Kelvi: A Morphological Parser to Support Tamil Literacy

Shankhalika Srikanth Sabrina Yu Sophia Chan Madeline Solis de Ovando

Overview

- Tamil is highly diglossic and has agglutinative properties.
- Kelvi.ca breaks down Tamil words to help learners recognize morphological patterns.

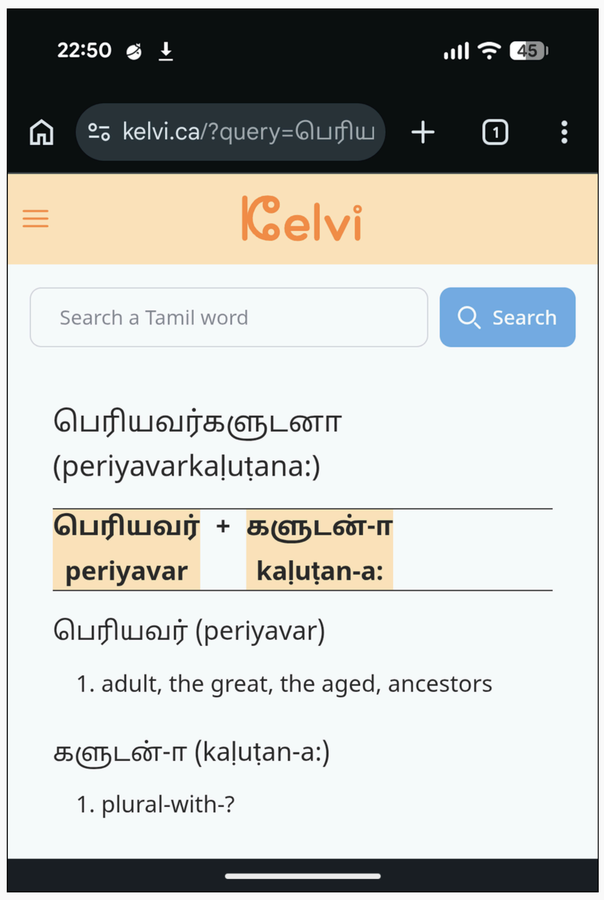


Fig: Translation: ”with the elders/adults/ancestors?”

