



AI-Driven Insights into Sangam-Era Tamil and English Words: Tracing Phonetic Correlations with Shared Contextual Meanings for Non-Native Tamil Learners

Authors:

Kalaivani Bharanitharan, Kirikujalambal Thillai Chidambaram

Dr. Swaminathan Subramanian

SILAMBU, non-Profit Organization, Michigan, USA

sswami@msu.edu +1 630 956 8225

APA Citation:

Name, Title, Journal of Indian Languages and Indian literature in English, 0(0),Page number; year of publication

Submission Date:

Acceptance Date:

Abstract

Language evolution is a dynamic process shaped by phonetic shifts, cultural exchanges, and historical migrations. While the etymological roots of many Indo-European languages are well-documented, the contributions of Sangam-era Tamil to modern languages remain underexplored. This paper introduces an AI-powered chatbot designed to help children and non-native Tamil speakers discover ancient Tamil words embedded in English, Germanic, and Latin languages through phonetic and semantic analysis. By leveraging advanced linguistic pattern recognition, AI-based etymology tracing, and references from Sangam literature, the chatbot offers an interactive and educational experience for young learners.

The chatbot identifies Tamil-origin words in English by applying systematic phonetic transformation rules, including:

- Consonant shifts (e.g., K → G, P → F/B, T → D, L → D).
- Vowel modifications reflecting natural linguistic evolution.
- Metathesis and syllabic restructuring, where sounds swap positions.

For instance, the English word “honey” (Old English: hunig) closely resembles the Tamil word “பூ நெய்” (phu ney, meaning honey in Sangam vocabulary). Similarly, the word “star” (German: Stern, Old English: steorra) shares phonetic and semantic similarities with the Tamil word “தாரை” (Thārai). Words like “swing” (Proto-Germanic: swungen) also bear a striking resemblance to the Tamil word “ஊஞ்சல்” (Wungel).

By drawing on Sangam-era texts, the chatbot establishes historical linguistic connections, providing children with an engaging platform to explore Tamil’s influence on global languages. This will help the learners to understand the Sangam era vocabulary and see the potential of Tamil language. This could lead to totally different research of language connection. This is a novel approach and there are no references in this context. So authors are unable to produce the references for this research paper.

1. Introduction

The origins of language development are deeply rooted in phonetic transformations and cultural interactions. Although Indo-European linguistic history is well documented, the contributions of Dravidian languages, particularly Sangam-era Tamil, remain less explored. Many Tamil-origin words have seamlessly integrated into English and Germanic languages through phonetic shifts and cultural exchanges over centuries.

This research introduces an AI-driven chatbot that enables young learners to discover Tamil etymological roots in English. The chatbot employs phonetic transformation analysis, corpus-based AI search, and cross-referencing with Sangam literature to validate linguistic links. This interactive learning tool not only enhances Tamil language preservation but also fosters appreciation for its historical significance in global linguistics. This paper does not claim any direct linguistic connections between Tamil and Germanic or Latin languages. However, the authors observed striking phonetic patterns where the meanings of words remain the same.

This paper surveys existing AI-based techniques, evaluates their contributions, and outlines potential research directions using state-of-the-art AI methodologies [1-7].

1.1. Methodology

The application of Artificial Intelligence (AI) and Machine Learning (ML) in language learning has opened new avenues for analyzing its phonetics, structure and usage. The chatbot integrates three core mechanisms:

Phonetic Transformation Analysis

The AI system identifies Tamil-origin words based on established phonetic transformation rules:

- Consonant effects: The consonants are lost at the beginning or added at the end
- Vowel Modifications: Vowels are lost at the beginning or affinity to consonants G/H or S
- Metathesis and transposition effect: Sound rearrangements where phonemes shift positions, or letters swap places

Corpus-Based AI Search

The chatbot incorporates an AI-based search mechanism trained on a shown on Figure 1

- A verified database of Tamil-English word pairs.
- Computational linguistic models recognizing phonetic patterns.
- AI-assisted confidence scoring to determine potential linguistic correlations.

Sangam Literature Cross-Referencing

The AI system retrieves historical word usage from digitized Sangam-era texts to validate Tamil's influence in modern languages. This historical validation ensures accurate etymological tracing and reinforces linguistic connections with literary

Limitations:

Clustering only reveals structural patterns, phonetics and meaning, not the linguistics connection. Requires validation from linguistic data.

