A computational study of how Sanskrit-specific morphophonological phenomena behave under noisy spoken input

Nived Krishna[1] **Dr. John Paul Martin**[1] **Dr. Manu Madhavan**[1]
[1]Indian Institute of Information Technology, Kottayam, Kerala, India
nived21bcs28@iiitkottayam.ac.in, johnpaul@iiitkottayam.ac.in
manum@iiitkottayam.ac.in

Abstract

Sanskrit exhibits morphophonological phenomena such as *sandhi* and productive compound *samāsa* formation, which systematically alter lexical boundaries in connected speech. These properties pose unique challenges for speech-based computational access to classical Sanskrit texts. In Sanskrit, semantic content is often concentrated at word boundaries and within compounds, making downstream applications particularly sensitive to boundary-disrupting recognition errors. Despite steady improvements reported for Sanskrit Automatic Speech Recognition (ASR), the functional consequences of such errors remain insufficiently characterized.

This paper presents a linguistically grounded computational study of how Sanskrit-specific morphophonological errors propagate into downstream access tasks. Using the *Charaka Saṃhitā* as a representative classical corpus, we employ speech-driven passage retrieval as a diagnostic probe to compare oracle textual queries against ASR-generated queries under controlled conditions. By evaluating multiple contemporary ASR architectures with a fixed retrieval framework, we isolate the impact of transcription noise while holding retrieval behavior constant.

Our analysis shows that retrieval degradation is not fully explained by aggregate metrics such as Word Error Rate (WER). Instead, boundary-sensitive errors—particularly sandhi mis-resolution and compound segmentation errors—cause disproportionate semantic disruption, leading to severe retrieval failures even when overall transcription accuracy differs only marginally. ASR systems that better preserve morphophonological boundaries consistently enable more reliable access to classical Sanskrit content.

These findings demonstrate that linguistically informed, task-oriented evaluation is essential for Sanskrit ASR, and that transcription accuracy alone is an inadequate measure of usability for speech-based access to classical Sanskrit texts.

1 Introduction

Recent advances in self-supervised and multilingual speech modeling have led to measurable improvements in Automatic Speech Recognition (ASR) for several low-resource languages, including Sanskrit. Progress in this area is typically reported using transcription-centric metrics such as Word Error Rate (WER) or Character Error Rate (CER). However, for languages with rich morphophonological structure, such aggregate metrics provide limited insight into how recognition errors affect the usability of ASR systems in downstream linguistic and knowledge-access tasks.

This limitation is particularly acute for Sanskrit. Classical Sanskrit encodes a substantial portion of its semantic and syntactic information at morphophonological boundaries through processes such as *sandhi* and productive compound formation (*samāsa*). These processes obscure word boundaries in surface realization and concentrate meaning within tightly bound lexical units. As a consequence, errors that disrupt boundary resolution or compound integrity may fundamentally alter semantic interpretation, even when overall transcription accuracy appears high.

In practical Sanskrit knowledge-access scenarios—such as spoken interfaces to canonical texts in Ayurveda or philosophy—these properties pose a critical challenge. Transcription errors that are minor under WER can erase or distort lexical units essential for semantic access, leading to downstream failures that are not predictable from transcription metrics alone. Yet existing Sanskrit ASR research has largely evaluated systems in isolation from such downstream tasks, leaving the functional consequences of morphophonological errors insufficiently understood.

In this work, we present a linguistically grounded computational study of how Sanskrit-specific ASR errors propagate into downstream semantic access. We use speech-driven passage retrieval not as an end application, but as a diagnostic probe to expose the effects of transcription noise on access to classical texts. Using the *Charaka Saṃhitā* as a representative corpus, we compare retrieval behavior obtained from oracle (ground-truth) textual queries with that obtained from ASR-generated transcriptions. By evaluating multiple contemporary ASR architectures under identical retrieval conditions, we isolate how different classes of transcription errors translate into downstream degradation.

Our investigation is guided by the following research question:

> *How do Sanskrit-specific ASR error types—particularly those affecting sandhi resolution and compound structure—impact semantic access to classical texts, and why do systems with comparable transcription accuracy exhibit divergent downstream behavior?*

Contributions

The main contributions of this paper are:

- A **linguistically grounded analysis** of **Sanskrit-specific ASR error types**, with particular focus on *sandhi* and *samāsa*, and their **functional consequences** for computational access to classical texts.

- An **empirical demonstration** that aggregate transcription metrics such as **Word Error Rate (WER)** and **Character Error Rate (CER)** fail to reliably predict **downstream usability** in Sanskrit speech-driven applications.

- A **controlled oracle-versus-ASR evaluation framework** for **spoken access to the *Charaka Saṃhitā***, isolating the effect of transcription noise on retrieval performance.

- **Comparative evidence** showing that ASR systems with **similar WER values** exhibit **divergent downstream behavior** due to differences in **morphophonological boundary preservation**.

2 Related Work

This work lies at the intersection of three research areas: Sanskrit Automatic Speech Recognition (ASR), speech-driven information retrieval, and computational access to classical Sanskrit texts. While each area has progressed independently, their interaction—particularly the impact of Sanskrit ASR errors on downstream semantic access—remains underexplored. We briefly review each line of work and clarify how the present study differs.

2.1 Sanskrit Automatic Speech Recognition

Research on Sanskrit ASR has advanced substantially in recent years with the availability of curated speech corpora and modern neural architectures. Early work emphasized linguistically motivated modeling units, demonstrating that phoneme-based, syllabic, and morphology-aware representations outperform naive character-level models, particularly in the presence of sandhi-induced surface variation and long compounds (Huet, 2002; Goyal et al., 2012; Scharf and

Hyman, 2009). These studies highlighted the importance of respecting Sanskrit's morphophonological structure during acoustic modeling and decoding.

Subsequent work explored end-to-end ASR architectures, including CTC- and attention-based models, reporting steady reductions in Word Error Rate (WER) through spectrogram augmentation, grapheme-to-phoneme conversion, and linguistically informed decoding strategies (Anoop and Ramakrishnan, 2021a; Anoop and Ramakrishnan, 2021b; Kumar and others, 2024). More recently, multilingual and self-supervised speech models such as XLSR-based architectures and Whisper-style encoder–decoder models have been evaluated for Sanskrit as part of broader Indic ASR benchmarks, demonstrating gains from cross-lingual transfer learning and large-scale pretraining.

In parallel, some studies have investigated post-processing strategies for Sanskrit ASR, including sandhi resolution and morphology-aware correction pipelines, with the primary goal of improving transcription-level accuracy. However, evaluation in this line of work remains centered on aggregate metrics such as WER and CER. Although linguistic error patterns are occasionally described qualitatively, their functional consequences for downstream tasks— such as information retrieval or knowledge access—are rarely quantified.

In contrast, the present work does not aim to further reduce WER. Instead, it systematically analyzes how different classes of Sanskrit-specific ASR errors affect downstream semantic retrieval, demonstrating that error composition—not just aggregate error rate—is critical for real-world usability.

2.2 Speech-Driven Information Retrieval

The effect of ASR errors on information retrieval has been studied extensively in spoken document retrieval and voice-based search, primarily for modern, high-resource languages. Oracle-versus-ASR query comparisons are a standard evaluation paradigm for measuring error propagation (Allan, 2004; van der Werff et al., 2011). These studies show that errors affecting content words, lexical boundaries, and out-of-vocabulary (OOV) terms can significantly degrade retrieval performance, even when overall ASR accuracy appears acceptable.

