220769

24. Qiang L, Meiqi S: Transformer-based models for natural language processing: A review of methods and applications. 2026 5th International Conference on Electronics Technology and Artificial Intelligence (ETAI), Harbin, China. 2026, 32-36. 10.1109/ETAI68332.2026.11485421

25. Groenwold S, Ou L, Parekh A, Honnavalli S, Levy S, Mirza D, Wang WY: Investigating African-American vernacular English in transformer-based text generation. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020, 5877-5883. 10.18653/v1/2020.emnlp-main.473

26. Nasution AU, Debataraja CS, Ningsih FR, Nasution UA, Daulay SH: The role of culture in enhancing students' language acquisition: Teacher's strategy. Culturalistics: Journal of Cultural, Literary, and Linguistic Studies. 2023, 7:85-93. 10.14710/ca.v7i2.20031

27. Lozano-Ruiz A, Fasfous AF, Ibanez-Casas I, Cruz-Quintana F, Perez-Garcia M, Pérez-Marfil MN: Cultural bias in intelligence assessment using a culture-free test in Moroccan children. Archives of Clinical Neuropsychology. 2021, 36:1502-1510. 10.1093/arclin/acab005

28. Nakayiza J: The sociolinguistic situation of English in Uganda: A case of language attitudes and beliefs. Ugandan English: Its Sociolinguistics, Structure and Uses in a Globalising Post-protectorate. Meierkord C, Isingoma B, Namyalo S (ed): John Benjamins Publishing Company, Amsterdam; 2016. 75-94. 10.1075/veaw.g59.04nak

29. Trisha BJ, Ghosh R, Das S: Integrating local culture and multilingualism in English language teaching in Bangladesh: Bridging global standards and local practices. Studies in Educational Management. 2025, 17:44-61. 10.32038/sem.2025.17.03

30. Ssempuuma J: Morphological and Syntactic Feature Analysis of Ugandan English: Influence from Luganda, Runyankole-Rukiga, and Acholi-Lango. Peter Lang Verlag, Berlin, Germany; 2019. 10.3726/b15305

31. El Mekki A, Atou H, Nacar O, Shehata S, Abdul-Mageed M: NileChat: towards linguistically diverse and culturally aware LLMs for local communities. Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing. 2025, 10967-10991. 10.18653/v1/2025.emnlp-main.556

32. Zhang C, Tao M, Huang Q, Lin J, Chen Z, Feng Y: MC2: towards transparent and culturally-aware NLP for minority languages in China. Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2024, 8832--8850. 10.18653/v1/2024.acl-long.479

33. Li C, Chen M, Wang J, Sitaram S, Xie X: CultureLLM: incorporating cultural differences into large language models. 38th Conference on Neural Information Processing Systems (NeurIPS 2024). 2024, 1-40.

34. Nekoto W, Marivate V, Matsila T, et al.: Participatory research for low-resourced machine translation: a case study in African languages. Findings of the Association for Computational Linguistics: EMNLP 2020. 2020, 2144-2160. 10.18653/v1/2020.findings-emnlp.195

35. Tao Y, Viberg O, Baker RS, Kizilcec RF: Cultural bias and cultural alignment of large language models. PNAS Nexus. 2024, 3:pgae346. 10.1093/pnasnexus/pgae346

36. Adelani DI: Natural language processing for African languages [PREPRINT]. arXiv. 2025, 10.48550/arXiv.2507.00297

37. Duncan S, Leoni G, Steven L, Mahelona K, Jones PL: Fit for our purpose, not yours: Benchmark for a low-resource, Indigenous language. 38th Conference on Neural Information Processing Systems (NeurIPS 2024) Track on Datasets and Benchmarks. 2024, 1-19.