2. Implementation Strategy

This AI bot development has planned for multiple phases.

Data Preparation Phase

Corpus Development: Extract and digitize Sangam literature from historical texts.

Training Data Creation: Annotate Tamil-English word pairs with phonetic transformation rules.

Model Development Phase

LLM Adaptation: Fine-tune the AI model using Tamil etymology datasets.

Rule Engine Implementation: Develop algorithms for phonetic transformation rules.

Retrieval System Integration: Link etymological findings with Sangam literature references.

User Experience Development

Educational Progression Design: Develop interactive lessons for step-by-step learning.

Interface Development: Implement a chatbot with visual phonetic transformation displays.

3. Tamil-English Phonetic Transformation Rules for AI Chat Bot Examples

Consonant Effects

- Tamil ‘பூ நெய்’ → English ‘Phu Nei’ □ Huney □ Honey (P→ vanished)
- Tamil ‘கிழார்’ → English ‘KLor □ Lord ’ (K→ vanished)
- Tamil ‘கசிவு’ → English ‘kaseev’ □ Seep (K→ vanished & V → P transformation)
- Tamil ‘கவி’ → English ‘scop’ (S-prefixing transformation & V → P transformation)
- Tamil ‘முடம்’ → English ‘Mute’ (Lost “m” at the end)
- Tamil ‘சுவை’ → English ‘savor’ (V → P transformation)
- Tamil ‘எழுது’ → English ‘edit’ (Phoneme shift L->D)

Vowel Modifications

- Tamil ‘பழம்’ → English ‘pple’ □ Apple (A→ Added)
- Tamil ‘அரிசி’ → Latin ‘Arice’ □ Rice (A→ Vanished)
- Tamil ‘இஞ்சி’ → Latin ‘ginger’ (E/I/Y vowel affinity to G)
- Tamil ‘ஈனு’ → German ‘gene’ (E/I/Y vowel affinity to G/J)
- Tamil ‘ஊக்கு’ → English ‘hook’ (O/U vowel affinity to H)
- Tamil ‘ஊளை’ → English ‘howl’ (O/U vowel affinity to H)

Structural Changes

- Tamil ‘வளம்’ → Tamil ‘Wellam’ □ Wela (old English) □ Weal - Metathesis effect; L placement
- Tamil ‘மகிழு’ → Tamil ‘ma-gilu’ □ gliu (old English) --Metathesis effect; L placement
- Tamil ‘என்பு’ → Tamil ‘N-bo’ □ bo-n □ bone - Transposed effect
- Tamil ‘மடை’ → Tamil ‘Ma-Da’ □ Da-Ma □ Damaz □ Dam - Transposed effect

Each method provides partial insights, but no approach has fully bridged structural analysis with semantic meaning.

4. Educational Significance and Impact

The chatbot serves as an educational tool for:

- **Enhancing Tamil Linguistic Awareness:** Helps young learners discover Tamil’s presence in global languages.
- **Bridging Ancient and Modern Languages:** Strengthens linguistic and cultural ties.
- **AI-Powered Language Learning:** Uses AI to make etymology interactive and accessible.
- **Preserving Tamil Heritage:** Encourages language preservation through digital tools.

5. Conclusions

AI and ML have significantly advanced Language learning methodologies. The AI-driven chatbot enhances Tamil linguistic exploration by unveiling Tamil-origin words in English through phonetic transformations and historical references (Figure 2). This research underscores Tamil’s deep-rooted presence in global linguistics, reaffirming its status as one of the world’s most influential classical languages.

By making etymology interactive and accessible, this chatbot fosters greater appreciation for Tamil’s historical significance while preserving its linguistic heritage for future generations.

6. Reference

1. Rajasekar, N. (Year). Tamil Language Processing with Artificial Intelligence: Building NLP Applications.
2. Rajendran, V. (2023). Deep Learning for Tamil Computing.
3. Tamil Virtual Academy. (2022). Tamil Computing: From Research to Applications.
4. Sobha, L., & Ram, V. S. (Eds.). (2021). Computational Linguistics for Dravidian Languages.
5. Anand Kumar, M. (2019). Tamil Morphological Analyzer: Design and Implementation.
6. Jeyapradha, M. Tamil Computing - Techniques, Benefits, and Challenges.
7. Sobha, L., & Ram, V. S. (Eds.). (2021). Computational Linguistics for Dravidian Languages.