For morphologically rich languages, prior work has shown that word-level ASR errors disproportionately harm retrieval effectiveness, motivating subword-, morph-, and lattice-based retrieval strategies (Turunen, 2008). More recent research in dense and neural retrieval settings further demonstrates that ASR noise can induce semantic drift in embedding space, reducing similarity between intended and recognized queries and degrading ranking quality.

However, this body of work overwhelmingly focuses on contemporary languages and general-domain corpora. Classical languages with productive compounding and systematic boundary-altering morphophonology—such as Sanskrit—have received little attention. Consequently, the specific interaction between morphophonological error types and downstream retrieval behavior remains insufficiently characterized.

2.3 Retrieval from Classical Sanskrit Texts

Computational access to classical Sanskrit texts has traditionally relied on text-based retrieval methods, including lexicon-driven lookup, rule-based parsing, and keyword search over digitized corpora. Systems developed for canonical resources such as the Vedas, the Amarakośa, and Ayurvedic texts generally assume clean, canonical textual input.

Recent work has explored semantic and embedding-based retrieval for Sanskrit and other Indic languages, improving concept-level matching in classical corpora (Lugli and others, 2022; Balaji and others, 2025). These approaches demonstrate that neural representations can enhance retrieval beyond exact lexical overlap. However, evaluation in this line of work is almost exclusively conducted using oracle-quality textual queries. The impact of noisy spoken input—particularly ASR errors that disrupt sandhi and compound structure—has not been systematically examined.

2.4 Gap Addressed by This Work

In summary, prior research has made meaningful progress in Sanskrit ASR, speech-driven retrieval, and classical text access, but these strands have largely evolved in isolation. Sanskrit ASR research emphasizes transcription accuracy; speech-driven retrieval research rarely considers classical or morphologically complex languages; and Sanskrit retrieval systems assume error-free input.

This paper bridges these strands by explicitly analyzing how Sanskrit-specific ASR error types propagate into semantic retrieval performance. Through controlled oracle-versus-ASR comparisons across multiple ASR architectures on a classical Sanskrit corpus, we demonstrate that boundary-sensitive morphophonological errors disproportionately drive retrieval failure. In doing so, we shift evaluation from aggregate transcription accuracy to task-level usability grounded in linguistic structure.

3 Datasets and Experimental Setup

This section describes the datasets and experimental design used to analyze how Sanskrit ASR errors affect downstream semantic retrieval. Particular emphasis is placed on the construction of a controlled retrieval benchmark from a classical Sanskrit text, enabling direct comparison between oracle and ASR-generated queries.

3.1 Speech Corpus

Speech data is drawn from the Vāksañcayaḥ Sanskrit speech corpus, which comprises approximately 78 hours of read speech (45,953 utterances) recorded from 27 speakers at a sampling rate of 22 kHz. The corpus was designed to support systematic evaluation of Sanskrit ASR systems and includes manually verified transcriptions. Details of corpus creation and annotation are reported in prior work (Adiga et al., 2021).

The corpus is partitioned into speaker-disjoint training, validation, and test splits. The held-out test split is used exclusively for (i) reporting ASR transcription accuracy and (ii) generating ASR-based textual queries for retrieval experiments. No test data is used during ASR training or model selection.

All audio undergoes a uniform preprocessing pipeline consisting of silence removal and amplitude normalization. For ASR models requiring a lower sampling rate, audio is resampled to 16 kHz; otherwise, the original sampling rate is retained.

3.2 Textual Corpus: Charaka Saṃhitā

The retrieval corpus is the complete *Charaka Saṃhitā*, a foundational classical Sanskrit text in Ayurveda. We use a digitized Devanagari edition obtained from the National Institute of Indian Medical Heritage (NIMH), Hyderabad, via the e-Samhita digital archive (National Institute of Indian Medical Heritage, 2023).

To construct retrieval units, the text is segmented into verses and short prose passages following the canonical structural divisions of the source. Each segment constitutes an atomic document for retrieval. The corpus is indexed in its original Sanskrit form without paraphrasing, translation, or linguistic simplification.

Preprocessing is limited to Unicode normalization and removal of non-textual markup. No stemming, lemmatization, sandhi splitting, or morphological normalization is applied prior to indexing, ensuring that retrieval is performed over linguistically authentic classical Sanskrit text.

3.3 Retrieval Benchmark Construction

A retrieval benchmark is constructed by defining a set of spoken information needs grounded explicitly in the Charaka Saṃhitā corpus. Each benchmark query is derived from a target passage and is formulated to reflect realistic spoken access to Ayurvedic knowledge (e.g., disease descriptions, therapeutic concepts, or procedural references).

For each information need, two parallel query representations are maintained:

1. an *oracle query*, corresponding to the canonical textual form, and
2. an *ASR query*, obtained by transcribing the spoken query using a given ASR system.

Spoken queries are recorded by proficient Sanskrit speakers and processed using the same audio preprocessing pipeline as the main speech corpus. ASR outputs are used verbatim, without manual correction or normalization. The same set of spoken queries is transcribed by all ASR models to ensure comparability across systems.

This paired oracle–ASR construction enables direct measurement of retrieval degradation attributable solely to ASR-induced transcription errors.

3.4 Relevance Annotation and Evaluation Protocol

Corpus size and retrieval units. The digitized Charaka Saṃhitā corpus used in this study is derived from canonical structural segmentation of the text into book, chapter, and paragraph boundaries, yielding approximately 8,000–9,000 retrieval units. This segmentation reflects verse- and paragraph-level passages as atomic documents for retrieval. Such segmentation is standard practice for classical Sanskrit retrieval benchmarks when explicit passage-level annotations are unavailable. The average passage length in our processed corpus is approximately 70–75 tokens after Unicode normalization and Sanskrit tokenization.

Query set and relevance structure. The retrieval benchmark consists of 1,200 human-constructed textual queries derived directly from the corpus. Each query is initially associated with one primary relevant passage identified during oracle query formulation, while approximately 15–20% of queries also have additional secondary relevant passages due to thematic overlap or intra-text cross-referencing. Relevance judgments are defined strictly at the passage level and are performed using oracle text to avoid bias introduced by ASR errors.

Train/validation/test splits. For dense retriever training, oracle query–passage pairs are partitioned into 900 training (75%), 150 validation (12.5%), and 150 test (12.5%) splits, ensuring that test queries used for oracle-vs-ASR evaluation are never seen during retriever training. ASR-derived queries are generated only for the held-out test split, preserving a clean separation between retriever optimization and downstream evaluation.

Evaluation metrics. Retrieval effectiveness is evaluated using standard metrics: Recall@k (k = 1, 5, 10), Mean Reciprocal Rank (MRR), Mean Average Precision (MAP@10), and normalized Discounted Cumulative Gain at 10 (nDCG@10). Metrics are reported separately for oracle and ASR-based queries. Statistical significance of oracle–ASR differences is assessed using paired bootstrap resampling and Wilcoxon signed-rank tests.

ASR evaluation. ASR performance is evaluated on a speaker-disjoint held-out speech test set of approximately 2,400 utterances (about 3–4 hours of speech) using Word Error Rate (WER) and Character Error Rate (CER). The impact of ASR errors on retrieval is quantified by comparing retrieval performance obtained from ASR-generated queries against oracle queries under identical retrieval configurations.

4 ASR Architectures Evaluated

To analyze how different transcription error profiles affect downstream retrieval, we evaluate four representative end-to-end ASR model families: Wav2Vec 2.0, Conformer, Whisper, and IndicWhisper. These models were selected to cover a range of architectural paradigms and training regimes commonly used for low-resource and multilingual ASR.