38. Boumediene H, Kaid Berrahal F, Bava Harji M: Natural language processing (NLP) for language, culture, and ethics. International Journal of Innovative Research in Education. 2025, 12:117-134. 10.18844/ijire.v12i2.9852

39. Chang CC, Li CF, Lee CH, Lee HS: Enhancing low-resource minority language translation with LLMs and retrieval-augmented generation for cultural nuances. Intelligent Systems and Applications. Arai K (ed): Springer, Cham; 2025. 1567: 190-204. 10.1007/978-3-032-00071-2_12

40. Katushemererwe F, Caines A, Buttery P: Building natural language processing tools for Runyakitara. Applied Linguistics Review. 2021, 12:585-609. 10.1515/applirev-2020-2004

41. Gallegos IO, Rossi RA, Barrow J, et al.: Bias and fairness in large language models: a survey. Computational Linguistics. 2024, 50:1097-1179. 10.1162/coli_a_00524

42. Visser R, Grobler T, Dunaiski M: Insights into low-resource language modelling: improving model performances for South African languages. JUCS - Journal of Universal Computer Science. 2024, 30:1849-1871. 10.3897/jucs.118889

43. Akera B, Mukiibi J, Naggayi LS, et al.: Machine translation for African languages: Community creation of datasets and models in Uganda. AfricaNLP Workshop. 2022, 1-10.

44. Wairagala EP, Mukiibi J, Babirye C, Katumba A, Nakatumba-Nabende J: The Makerere gendered corpus: a gendered English to Luganda parallel corpus (one) [Data set]. (2022). https://doi.org/10.5281/zenodo.5864560.

45. Mars M: From word embeddings to pre-trained language models: a state-of-the-art walkthrough. Applied Science. 2022, 12:8805. 10.3390/app12178805

46. Raychawdhary N, Bhattacharya S, Seals C, Dozier GV: Empowering sentiment analysis in African low-resource languages through transformer models and strategic language selection. IEEE Access. 2025, 13:147859-147873. 10.1109/access.2025.3599480

47. Mabokela KR, Celik T, Primus M: Sentiment analysis in African languages: Evaluating generative AI and Afrocentric language models. South African Computer Science and Information Systems Research Trends. Gerber A (ed): Springer, Cham; 2026. 2583:259-276. 10.1007/978-3-031-96262-2_17

48. Joshi S, Khan MS, Dafe A, Singh K, Zope V, Jhamtani T: Fine-tuning LLMs for low resource languages. 2024 5th International Conference on Image Processing and Capsule Networks (ICIPCN), Dhulikhel, Nepal. 2024, 511-519. 10.1109/ICIPCN63822.2024.00090

49. Kanjirangat V, Samardžić T, Dolamic L, Rinaldi F: Tokenization and representation biases in multilingual models on dialectal NLP tasks. Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing. 2025, 23992-24010.

50. Nakatumba-Nabende J, Babirye C, Nabende P, et al.: Building text and speech benchmark datasets and models for low-resourced East African languages: experiences and lessons. Applied AI Letters. 2024, 5:e92. 10.1002/ail2.92

51. Suryanto TLM, Wibawa AP, Hariyono H, Nafalski A: Comparative performance of transformer models for cultural heritage in NLP tasks. Advance Sustainable Science Engineering and Technology. 2025, 7:1-13. 10.26877/asset.v7i1.1211

52. Pooja S, Gokul G, Linkesh Mani K, Raj Kumar AS, Amutha Bharathi R: Context-aware AI chatbot using transformer-based models for intelligent user interactions. Proceedings of the 3rd International Conference on Futuristic Technology. 2025, 3:667-674. 10.5220/0013639800004664

53. Liu S, Zhang X, Zhang S, Wang H, Zhang W: Neural machine reading comprehension: methods and trends. Applied Sciences. 2019, 9:3698. 10.3390/app9183698

54. Sen J, Waghela H, Pandey R: Contextual contrastive search for improved text generations in large language models. 2025 6th International Conference for Emerging Technology (INCET), BELGAUM, India. 2025, 1-9. 10.1109/INCET64471.2025.11140308

55. Humayun MA, Yassin H, Shuja J, Alourani A, Abas PE: A transformer fine-tuning strategy for text dialect identification. Neural Computing and Applications. 2023, 35:6115-6124. 10.1007/s00521-022-07944-5

56. Zhang H, Song H, Li S, Zhou M, Song D: A survey of controllable text generation using transformer-based pre-trained language models. ACM Computing Surveys. 2023, 56:1-37. 10.1145/3617680

57. Luo X: Cross-cultural adaptation framework for enhancing large language model outputs in multilingual contexts. Journal of Advanced Computing Systems. 2023, 3:48-62. 10.69987/jacs.2023.30505

58. Liu CC, Gurevych I, Korhonen A: Culturally aware and adapted NLP: A taxonomy and a survey of the state of the art. Transactions of the Association for Computational Linguistics. 2025, 13:652-689.

