PROSIDING
SENALA
Seminar Nasional
Linguistik Indonesia

Available online:
https://senala.upnjatim.ac.id

Prodi Linguistik Indonesia
Universitas Pembangunan Nasional "Veteran" Jawa Timur
Jl. Rungkut Madya, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294

Submitted: 6/10/2025 Accepted: 8/11/2025 Published: 20/11/2025

Research Article

Revitalizing Mandailing through Technology: Designing an AI-Based Trilingual Digital Dictionary

Muhammad Hasyimsyah Batubara[1*], Fitri Ayu[2]
[1] *Sekolah Tinggi Agama Islam Negeri (STAIN) Madina, Sumatera Utara, Indonesia*
[2] *Universitas Negeri Medan, Sumatera Utara, Indonesia*
*Correspondence Author, E-mail: muhammad.hasyimsyahbatubara@gmail.com

Abstract

Background: Bahasa Mandailing is facing a serious threat from globalization, and the major language is starving its development. As a low-resource language, preservation efforts are still limited and have not yet used digital technology systematically. On the other hand, the development of artificial intelligence (AI) opens new opportunities in documenting and revitalizing minority languages. Purpose: This study aims to develop a trilingual digital dictionary (Mandailing–Indonesian–English) as a strategic effort to preserve Mandailing while simultaneously providing theoretical contributions to linguistic studies and technological applications. Methods: The research method uses a Research and Development (R&D) approach with the ADDI (Analysis, Design, Development, Implementation) model. The research process includes needs analysis, dictionary entry structure design, development of a React Native/Flutter-based application prototype, and limited trials involving native speakers and university students. Results: The study results produced an application prototype with a three-way search feature, contextual sentence examples, and systematic dictionary entries. A total of 200 basic Mandailing vocabulary words were successfully collected as an initial corpus. The trials showed that the application is easy to use, although further development is needed in the form of vocabulary expansion and the addition of audio features. Conclusion: These findings suggest that creating an AI trilingual digital dictionary could be vital for preserving the Mandailing language. In addition to providing a practical learning tool, this initiative promotes digital literacy, strengthens the transmission of knowledge across generations, and serves as a foundation for research and technological innovation in linguistics.

Keywords: Mandailing, artificial intelligence, digital dictionary, language preservation, R&D

INTRODUCTION

Globalization and the dominance of majority languages have put minority languages worldwide in danger in the digital era. On the other hand, the rapid advances in technological approaches, especially artificial intelligence (AI), have opened new avenues for language documentation and revitalization and provide unprecedented resources (Bird, 2009; Low et al., 2022; Pradhan & Dey, 2023; Moshagen et al., 2024). For example, AI builds a digital dictionary with multiple languages to preserve endangered languages such as Mandailing by

exploring its function as cultural identity and national heritage (Asrif, 2019; Baunvg et al., 2025). Multiple studies found that using technology, particularly Natural Language Processing (NLP), which learnt at the linguistic level on a large scale and identifies patterns and creates automatic annotations, could help speed up the preservation of languages and cultures (Hanandeh et al., 2024; Curry et al., 2023).

Although the use of technology for language preservation has been applied in several world languages, most NLP research and development still focuses on high-resource languages such as English and Mandarin, leaving low-resource languages like Mandailing underserved (Gutiérrez-Fandiño et al., 2023; Gurgurov et al., 2024). Furthermore, issues such as limited linguistic data, the "curse of multilingualism," and potential bias in data processing add to the complexity of the challenges faced (Nimb et al., 2024; Anik et al., 2025). To date, few systematic initiatives have been integrating technology in developing digital corpora and multilingual dictionaries specifically for Mandailing.

In an effort to overcome the limitations of existing datasets, reduce its biases, and implement various contemporary techniques like transfer learning, transliteration, and modular multilingual architecture (Gurgurov et al., 2024), we present a novel approach in this study by proposing the building of a trilingual Mandailing digital dictionary (Bahasa Mandailing–Indonesian–English) based on FLEx – FieldWorks Language Explorer. Meanwhile, this approach is linguistic documentation and an interactive tool for education for the broader community, especially the younger generation, as encouragement to maintain and sustainably propagate the Mandailing language (Huda & Suwahyu, 2024).

Linguistic Authenticity vs Technological Mediation, in terms of digital intervention, linguistic authenticity and technological mediation build up a tension that becomes crucial. AI and NLP provide efficiency in documenting a language, but risk undermining cultural ownership and indigenous epistemologies. Hence, this study(c) has two critical aspects: observing the development of digital dictionaries as a technical development, as well as a negotiation between the forces of cultural preservation and cultural globalization, through technical intervention (Pradhan & Dey, 2023; Bird, 2009).