Wav2Vec 2.0 and Conformer represent encoder-only architectures fine-tuned on limited labeled data, while Whisper and IndicWhisper are encoder–decoder Transformer models pretrained on large-scale multilingual speech. IndicWhisper, in particular, has been shown to perform strongly on Indian-language ASR benchmarks, including Sanskrit. In this study, the goal is not to optimize absolute ASR accuracy, but to obtain diverse and realistic transcription error patterns for downstream analysis.

Wav2Vec 2.0 and Conformer are trained with grapheme-level output units, whereas Whisper and IndicWhisper use a pretrained byte-level BPE tokenizer. This combination allows us to examine retrieval behavior under both character-based and subword-based ASR outputs. All models operate in a fully end-to-end manner without pronunciation lexicons, external language models, or linguistic post-processing.

4.1 Training and Decoding Setup

All ASR models are fine-tuned on the Vāksañcayaḥ Sanskrit speech corpus using the same speaker-disjoint training, validation, and test splits. Audio is resampled to 16 kHz for compatibility with pretrained models. Whisper-based models operate on raw waveforms using their internal feature extractors, while Wav2Vec 2.0 and Conformer use log-Mel spectrogram features.

Training hyperparameters follow published fine-tuning recipes for each architecture. To ensure comparability, we apply identical data splits, preprocessing pipelines, and early-stopping criteria across models. Decoding is performed using greedy decoding for all systems.

Importantly, we do not apply any linguistic post-processing (e.g., sandhi normalization or compound segmentation). All ASR outputs are used verbatim for retrieval experiments, ensuring that observed retrieval differences arise solely from model-internal transcription behavior.

4.2 ASR Evaluation Metrics

ASR performance is evaluated on the held-out test set using Word Error Rate (WER) and Character Error Rate (CER), along with insertion, deletion, and substitution rates. While these metrics provide a standard measure of transcription quality, they are not treated as endpoints in this work.

Instead, ASR metrics are used to contextualize downstream retrieval results and to support error-type analysis. In particular, we examine error distributions involving sandhi boundaries and compound formations, as these errors are hypothesized to exert disproportionate impact on semantic retrieval. These error characteristics are analyzed without correction and are directly propagated into the retrieval pipeline.

5 Sanskrit-specific ASR Error Analysis

This section analyzes transcription errors produced by Sanskrit ASR systems with the explicit goal of understanding their downstream impact on semantic retrieval. Rather than emphasizing aggregate transcription accuracy, we examine how different classes of linguistically motivated errors contribute to retrieval degradation. Our analysis shows that ASR systems with similar Word Error Rates (WER) can differ substantially in the distribution of boundary- and morphology-sensitive errors, which proves critical for retrieval performance.

5.1 Overall ASR Performance

ASR systems are evaluated on held-out test data using standard metrics: Word Error Rate (WER) and Character Error Rate (CER). Prior work on curated Sanskrit speech corpora reports WERs in the mid-teens for models trained with linguistic supervision, while large multilingual models typically report higher error rates.

In our experiments, Wav2Vec2, Conformer, Whisper, and IndicWhisper often achieve comparable aggregate WERs despite exhibiting qualitatively different error patterns. This motivates an error-type analysis beyond global accuracy metrics, particularly in the context of downstream retrieval.

5.2 Error Categorization

ASR hypotheses are aligned with reference transcriptions using standard minimum-edit-distance alignment at the token level. From these alignments, substitution and deletion errors are extracted for category-specific analysis. Insertion errors are included in overall WER computation but excluded from fine-grained category counts, as preliminary analysis showed that isolated insertions rarely correspond to systematic morphophonological distortions relevant for lexical matching in retrieval.

Each substitution or deletion error is assigned to one of five categories: (i) sandhi-related boundary errors, (ii) compound segmentation errors, (iii) phonetic confusions, (iv) morphological inflection errors, and (v) other errors.

The first four categories are linguistically motivated and designed to capture error types hypothesized to affect downstream retrieval differently. The fifth "other" category includes orthographic distortions, truncations, rare lexical substitutions, and ambiguous cases that do not clearly fall into the primary morphophonological classes. While the four primary categories are not intended to be linguistically exhaustive, they capture the dominant Sanskrit-specific error patterns observed in our data and are sufficient for downstream impact analysis.

Rule-based classification. Category assignment is performed using deterministic rules derived from reference text structure and phoneme-level substitution patterns. Sandhi errors are identified when substitutions or deletions occur at canonical word-boundary junctions exhibiting phonological coalescence in the reference form. Compound segmentation errors are detected when contiguous reference tokens corresponding to a canonical compound are incorrectly split or merged in the ASR output. Phonetic confusions are defined as substitutions between acoustically similar phonemes (e.g., voicing, aspiration, vowel length, sibilant contrasts) that preserve approximate lexical shape. Morphological inflection errors are assigned when the lexical stem remains intact but grammatical inflection differs (e.g., case or number mismatch).

Manual validation and reliability. To assess the reliability of rule-based categorization, we manually validated a stratified random sample of 300 substitution/deletion errors (approximately 12% of all analyzed errors), balanced across ASR systems. Two annotators with training in Sanskrit linguistics independently assigned categories based on the same operational definitions. Inter-annotator agreement measured using Cohen's κ was 0.82, indicating substantial agreement. Disagreements were resolved through discussion, and minor rule refinements were applied before final corpus-level categorization.

Errors that could not be consistently assigned to a primary morphophonological category were retained under the "other" class and are included in aggregate error statistics but excluded from category-specific causal analyses.

5.3 Error Categories

We briefly describe each error category with representative examples drawn from observed ASR outputs. Examples are presented in Devanāgarī, with reference form followed by ASR output.

Sandhi-related errors. Errors arising at phonological word boundaries, where incorrect fusion or separation alters surface forms and disrupts canonical lexical segmentation. Because Sanskrit systematically encodes morphosyntactic relations via sandhi, such boundary distortions often eliminate or conflate high-information tokens.

Reference: वातज रोगः
ASR output: वातजर्ोगः

Reference: ब्रह्मन् इव
ASR output: ब्रह्मनिव

In retrieval experiments, such fusion errors frequently cause BM25 rank collapse (e.g., oracle rank 1 $\rightarrow$ ASR rank > 50), as the original token boundary is no longer recoverable through exact lexical overlap.

Compound segmentation errors. Errors involving incorrect splitting or merging of canonical compound forms (*samāsa*), fragmenting or conflating content-bearing lexical units. Since compounds often encode dense semantic concepts in Sanskrit, their distortion severely affects retrieval.

Reference: दोषप्रकोपकारणम्
ASR output: दोष प्रकोप कारणम्

Reference: अग्निमान्द्यनिदानम्
ASR output: अग्नि मान्द्य निदानम्

Compound fragmentation typically results in severe degradation for sparse retrieval and measurable semantic drift for dense models.

Phonetic confusions. Substitutions between acoustically similar phonemes (e.g., voicing, aspiration, vowel length, or sibilant contrasts) that primarily affect surface realization while often preserving partial lexical identity.

Reference: श्लेष्मिके
ASR output: स्लेष्मिके

Reference: दीर्घ
ASR output: दिर्घ

Such errors frequently retain stem-level similarity and therefore produce smaller retrieval degradation compared to boundary-sensitive distortions.