Systems Design

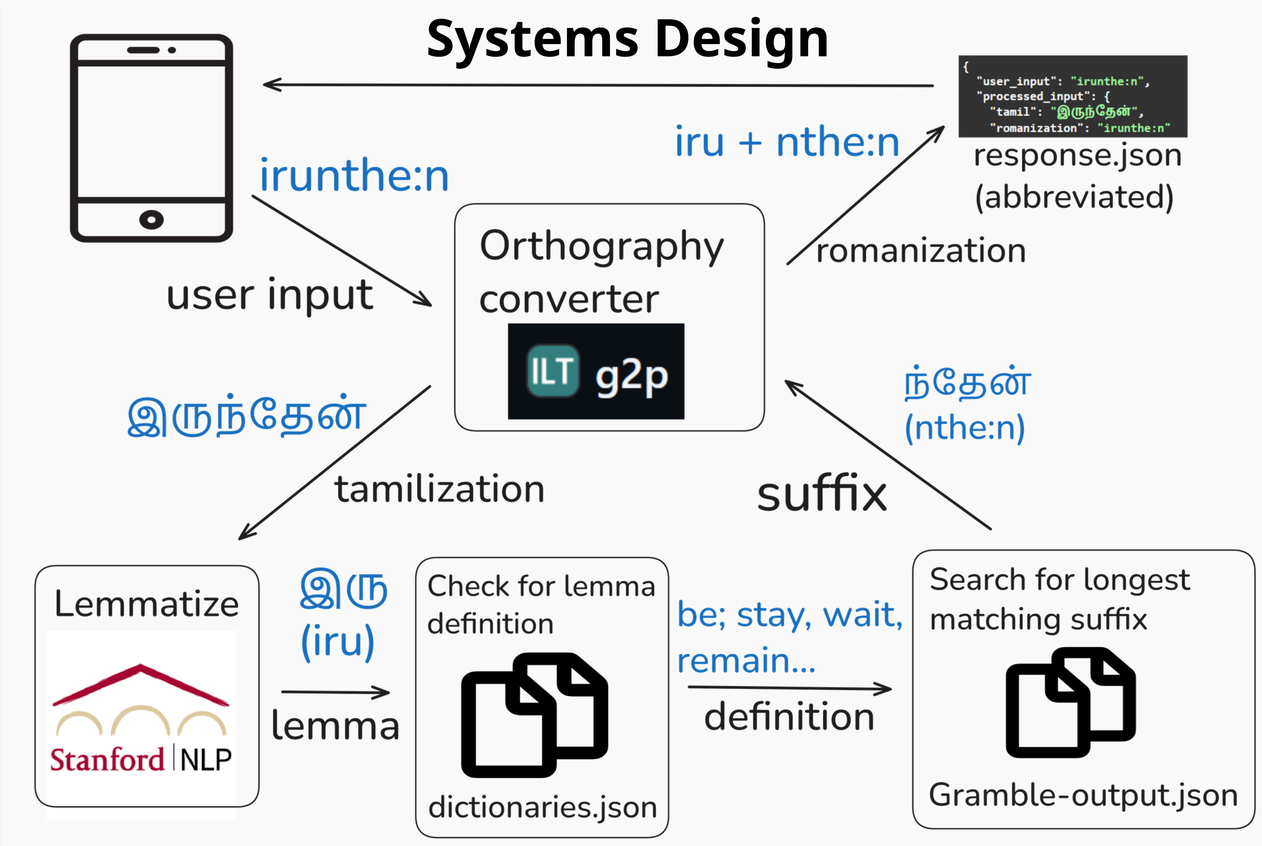


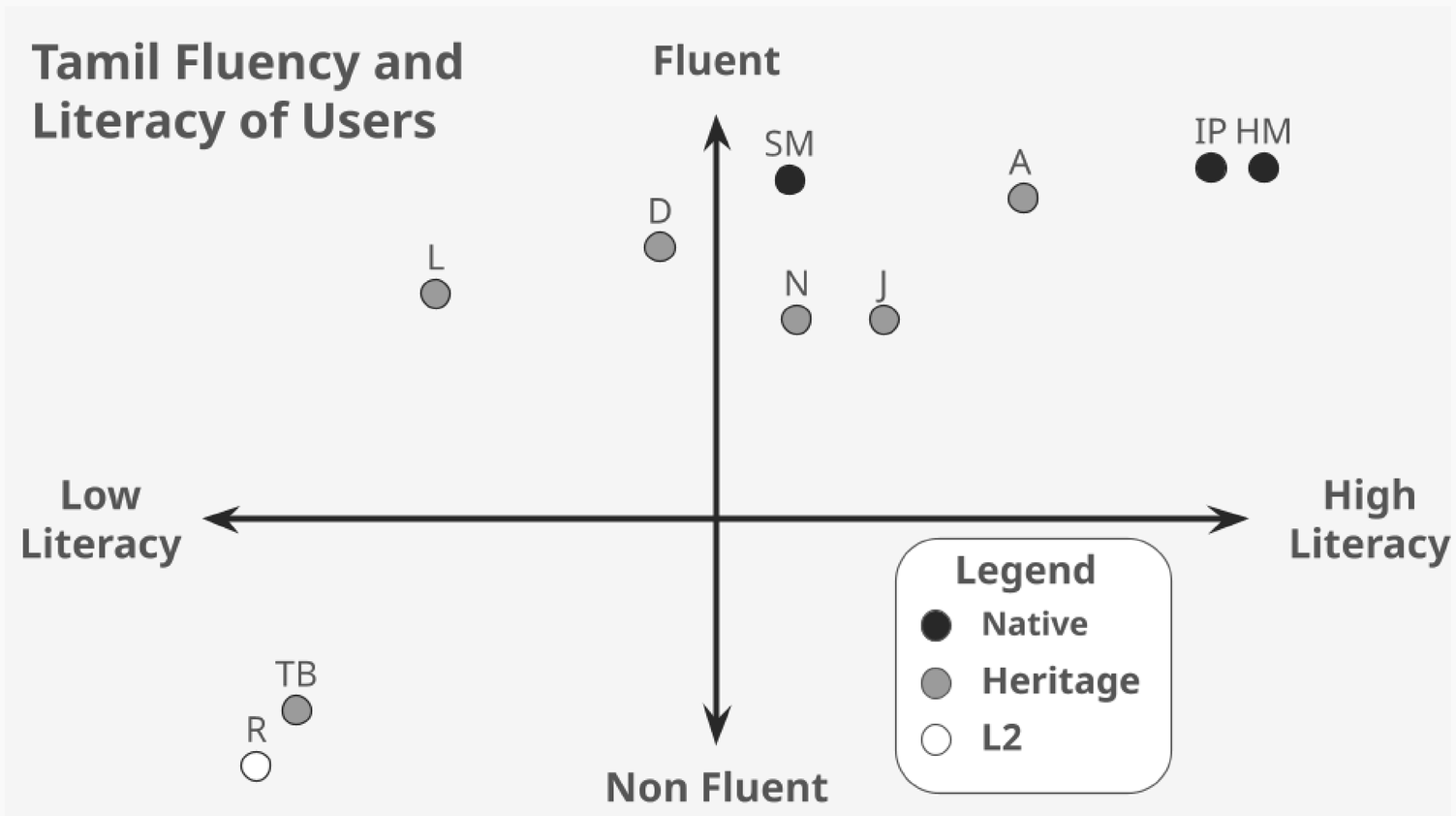
Fig: Diagram of Kelvi’s frontend and backend.