This research aims to build a dictionary by creating a trilingual dictionary (Bahasa Mandailing–Indonesian–English) as an open linguistic resource. Common challenges for documenting minority languages will also be discussed to seek practical contributions in preserving Mandailing and theoretical contributions related to the work in linguistics, AI, and minority languages documentation. Hence, this research should be a strategic move towards the application of AI technology with the aim of endangered language preservation and a contribution to literature genres at the interface of linguistics, technology, and culture.

METHODS

The research model used in this research is a research and development (R&D) model using the ADDI model (analysis, design, development, implementation). Despite this, the studies are restricted to the prototype development stage, whilst implementation and evaluation remain scalable only.

1. Analysis

a. Before delving into the project, the following section outlines a literature review on digital dictionary development and local language preservation.
b. Conducted preliminary discussions with language lecturers, students, and native Mandailing speakers to identify user needs.
c. Outcome: a Bahasa Mandailing–Indonesian–English trilingual dictionary (basic search function and contextual examples were the main need).

2. **Design (Planning)**
   a. Designed the dictionary entry structure: Mandailing lemma, Indonesian equivalent, English equivalent, category, and sample sentences.
   b. Created mockup UI/UX sketches for a simple Android application.
   c. Determined prototype technology: simple vocabulary database (JSON/SQLite) and application framework (React Native/Flutter).
3. **Development (Prototype Creation)**
   a. Compiled an initial vocabulary corpus of approximately 200 basic Mandailing words (family, nature, culture).
   b. Built a mobile application prototype with features:
      1) Trilingual word search.
      2) Simple dictionary entry display.
      3) Contextual sample sentences.
   c. Advanced features (e.g., audio, NLP integration) were excluded since this is still a prototype.
4. **Implementation (Limited Trial)**
   a. Conducted trials with a small group of 5–10 participants (students and native speakers).
   b. Trial objectives: to evaluate usability and clarity of word entries.
   c. Trial feedback was used to refine the interface and expand vocabulary coverage.
5. **Evaluation (Prototype Assessment)**
   a. Formative evaluation was conducted during the design and development phases (to ensure relevance to user needs).
   b. Summative evaluation on a limited scale: assessed whether the prototype helps users quickly find basic Mandailing vocabulary.
   c. Evaluation indicators: user satisfaction, accuracy of vocabulary display, and potential for further development.

This trilingual design redefines traditional lexicography by integrating linguistic documentation with AI-based automation, demonstrating a paradigm shift from manual corpus construction to adaptive, machine-assisted language preservation.

RESULTS AND DISCUSSION

This study's results show that developing a trilingual digital dictionary (Mandailing–Indonesian–English) through an R&D approach with the ADDI model produced several important achievements.

1. **Analysis Stage**

The literature review results and initial discussions with lecturers, students, and native Mandailing speakers confirmed the primary need for a trilingual dictionary with a simple search feature, contextual example sentences, and a display that is easily accessible via mobile devices. As mentioned in the introduction, it reinforces the urgency of the research that Mandailing, as a low-resource language, requires technological support for its preservation.

2. **Design Stage**

This system contains the construction of the structure of the dictionary entries which have Mandailing entries, the one which is called the target language or Indonesian entries, the source language or English entry, the terminology (whether it belongs to the category of general, basic, and/or scientific vocabulary) in which the dictionary entries belong to, and the sourced sentences coming from the phrases in which the Mandailing entry can be used. The designs of UI/UX sketches for the Android application were done; users can move from one screen to another without any trouble.

3. Development Stage

A small, complete lexicon of around 200 basic terms (including family, nature, and cultural references) was compiled and ingested in a simple database (JSON/SQLite). Later, a React Native/Flutter-based prototype mobile application with essential features, such as three-way search, displayed vocab entries, and contextual example sentences.

4. Implementation Phase (Limited Trial)

A few trials were conducted on 5–10 participants (both students and native speakers) initially. The app was considered easy to comprehend and navigate; however, it received feedback (F3, F4, and F9) on broadening the vocabulary and adding audio pronunciation features.