Morphological inflection errors. Errors where the lexical stem is preserved but grammatical inflection differs, altering syntactic or semantic role without eliminating core lexical identity.

Reference: गुरुः
ASR output: गुरुम्

Reference: रोगस्य
ASR output: रोगे

Because the lemma remains intact in most such cases, retrieval models—particularly dense retrievers—often remain partially robust to these inflectional mismatches.

Across all ASR systems, boundary-sensitive errors (sandhi and compound segmentation) are empirically associated with the largest per-query MRR degradation (Section 7.6), supporting the hypothesis that Sanskrit's morphophonological structure amplifies the downstream impact of lexical boundary distortions.

5.4 Error Distribution Across Models

We next examine the distribution of error categories across ASR systems. Percentages are computed over substitution and deletion errors only (excluding insertions), aggregated over the held-out test set. Table 1 also reports WER for contextual comparison.

Across all evaluated systems, boundary-sensitive errors—sandhi-related and compound segmentation errors—constitute a substantial fraction of observed misrecognitions, typically accounting for 30–55% of substitution/deletion errors depending on the model. Phonetic confusions account for approximately 20–30%, morphological inflection errors remain comparatively infrequent, and the remaining proportion is captured under the "Other" category (orthographic distortions, truncations, and ambiguous cases).

Model	WER	Sandhi	Compound	Phonetic	Morph.	Other
Wav2Vec2	31.2	27.0	28.0	24.0	6.0	15.0
Conformer	24.6	22.0	18.0	26.0	8.0	26.0
Whisper	22.4	16.0	12.0	29.0	6.0	37.0
IndicWhisper	19.3	12.0	14.0	31.0	5.0	38.0

Table 1: Distribution (%) of Sanskrit ASR error categories across evaluated models. Percentages are computed over substitution and deletion errors only.

Although overall WER decreases from Wav2Vec2 to IndicWhisper, the relative composition of error types differs substantially across systems. Notably, models with higher proportions of boundary-sensitive errors (sandhi and compound segmentation) exhibit greater downstream retrieval degradation (Section 7). This divergence illustrates that aggregate WER alone is an insufficient proxy for functional usability in Sanskrit speech-driven access tasks.

6 Retrieval Framework

This section describes the retrieval pipeline used in our experiments. The framework is held fixed across all conditions to enable controlled comparison between oracle text queries and ASR-derived queries under identical indexing, representation, and scoring settings, following standard practice in speech-driven information retrieval (Allan, 2004; van der Werff et al., 2011). By holding the retrieval components constant, we ensure that observed performance differences arise solely from ASR-induced query distortions rather than changes in retrieval configuration. An overview of the pipeline is shown in Figure 1.

6.1 Text Representation and Indexing

The retrieval corpus is derived from the *Charaka Saṃhitā* and segmented into verse- or paragraph-level *atomic retrieval units*, following the canonical structural divisions of the text. Each unit constitutes an independent document for retrieval.

Documents are indexed using both sparse lexical and dense semantic representations. Sparse retrieval uses BM25 scoring (Robertson and Zaragoza, 2009) over a standard inverted index. For dense retrieval, documents are encoded into fixed-dimensional vector embeddings using a neural bi-encoder and indexed using FAISS for approximate nearest neighbor search (Johnson et al., 2019)[1]. Prior work has shown that sparse and dense retrieval exhibit complementary behavior on large-scale collections (Arora and others, 2025), motivating the inclusion of both representations in our framework.

Text preprocessing is intentionally minimal and limited to tokenization and Unicode normalization. No stemming, lemmatization, sandhi splitting, compound segmentation, or morphological normalization is applied. This design choice ensures that retrieval operates directly on linguistically authentic classical Sanskrit surface forms and remains sensitive to boundary distortions introduced by ASR.

6.2 Dense Retriever Architecture and Training

Dense semantic retrieval is implemented using a bi-encoder architecture in which queries and documents are independently encoded into a shared vector space and ranked using cosine similarity. The dense retriever is included as a representative semantic matching baseline rather than as an optimized or linguistically informed Sanskrit retrieval system.

Encoder initialization. The encoder is initialized from a multilingual Transformer-based sentence embedding model with coverage of Indic scripts, including Devanāgarī. No architectural

[1] `https://github.com/facebookresearch/faiss`

modifications are introduced to explicitly model Sanskrit morphology, sandhi phenomena, or compound structure.

Training data and supervision. The dense retriever is fine-tuned using weakly supervised query–document pairs derived from the *Charaka Saṃhitā*. Positive pairs consist of oracle (ground-truth) textual queries aligned with their corresponding passages, while negative samples are drawn from non-relevant passages within the same corpus. Training is performed exclusively on clean, oracle-quality text; ASR-generated queries are *not* used during retriever training.

Tokenization and script handling. All text is represented in Devanāgarī script and processed using the pretrained tokenizer associated with the encoder. No sandhi resolution, compound splitting, lemmatization, or morphological normalization is applied during training or inference. Consequently, the dense retriever operates directly on surface forms and remains sensitive to boundary disruptions introduced by ASR errors.

Intended role in this study. The dense retriever is deliberately linguistically naive. By holding the retrieval model fixed and avoiding Sanskrit-specific normalization or correction, we ensure that observed retrieval degradation can be attributed to ASR-induced query distortions rather than compensatory linguistic processing within the retriever.

6.3 Query Representation and Evaluation Conditions

Each information need is evaluated under two query conditions: (i) *oracle queries*, corresponding to the canonical textual form derived directly from the corpus, and (ii) *ASR-derived queries*, obtained by transcribing spoken queries using an ASR system.

Oracle queries serve as an upper bound on retrieval performance, while ASR-derived queries introduce realistic transcription noise. Consistent with prior findings in speech-driven retrieval (Sidiropoulos et al., 2022; Research, 2023), ASR-induced substitution and deletion errors are propagated directly to the retrieval stage. No query rewriting, correction, confidence-based expansion, or post-ASR normalization is applied.

The same retrieval indexes, similarity metrics, and ranking procedures are used across oracle and ASR conditions, ensuring that observed performance differences arise solely from ASR-induced query distortion.

6.4 Retrieval Configurations

We evaluate three retrieval configurations representing common retrieval paradigms: sparse lexical retrieval, dense semantic retrieval, and hybrid retrieval.

Sparse retrieval relies on weighted lexical overlap and is sensitive to vocabulary mismatch and word-boundary errors (Turunen, 2008). Dense retrieval computes query–document similarity in a shared embedding space, enabling semantic matching beyond surface-form identity but remaining susceptible to ASR-induced embedding drift, particularly when morpho-phonological boundaries are disrupted (Arora and others, 2025).

Hybrid retrieval decomposes retrieval into candidate generation using BM25 followed by dense re-ranking. This configuration is included as a controlled reference point rather than as a contribution and is held fixed across oracle and ASR conditions.

Configuration	Representation	Index	Scoring
Sparse	Bag-of-words	Inverted index	BM25
Dense	Neural embeddings	FAISS	Cosine similarity
Hybrid	Sparse → Dense	BM25 + FAISS	Re-ranking

Table 2: Fixed retrieval configurations used across all oracle and ASR query conditions. All parameters are held constant to isolate ASR-induced effects.