User Research

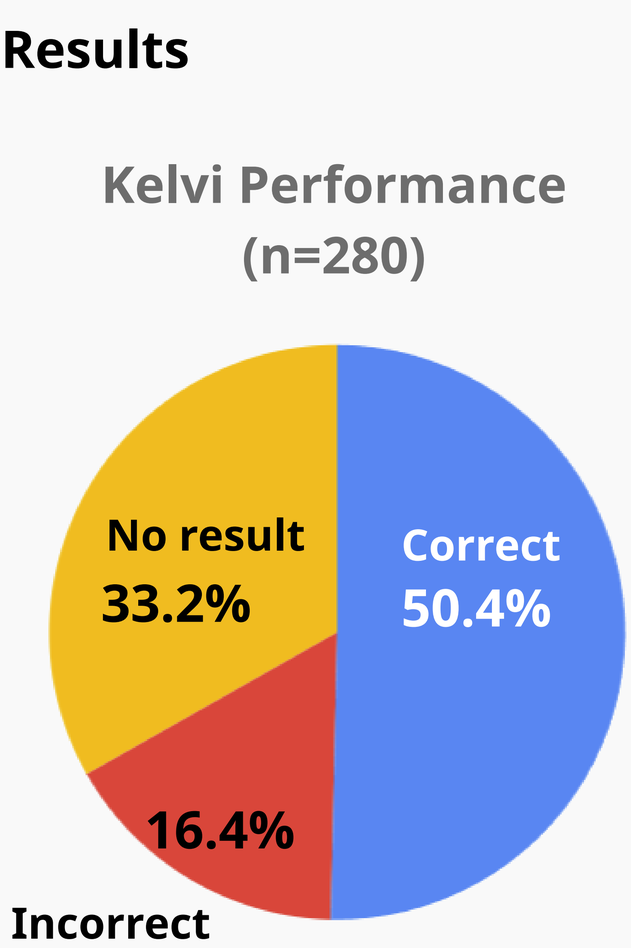
- We interviewed (n=10) adult Tamil speakers.
- Most common method reported for looking up Tamil words was Google Translate (n=8).
- After interacting with our medium fidelity prototype, 70% of users reported they would use the tool if it existed.

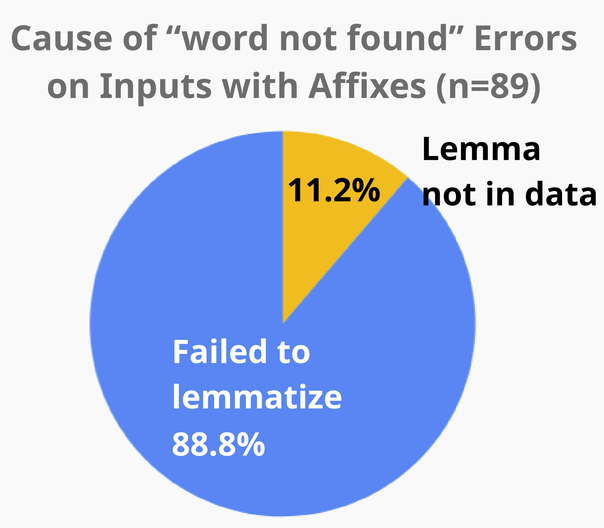
Based on exploratory interviews, we established the following user requirements:

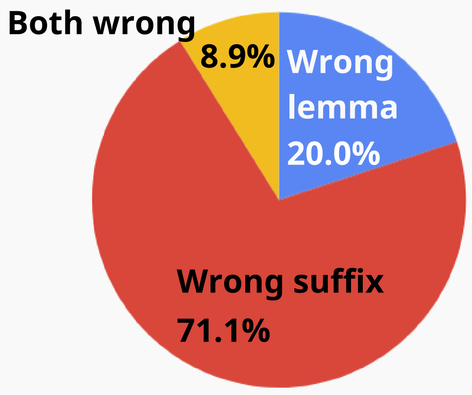
1. Intuitively understand the morphological breakdown even as an L2 learner.
2. Not overwhelm a beginner learner.
3. Is accessible on the phone.



Results







Discussion and Future Work

Problem	Solution
Lack of root-word (i.e. dictionary) data	Search for more open-source data
Incomplete lemmatization	Compensate with Gramble
Homophonous morphemes	Use PoS tags where available in dictionary data / provide multiple definitions
Romanization input is lower performing	Incorporating transliteration models, such as the one developed by Madhani et al. (2023)
Spoken register forms are underrepresented	Add to Gramble through combination of linguistic research and native-speaker expertise

Access: intentionally open source so can be accessed by language community members as well as used by other minoritized and/or low-resource languages

Acknowledgements

This project was made possible with support from the Audrey Wearn Language Prize and the Future Design School. We would also like to thank Fizza Ahmed, Vasumathi Srikanth, Srikanth Mangalam, Aidan Pine, Patrick Littell, Sowmya Vajjala, Roland Kuhn, and the Indigenous Languages Technology project team at the National Research Council of Canada. An additional thanks to Mark Zanon, Monique Yu, and Johnathan Steffen. Finally, thank you to the anonymous reviewers for their feedback.

References

1. Yash Madhani, Sushane Parthan, Priyanka Bedekar, Gokul Nc, Ruchi Khapra, Anoop Kunchukuttan, Pratyush Kumar, and Mitesh Khapra. 2023. Aksharantar: Open Indic-language transliteration datasets and models for the next billion users. In Findings of the Association for Computational Linguistics: EMNLP 2023, pages 40–57, Singapore. Association for Computational Linguistics.