59. Li Z, Li X, Tao C, et al.: RetriEVAL: Evaluating text generation with contextualized lexical match. WSDM '25: Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining. 2025, 934-943. 10.1145/3701551.3703581

60. International corpus of English - Uganda. (2022). Accessed: January 22, 2026: https://www.ice-corpora.uzh.ch/en/joinice/Teams/iceug.html.

61. Isingoma B, Meierkord C: Capturing the lexicon of Ugandan English: ICE-Uganda, its limitations, and effective complements. Corpus Linguistics and African Englishes. Esimaje AU, Gut U, Antia BE (ed): John Benjamins Publishing Company, Amsterdam; 2019. 293-328. 10.1075/scl.88.13isi

62. Biber D: Representativeness in corpus design. Literary and Linguistic Computing. 1993, 8:243-257. 10.1093/llc/8.4.243

63. McEnery T, Hardie A: Corpus Linguistics: Method, Theory and Practice. Cambridge University Press, Cambridge; 2011. 10.1017/CBO9780511981395

64. Gries ST: What is corpus linguistics?. Language and Linguistics Compass. 2009, 3:1225-1241. 10.1111/j.1749-818x.2009.00149.x

65. Atkins S, Clear J, Ostler N: Corpus design criteria. Literary and Linguistic Computing. 1992, 7:1-16. 10.1093/llc/7.1.1

66. Kytö M: Corpora and historical linguistics. Revista Brasileira de Linguística Aplicada. 2011, 11:417-457. 10.1590/s1984-63982011000200007

67. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I: Language models are unsupervised multitask learners. OpenAI Blog. (2019). https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.

68. Sanh V, Debut L, Chaumond J, Wolf T: Distilbert, a distilled version of bert: smaller, faster, cheaper, and lighter [PREPRINT]. arXiv. 2020, 10.48550/arXiv.1910.01108

69. Black S, Gao L, Wang P, Leahy C, Biderman S: GPT-Neo: Large scale autoregressive language modeling with meshtensorflow. (2021). https://zenodo.org/records/5297715.

70. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. (2022). Accessed: January 21, 2026: https://sagorbrur.github.io/2022/04/bart-paper-review/.

71. Raffel C, Shazeer N, Roberts A, et al.: Exploring the limits of transfer learning with a unified text-to-text transformer. The Journal of Machine Learning Research. 2020, 21:5485-5551.

72. Wei J, Gao Y: LongSum: An efficient transformer for long document summarization. Database Systems for Advanced Applications. DASFAA 2024. Onizuka M, Lee J-G, Tong Y, et al. (ed): Springer, Singapore; 2025. 14851:406-415. 10.1007/978-981-97-5779-4_27

73. Kozhirbayev Z: Enhancing neural machine translation with fine-tuned mBART50 pre-trained model: An examination with low-resource translation pairs. Ingénierie des Systèmes d'Information. 2024, 29:831-838. 10.18280/isi.290304

74. Tiedemann J, Aulamo M, Bakshandaeva D, et al.: Democratizing neural machine translation with OPUS-MT. Language Resources and Evaluation. 2024, 58:713-755. 10.1007/s10579-023-09704-w

How to cite this article:
Byansi D, Isingoma B, Nakatumba-Nabende J, et al. (July 15, 2026) Cultural and Linguistic Contextualization in Ugandan English: Training Transformer Models for Enhanced Local Relevance. Cureus J Comput Sci 3 : es44389-026-00167-y. DOI https://doi.org/10.7759/s44389-026-00167-y