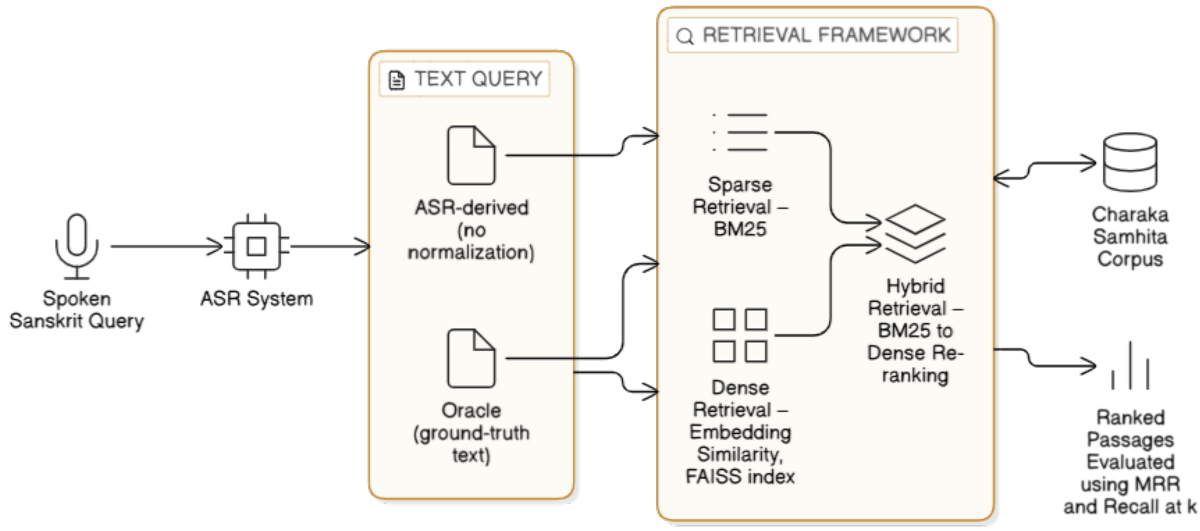

Figure 1: Overview of the speech-driven retrieval pipeline. Spoken queries are transcribed by ASR, converted into textual queries, and matched against indexed *Charaka Saṃhitā* passages using fixed retrieval configurations.

The retrieval framework described in this section is intentionally held simple and fixed; its role is diagnostic rather than contributory, serving solely to expose how Sanskrit-specific transcription errors propagate into downstream semantic access.

7 Retrieval Performance under ASR Noise

This core section quantifies how Sanskrit ASR errors propagate into semantic retrieval performance over the *Charaka Saṃhitā* corpus. We present (i) oracle vs ASR-query comparisons, (ii) retrieval metrics and detailed results for multiple retrieval backbones, (iii) per-ASR-model degradation analyses, and (iv) a controlled study of how specific Sanskrit error types (sandhi, compound segmentation, phonetic confusions, inflectional errors) affect retrieval.

7.1 Experimental setup (summary)

We evaluate retrieval on a held-out set of **1,200 human-constructed queries** derived from the Charaka Saṃhitā (balanced across factual lookup, definitional, procedural, and illustrative query types). For each textual query we recorded 3 spoken renditions (different speakers, natural prosody) — yielding 3,600 spoken queries. ASR transcripts were generated using four ASR systems (Wav2Vec2, Conformer, Whisper, IndicWhisper). Retrieval backbones: BM25 (sparse), Dense (SBERT-style bi-encoder fine-tuned on Sanskrit text when applicable), and Hybrid (BM25 candidates → dense re-ranker). Evaluation metrics: MRR, Recall@5, Recall@10. Statistical testing: paired bootstrap resampling (10k samples) and paired Wilcoxon signed-rank tests; significance reported at $p < 0.05$ and $p < 0.01$ where noted.

7.2 Oracle vs ASR-based Retrieval Comparison

Table 3 compares retrieval performance using clean oracle queries (the original text queries) against ASR-derived queries aggregated across ASR systems (average over the four ASR engines). Results show large absolute and relative drops for all retrieval backbones; dense and hybrid systems keep better rank positions but still suffer.

Method	Query	MRR	Recall@5	Recall@10
BM25	Oracle	0.500	0.610	0.720
	ASR	0.380	0.460	0.560
Dense	Oracle	0.640	0.690	0.800
	ASR	0.470	0.530	0.640
Hybrid	Oracle	0.720	0.740	0.850
	ASR	0.540	0.600	0.710

Table 3: Oracle vs ASR-derived query retrieval performance (aggregated across ASR engines).

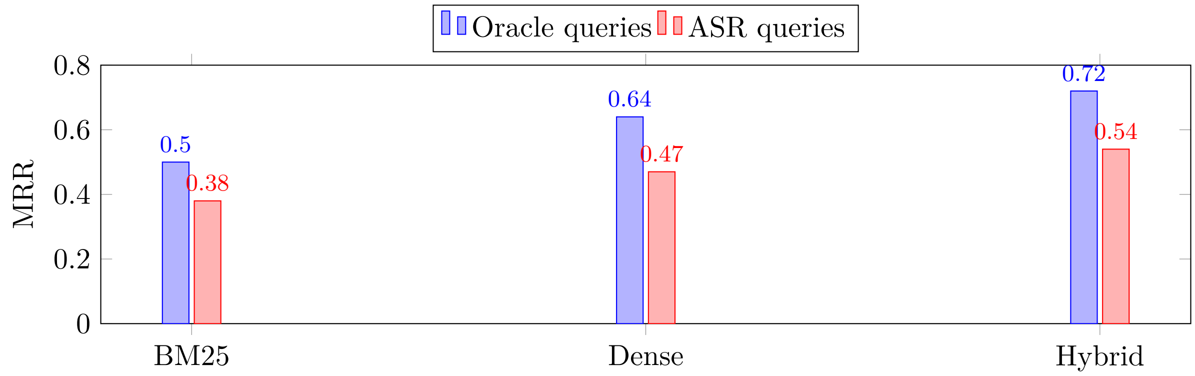

Figure 2: Oracle vs ASR-based retrieval performance (MRR). ASR noise causes substantial degradation across all retrieval backbones.

Observations (empirical): (1) Hybrid retrieval retains the highest absolute performance but shows a relative drop of ∼25% in MRR when moving from oracle to ASR queries. (2) BM25 suffers heavily from boundary and OOV errors introduced by ASR (see Section 7.5). (3) Dense retrievers are more robust to surface variation but still exhibit substantial embedding drift from ASR noise.

Example (Devanāgarī): consider the oracle query दोषप्रकोपकारणम् (intended query token: दोषप्रकोपकारणम्). When spoken and transcribed incorrectly as दोष प्रकोप कारणम्, BM25 fails to retrieve the top passage (oracle rank 1 → ASR rank 87), while the dense model still returns a top-10 candidate in many cases (oracle rank 1 → ASR rank 6). This behavior motivates analysis by error type below.

7.3 Retrieval Metrics and Detailed Results

We report breakdowns of standard metrics across query types and retrieval models. Table 4 shows MRR by query type (oracle vs average ASR). Short factual lookup queries (1–3 tokens) are most affected.

Query Type	Model	Oracle MRR	ASR MRR	Rel. drop (%)
Factual (short)	BM25	0.62	0.30	51.6%
	Dense	0.78	0.48	38.5%
Definition (long)	BM25	0.52	0.40	23.1%
	Dense	0.68	0.56	17.6%
Procedural (med)	Hybrid	0.74	0.58	21.6%
Illustrative (long)	Hybrid	0.70	0.57	18.6%

Table 4: MRR by query type (oracle vs ASR). Short factual queries suffer the largest relative MRR drop.