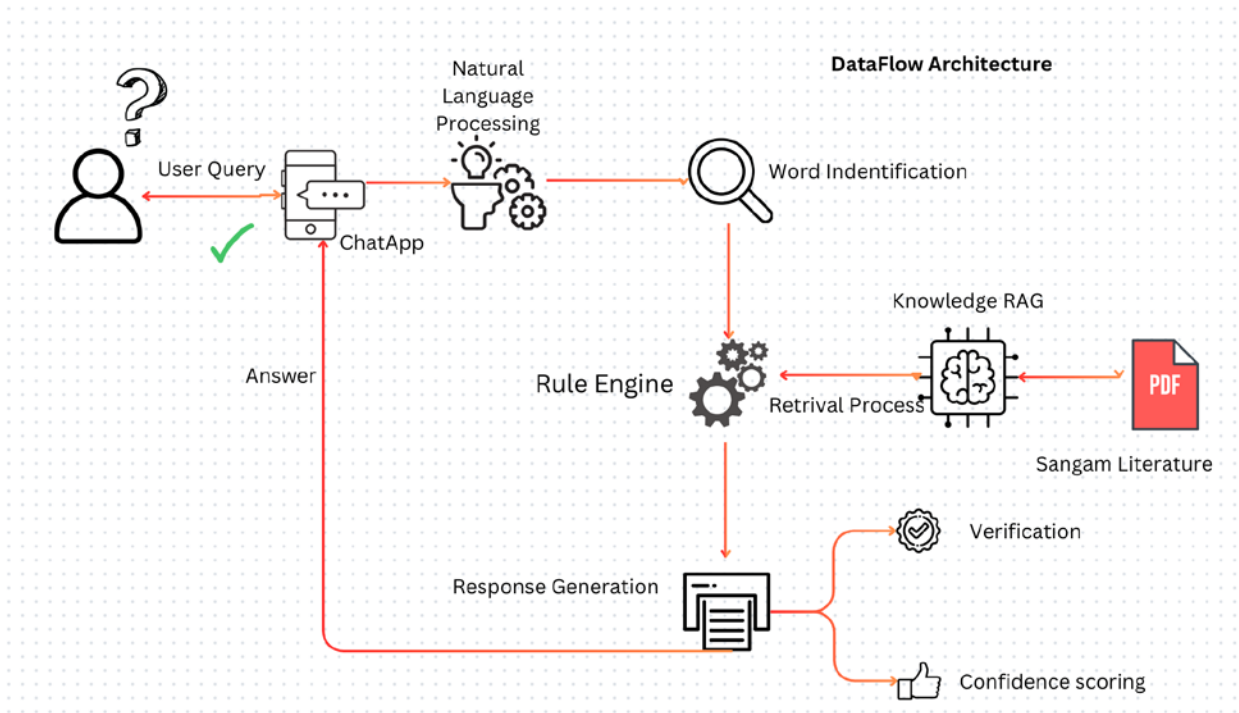


Figure 1: Data Flow Architecture

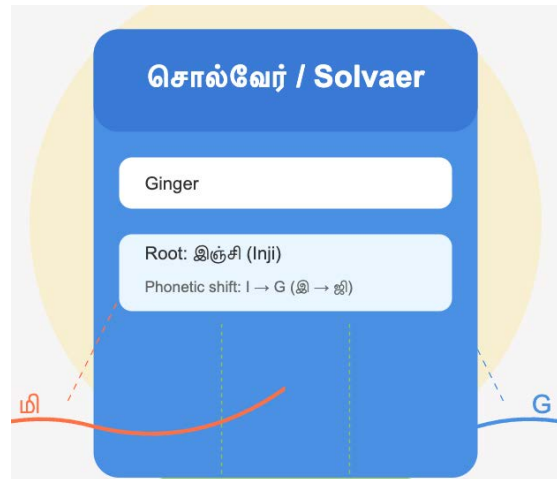


Figure 2: Proposed AI Bot