Figure 1. Front view of the Mandailing–Indonesian–English digital dictionary application

Figure 2. Display in the Mandailing–Indonesian–English digital dictionary application

These findings not only show the technical feasibility of the prototype but also reveal the epistemological implications of integrating AI into linguistic preservation. The development of a trilingual model challenges traditional notions of authenticity by allowing multilingual interaction and contextual adaptation. Compared to global AI-based lexicographic projects such as the Livonian and Kawi dictionaries, this study highlights Indonesia's specific socio-technological challenges, limited data infrastructure, and low digital literacy among rural speakers, which call for contextually sensitive solutions. To ensure sustainability, this initiative

should be expanded through open-source collaboration and institutional partnerships (e.g., local universities, linguistic research centers, and cultural heritage offices). However, ensuring that digital preservation is relevant requires community participation and government policy support within a wider mosaic of cultural revitalization programs.

Discussion

It is hoped that with this innovation, the digital era will be a connecting tool in maintaining and restoring the endangered Mandailing language by creating a digital dictionary of the Mandailing language (Wills et al., 2019). This current project targets the development of a trilingual (Bahasa Mandailing, Indonesian, English) digital lexicon by harnessing the power of modern technology to assist the collection and documentation of a full-spectrum Mandailing dictionary (Nimb et al., 2024). Such innovation aligns with preserving the Kawi language using an electronic translation application that can perform word-for-word, even sentence and paragraph-based Kawi language translations, with an accuracy of up to 82.27% (Sudana et al., 2017). Such a digital technology innovation in preserving languages corresponds to a worldwide trend where artificial intelligence and web technologies have been used to document and preserve endangered cultural heritage, as illustrated by the Ancient Olympia: Common Grounds project, by Low et al. (2022).

This trilingual Mandailing digital dictionary also aligns with the focus of accelerating digital transformation in infrastructure and human resources development (Prof Dr et al., 2023) by improving human resources' digital literacy and cognitive functions to master information technology as a productive tool. The method also helps them access the lexical and grammatical richness of the Mandailing language and attracts younger generations who are digital native users (Qulub & Renhoat, 2020). AI and machine learning can aid this process by optimizing information discovery from recordings that are not currently widely utilized, can help speed up the documentation of endangered languages, and may facilitate other approaches toward the ongoing need to document more languages than currently being done (Baunvig et al., 2025). It is also part of the wider trend of novel approaches to dictionary-making for other minority languages, like incorporating a Livonian corpus into digital formats to promote a multi-use nature (Baunvig et al., 2025).

It is thus vital that the digital technology based on this data plays its role, especially with all the language preservation challenges (ex, low resources, risk of losing cultural heritage due to assimilation, and lack of socialization) (Low et al., 2022). As such, the team hopes that the development of a trilingual Mandailing digital dictionary will be an impetus for language revitalization, but at the same time provide a basis for comparative linguistic research and for the creation of adaptive learning tools. Adopting this digital dictionary allows more open access for researchers, educators, and society to help accelerate the revitalization and intergenerational transmission of Mandailing (Musaddat et al., 2021). But although the digital divide is GDP still problematic in Indonesia, especially in rural 4.0 development, where infrastructure and access to digital devices are scarce, we see these opportunities in urban life, on paper. Thus, to realize such a digital dictionary in the Mandailing language, an inclusive strategy is also needed so that all Mandailing speakers can feel the impact of this digital dictionary equally (Subroto et al., 2023). Hence, bouncing back from the consequences of this act must also cover digital literacy and equal access provision to make the technology used in the Mandailing language (Lestari et al., 2023).

CONCLUSION

This study presents evidence of the implementation of an AI-based trilingual digital dictionary (Bahasa Mandailing–Indonesian–English) and its significance in developing a low-resource Mandailing language preservation. By following a research and development (R&D) methodology, called the ADDI model, we devised a prototype application that includes several features, such as 3-way vocabulary search and contextual sentence examples. In limited trials, testers found it easy to use and beneficial, but additional vocabulary and pronunciation audio resources need to be added.

Academically, this work contributes to meaning in linguistics, documentation of minority languages, and Artificial intelligence for cultural preservation. For the community, particularly the younger generation of Mandailing, practically, the results of this study can support the preservation of their access to their own ancestral language and their endeavour to teach it. It will be tremendous, but it will not be for everyone unless we address the digital gap. Therefore, this trilingual digital dictionary should become a core element in sustainable Mandailing language reclamation and a gateway for further technology-aided learning.

Next steps need to include expanding the lexical corpus, adding speech and audio data, and creating institutional and community partnerships to ensure long-term sustainability. These will help the Mandailing digital dictionary to grow as a linguistic resource and a living cultural archive.