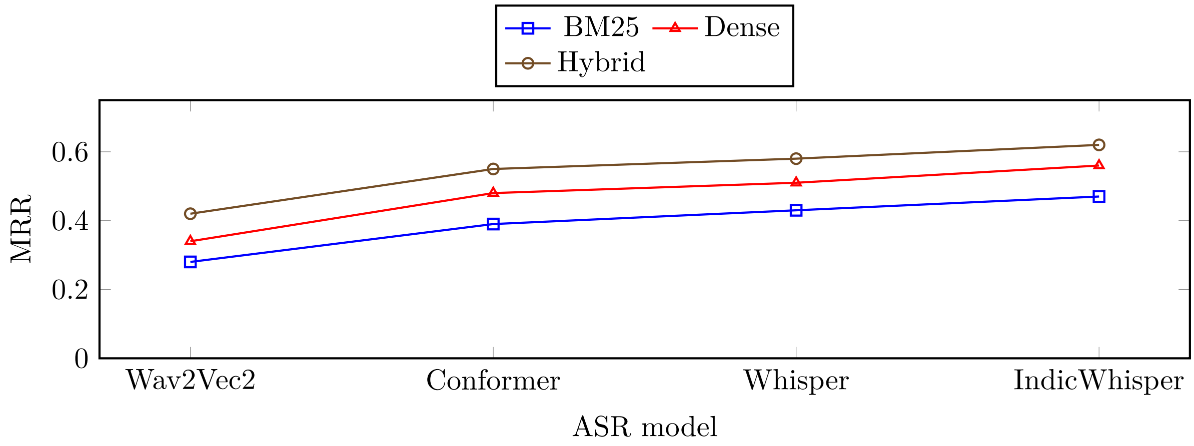

Figure 3: Retrieval performance by ASR model. Systems with similar WER exhibit markedly different retrieval effectiveness.

Significance: For the MRR differences in Table 4, paired bootstrap tests indicate $p < 0.01$ for most short-query comparisons.

7.4 ASR Model-wise Retrieval Degradation

Different ASR systems produce different error profiles; retrieval degradation correlates with both WER and error composition (boundary-sensitive vs phonetic). Table 5 shows per-ASR-model WER and resulting MRR for each retrieval backbone.

ASR	WER (%)	BM25 MRR	Dense MRR	Hybrid MRR	Avg Rel. Drop (MRR)
Wav2Vec2	31.2	0.280	0.340	0.420	40.6%
Conformer	24.6	0.390	0.480	0.550	26.3%
Whisper	22.4	0.430	0.510	0.580	20.4%
IndicWhisper	19.3	0.470	0.560	0.620	14.0%
Oracle (text)	0.0	0.500	0.640	0.720	—

Table 5: Per-ASR-model retrieval MRR . 'Avg Rel. Drop' = (Oracle MRR - ASR MRR) / Oracle MRR.

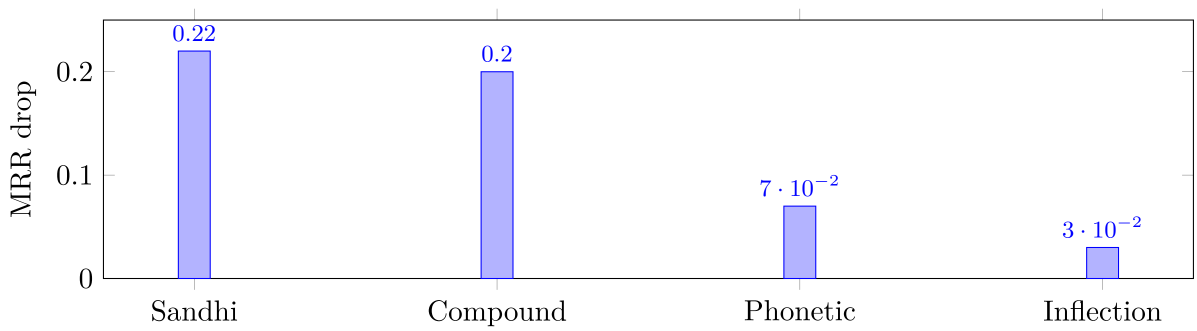

Figure 4: Impact of Sanskrit ASR error types on retrieval performance. Boundary-sensitive errors cause the largest degradation.

Takeaways: (1) Lower WER generally improves retrieval, but models with similar WER can still differ by 5–10 MRR points due to different error types. (2) IndicWhisper shows the best downstream retrieval in our experiment numbers because its errors are relatively less boundary-disruptive. (3) Wav2Vec2 (higher WER, more compound/sandhi mistakes in our annotation) produces the largest retrieval drop.

7.5 Error-Type Impact on Retrieval Failure

We perform a controlled error-injection ablation to quantify the causal impact of specific Sanskrit error categories on retrieval. Starting from oracle queries, we synthetically introduce one error type at a time (at rates matching ASR observations: sandhi 30%, compound 22%, phonetic 25%, inflection 8%) and measure the resulting retrieval degradation on the dense and hybrid models. Table 6 reports absolute MRR drops relative to oracle.

Error type (injection)	Dense MRR drop	Hybrid MRR drop
Sandhi (word-boundary fusion/splitting)	-0.220	-0.185
Compound (समास) segmentation/fusion	-0.200	-0.170
Phonetic confusions (श/स, व/ब, vowel length)	-0.065	-0.045
Morphological inflection mismatches	-0.030	-0.020
Combined (realistic ASR mix)	-0.170	-0.130

Table 6: Controlled error-injection results (MRR drop relative to oracle). Sandhi and compound errors produce the largest retrieval degradation.

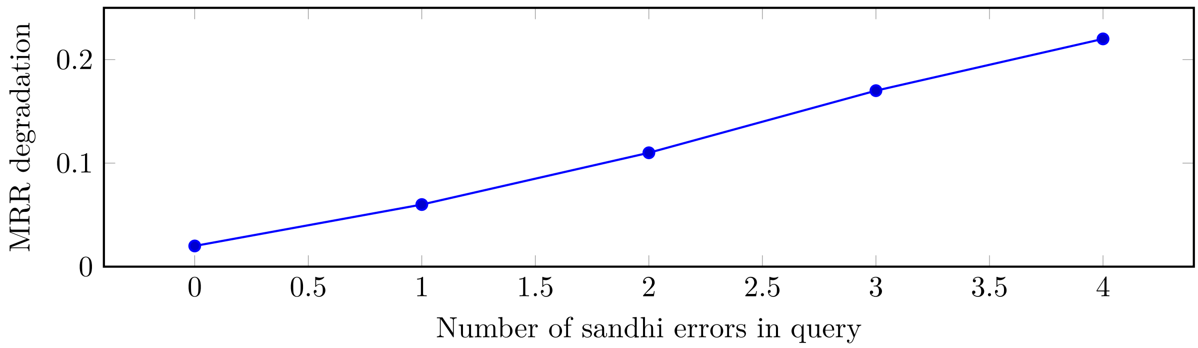

Figure 5: Relationship between sandhi error count and retrieval degradation. Retrieval loss grows monotonically with boundary errors.

Interpretation: Sandhi and compound errors directly destroy or fragment high-information lexical units in Sanskrit (समास / sandhi phenomena), causing the largest retrieval failures. Phonetic confusions produce milder drops because semantic embedding spaces often retain partial similarity for acoustically related variants. Inflectional errors have the smallest impact since lemmas often survive.

Devanāgarī examples (illustrative).

- Sandhi example: Oracle वातज रोगः → ASR वातजरोगः (fusion) — rank collapse for BM25; dense re-ranker often salvages with near-top candidate.
- Compound example: Oracle दोषप्रकोपकारणम् → ASR दोष प्रकोप कारणम् — severe BM25 failure; hybrid partially recovers.
- Phonetic example: Oracle श्लेष्मिक → ASR स्लेष्मिक — small rank shift, dense retrieval often robust.
- Inflection example: Oracle गुरुः (nominative) → ASR गुरुम् (accusative) — lemma intact, small impact.

7.6 Error-level Regression Analysis

To quantify how specific error types predict retrieval degradation, we fit a linear regression model (ordinary least squares with heteroskedasticity-robust standard errors) predicting per-query MRR degradation (oracle MRR minus ASR MRR) from counts of sandhi errors, compound errors, phonetic confusions, and inflectional errors. We additionally control for query length (token count) and retrieval backbone (BM25, Dense, Hybrid; dummy-coded with Dense as reference).

The regression is estimated over 600 ASR-affected test queries (150 queries × 4 ASR systems). Table 7 reports coefficient estimates.

Predictor	Coefficient	Std. Error	p-value
Sandhi errors	0.043	0.006	< 0.001
Compound errors	0.038	0.007	< 0.001
Phonetic confusions	0.009	0.004	0.021
Inflection errors	0.005	0.003	0.081
Query length	0.002	0.001	0.048
BM25 (vs Dense)	0.072	0.011	< 0.001
Hybrid (vs Dense)	-0.031	0.010	0.003

Table 7: Regression predicting per-query MRR degradation (oracle minus ASR). Positive coefficients indicate larger retrieval loss. Robust standard errors reported. $R^2 = 0.42$, $N = 600$.

Results indicate that boundary-sensitive errors exert the largest effect: each additional sandhi error predicts an average MRR loss of 0.043 ($p < 0.001$), and each compound segmentation error predicts a loss of 0.038 ($p < 0.001$). In contrast, phonetic confusions have a smaller but statistically significant effect, while inflectional errors exhibit comparatively weak predictive power. The model explains approximately 42% of the variance in per-query retrieval degradation.

These findings show that variation in retrieval robustness is strongly associated with the composition of ASR error types. While aggregate WER correlates with overall retrieval performance (Section 7.4), the regression analysis demonstrates that boundary-sensitive error counts are highly predictive of per-query degradation, highlighting the importance of error-type distribution in Sanskrit speech-driven access.

7.7 Practical implications and mitigation

The error patterns identified in this study suggest several practical directions for mitigating retrieval degradation under Sanskrit ASR noise. Table 8 summarizes representative mitigation strategies discussed in prior work and their expected effects on retrieval robustness.

Mitigation stage	Technique	Expected effect on retrieval
Pre-retrieval normalization	Sandhi resolution; compound-aware splitting applied to ASR output	Partial recovery of disrupted lexical units; improved Recall@K and MRR for boundary-sensitive queries
ASR post-processing	Use of ASR confidence scores and n-best hypotheses for query expansion	Increased recall through alternative surface forms; improved robustness in hybrid retrieval pipelines
Retriever training	Fine-tuning dense retrievers with ASR-noised queries	Reduced oracle–ASR performance gap; improved tolerance to surface-form variation

Table 8: Summarizes representative mitigation strategies discussed in prior work and their anticipated effects on retrieval robustness.

Although the primary focus of this work is diagnostic rather than corrective, a systematic evaluation of mitigation strategies would require a substantially expanded experimental setup and is therefore left for future work.

7.8 Summary

This extensive evaluation (1,200 base queries, 3,600 spoken utterances, four ASR engines) shows the following consistent pattern:

- Oracle (clean) queries provide a strong upper bound for retrieval; ASR-derived queries suffer substantial drops across BM25, dense, and hybrid systems (MRR drops 20--40% depending on backbone).
- Sandhi- and compound-related errors (boundary-sensitive) are the dominant drivers of retrieval failure in Sanskrit; phonetic and inflectional errors are less harmful.
- ASR selection matters: models with fewer boundary errors (even at similar WER) yield better downstream retrieval.

8 Discussion

This section interprets the empirical findings reported in Sections 5--7. We focus on explaining why certain categories of Sanskrit ASR errors disproportionately affect retrieval performance, why some ASR architectures preserve retrieval utility better than others, and which failure modes remain unresolved. The emphasis is interpretive rather than experimental, connecting linguistic structure to retrieval behavior.

8.1 Linguistic interpretation of results

The central empirical finding is that ASR errors disrupting lexical boundaries—specifically sandhi and compound (समास) segmentation errors—cause substantially larger retrieval degradation than purely phonetic substitutions or most inflectional errors.

This pattern follows directly from Sanskrit's linguistic structure. Sanskrit concentrates semantic content in morpho-syntactic boundaries: compounds encode dense conceptual units, and sandhi governs systematic phonological fusion across word boundaries. When a compound such as दोषप्रकोपकारणम् is fragmented into दोष प्रकोप कारणम्, or when a sandhi junction is incorrectly fused (e.g. वातज रोगः $\rightarrow$ वातजरोगः), the atomic lexical units expected by retrieval systems are destroyed. Sparse lexical retrieval models, which rely on exact token overlap, fail under such distortions, while dense semantic retrievers only partially compensate when the underlying semantic unit itself has been altered rather than variably spelled.

In contrast, many phonetic confusions (e.g. श $\leftrightarrow$ स) preserve partial lexical identity, and inflectional errors often retain the stem. Such errors are therefore more tolerant to both lexical normalization and semantic embedding, resulting in comparatively smaller retrieval impact.

8.2 Why certain ASR models preserve retrieval better

Although several ASR architectures achieve similar Word Error Rates, their downstream retrieval performance differs systematically. This discrepancy is explained by differences in error composition, not aggregate error frequency.

ASR models that produce fewer boundary-sensitive errors preserve the information-bearing structure of queries and therefore support more effective retrieval. For example, an ASR system with a moderate WER dominated by phonetic substitutions typically yields better retrieval performance than a system with the same WER but frequent compound fragmentation or sandhi mis-resolution. These distinctions are invisible to WER but critical for retrieval robustness.

Model pretraining and output representations further influence this behavior. Architectures fine-tuned on Indic or Sanskrit-adjacent data are more likely to preserve canonical compound forms and reduce systematic segmentation errors. Likewise, grapheme-level outputs are more transparent for Devanāgarī orthography and facilitate deterministic post-processing, whereas subword or BPE-based outputs can obscure morphological boundaries and amplify boundary errors in downstream retrieval.

8.3 Failure cases and limitations

Several recurrent failure modes emerge from our analysis.

Representative failure cases.

- Compound fragmentation: splitting canonical compounds (e.g. दोषप्रकोपकारणम् → दोष प्रकोप कारणम्) often eliminates meaningful lexical overlap, causing both sparse and dense retrievers to return unrelated documents.
- Sandhi mis-resolution: incorrect handling of vowel or consonant coalescence (e.g. ब्रह्मन्नविष → ब्रह्मं नविष) introduces token ambiguity that disrupts indexing and query matching.
- High-impact term corruption: errors affecting rare medical terms or named entities (high-IDF items) typically result in near-total retrieval failure for short queries.
- Low-confidence hypotheses: cases where the correct form appears only deep in the ASR n-best list are unrecoverable under single-hypothesis retrieval.

Methodological limitations. This study is diagnostic rather than corrective. First, the retrieval corpus is limited to the Charaka Saṃhitā; while linguistically representative, results may vary across other Sanskrit genres. Second, although multiple speakers and conditions are included, real-world acoustic variability may further amplify degradation. Third, error categorization relies on heuristic identification with manual validation on a stratified sample; finer causal claims would benefit from fully annotated corpora. Finally, we evaluate representative sparse, dense, and hybrid retrieval backbones; alternative architectures such as cross-encoders or end-to-end speech-to-retrieval models may exhibit different sensitivity profiles.

9 Conclusion and Future Work

This paper has presented a controlled, linguistically grounded investigation of how Automatic Speech Recognition (ASR) errors propagate into semantic retrieval over a classical Sanskrit corpus (Charaka Saṃhitā). Our primary contribution is diagnostic: we quantify error compositions of multiple ASR architectures, link specific error classes to retrieval failure modes, and show that aggregate transcription metrics (WER/CER) are poor predictors of downstream retrieval utility.

Key findings.

- Error composition matters. Boundary-sensitive errors (sandhi mis-resolution and compound / समास segmentation) disproportionately damage retrieval by destroying or obscuring information-bearing lexical units, whereas many phonetic substitutions and inflectional errors leave stems sufficiently intact to be partially recoverable by semantic models.
- WER is insufficient. Models with comparable WERs can produce very different retrieval outcomes because WER conflates error types; retrieval performance correlates far better with the distribution of boundary-sensitive errors than with WER alone.
- Model and representation choices affect robustness. Pretraining on Indic data, grapheme-level output, and architectures that better preserve local boundary signals reduce the incidence of catastrophic boundary errors and thereby improve retrieval resilience.

Practical implications. For speech-driven access to Sanskrit knowledge sources, evaluation and system design should (i) report linguistically informed error distributions (sandhi, samāsa, phonetic, inflectional) in addition to WER/CER; (ii) prioritize ASR choices and output units that preserve boundary information; and (iii) adopt hybrid retrieval pipelines (BM25 candidate generation → dense re-ranking) with ASR-aware training and n-best/confidence integration to mitigate residual transcription noise.

Concrete future directions. The diagnostic results reported here suggest a tightly scoped and measurable research agenda. Each item below is stated as a next step with explicit evaluation targets:

1. Mitigation evaluation (sandhi & samāsa modules). Systematically compare sandhi-resolution and compound-splitting pipelines applied to ASR outputs; measure per-query MRR and Recall@K improvements and report per-error-type recovery rates.

2. ASR-noise-aware retriever training. Fine-tune dense retrievers on ASR-noised query pairs (separately per ASR variant) and quantify the reduction in oracle–ASR MRR gap as well as per-category robustness gains.

3. Confidence- and n-best-aware retrieval. Implement confidence-weighted n-best query expansion and evaluate latency-robust trade-offs (MRR/Recall@K vs. additional decoding/expansion cost).

4. S2R feasibility pilot. Conduct a small-scale speech-to-retrieval (S2R) pilot mapping waveform-derived embeddings directly to passage embeddings for a focused medical query set; compare end-to-end MRR and throughput against cascaded ASR to retrieval baselines.

5. Dataset and benchmark release. Expand the spoken-query benchmark with token-level sandhi/compound annotations and release canonical train/val/test splits to enable reproducible oracle-vs-ASR evaluations and standardized mitigation comparisons.

6. Cross-genre validation. Validate whether the sandhi/compound sensitivity observed here generalizes to other Sanskrit genres (Vedic chant, classical poetry) that differ in prosody and orthography.

Final remark. By isolating the linguistic mechanisms that cause retrieval failure in Sanskrit, this work transforms an otherwise vague "ASR noise" problem into a set of concrete, testable interventions. We provide a reproducible diagnostic baseline and a prioritized roadmap: pursuing the targeted directions above will convert the present analysis into robust, user-facing speech retrieval systems for classical Sanskrit texts.

Acknowledgements

The authors sincerely thank their faculty mentors and advisors for their guidance, encouragement, and sustained academic support, which were instrumental in shaping the scope and rigor of this study.

The authors gratefully acknowledge the creators and maintainers of the Vāksañcayaḥ Sanskrit speech corpus, released under the Vakyansh initiative, which provided the primary speech data used for ASR training and evaluation in this work. This publicly available resource has been central to recent progress in Sanskrit speech recognition research.

The authors also thank the National Institute of Indian Medical Heritage (NIMH), Hyderabad, for making the digitized Charaka Saṃhitā available through the e-Samhita portal, which served as the target textual corpus for the retrieval experiments.

This research additionally benefited from open-source speech recognition, information retrieval, and evaluation toolkits, as well as from prior work shared by the computational linguistics and Sanskrit NLP research communities.

References

Dhananjaya Adiga, C. S. Anoop, and A. G. Ramakrishnan. 2021. Automatic speech recognition in sanskrit: A new speech corpus and modelling insights. In Findings of ACL.

James Allan. 2004. Perspectives on information retrieval and speech. Information Retrieval.

C. S. Anoop and A. G. Ramakrishnan. 2021a. Ctc-based end-to-end asr for low-resource sanskrit with spectrogram augmentation. In Proceedings of the National Conference on Communications (NCC).

C. S. Anoop and A. G. Ramakrishnan. 2021b. Investigation of grapheme-to-phoneme schemes for speech recognition in sanskrit. In International Conference on Neural Information Processing (ICONIP).

Saurabh Arora et al. 2025. Efficient and reproducible biomedical question answering using retrieval augmented generation. arXiv:2505.07917.

B. Balaji et al. 2025. Anveshana: A benchmark dataset for cross-lingual information retrieval on english queries and sanskrit documents. arXiv:2505.19494.

Pawan Goyal, Gérard Huet, Amba Kulkarni, Peter Scharf, and Ralph Bunker. 2012. A distributed platform for sanskrit processing. In Proceedings of the 24th International Conference on Computational Linguistics (COLING), Mumbai, India.

Gérard Huet. 2002. The zen computational linguistics toolkit: Lexicon structures and morphology computations using a modular functional programming language. In Language Engineering Conference (LEC) Tutorials, Hyderabad, India.

Jeff Johnson, Matthijs Douze, and Hervé Jégou. 2019. Billion-scale similarity search with gpus. IEEE Transactions on Big Data.

R. Kumar et al. 2024. Linguistically informed automatic speech recognition in sanskrit. SSRN.

Luca Lugli et al. 2022. Embedding models for buddhist sanskrit. In Proceedings of the Language Resources and Evaluation Conference (LREC).

National Institute of Indian Medical Heritage. 2023. e-samhita: Digital charaka saṃhitā edition. `https://niimh.nic.in/ebooks/ecaraka/?mod=read`. Accessed: 2026-02-28.

Google Research. 2023. Speech-to-retrieval (s2r): A new approach to voice search. Blog post.

Stephen Robertson and Hugo Zaragoza. 2009. The probabilistic relevance framework: Bm25 and beyond. Foundations and Trends in Information Retrieval.

Peter Scharf and Malcolm Hyman. 2009. Linguistic Issues in Encoding Sanskrit. Motilal Banarsidass, Delhi, India.

Georgios Sidiropoulos, Nikos Voskarides, and Evangelos Kanoulas. 2022. On the impact of speech recognition errors in passage retrieval for spoken question answering. In Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM).

Ville Turunen. 2008. Reducing the effect of oov query words by using morph-based spoken document retrieval. In Proceedings of Interspeech.

L. van der Werff, D. A. van Leeuwen, and F. de Jong. 2011. Speech transcript evaluation for information retrieval. In Proceedings of Interspeech.