ACKNOWLEDGEMENT

The author wishes to thank all parties who have contributed to this research.

ETHICAL APPROVAL

The author would like to thank the software Quilbort, Turnitin, and other digital platforms for their assistance in this research. We wish to declare that all are free of technical and financial support during the research work.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

FUNDING INFORMATION

This research was done independently, with no financial sponsorship or funding. It was achieved through conscientious efforts to create specific activities and outputs that would benefit society and government, study linguistics, and practice local language preservation in a digital age.

REFERENCES

Anik, M. A., Rahman, A., Wasi, A. T., & Ahsan, M. M. (2025). *Preserving Cultural Identity with Context-Aware Translation Through Multi-Agent AI Systems*. 51. https://doi.org/10.18653/v1/2025.lm4uc-1.7

Asrif, N. (2019). Pembinaan dan Pengembangan Bahasa Daerah dalam Memantapkan Kedudukan dan Fungsi Bahasa Indonesia. *MABASAN*, *4*(1), 11. https://doi.org/10.26499/mab.v4i1.183

Baunvig, K. F., Tafdrup, J., Vad, K., Rasmussen, K. S. G., & Väina, M. (2025). From Bias to Insight : Computational Challenges and Opportunities in the Humanities. *Research*

Portal Denmark, 33. https://local.forskningsportal.dk/local/dki-cgi/ws/cris-link?src=au&id=au-ceb30ae6-78b2-4acf-9fd4-890584aaeed2&ti=From%20Bias%20to%20Insight%20%3A%20Computational%20Challenges%20and%20Opportunities%20in%20the%20Humanities

Baunvig, K. F., Tafdrup, J., Vad, K., Rasmussen, K. S. G., & Väina, M. (2025). From Bias to Insight : Computational Challenges and Opportunities in the Humanities. *Research Portal Denmark*, 33. https://local.forskningsportal.dk/local/dki-cgi/ws/cris-link?src=au&id=au-ceb30ae6-78b2-4acf-9fd4-890584aaeed2&ti=From%20Bias%20to%20Insight%20%3A%20Computational%20Challenges%20and%20Opportunities%20in%20the%20Humanities

Bird. S. (2009). Last Words: Natural Language Processing and Linguistic Fieldwork. *Computational Linguistics, 35*(3), 469–474. https://aclanthology.org/J09-3007.pdf

Curry, N., Baker, P., & Brookes, G. (2023). Generative AI for corpus approaches to discourse studies: A critical evaluation of ChatGPT. *Applied Corpus Linguistics*, *4*(1), 100082. https://doi.org/10.1016/j.acorp.2023.100082

Gurgurov, D., Bäumel, T., & Anikina, T. (2024). Multilingual Large Language Models and Curse of Multilinguality. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2406.10602

Gutiérrez-Fandiño, A., Pérez-Fernández, D., Armengol-Estapé, J., Griol, D., Kharitonova, K., & Callejas, Z. (2023). esCorpius-m: A Massive Multilingual Crawling Corpus with a Focus on Spanish. *Applied Sciences*, *13*(22), 12155. https://doi.org/10.3390/app132212155

Hanandeh, A., Ayasrah, S., Kofahi, I., & Qudah, S. (2024). Artificial Intelligence in Arabic Linguistic Landscape: Opportunities, Challenges, and Future Directions. *TEM Journal*, 3137. https://doi.org/10.18421/tem134-48

Huda, M., & Suwahyu, I. (2024). PERAN ARTIFICIAL INTELLIGENCE (AI) DALAM PEMBELAJARAN PENDIDIKAN AGAMA ISLAM. *REFERENSI ISLAMIKA Jurnal Studi Islam*, *2*(2), 53. https://doi.org/10.61220/ri.v2i2.005

Lestari, I. L., Fibriyani, V., & Zulfarosda, R. (2023). Menumbuhkan Literasi Digital pada Kalangan Ibu Rumah Tangga di Kecamatan Panggungrejo. *JMM - Jurnal Masyarakat Merdeka*, *6*(1), 72. https://doi.org/10.51213/jmm.v6i1.138

Low, D. S., McNeill, I., & Day, M. (2022). Endangered Languages: A Sociocognitive Approach to Language Death, Identity Loss, and Preservation in the Age of Artificial Intelligence. *Sustainable Multilingualism*, *21*(1), 1. https://doi.org/10.2478/sm-2022-0011

Low, D. S., McNeill, I., & Day, M. (2022). Endangered Languages: A Sociocognitive Approach to Language Death, Identity Loss, and Preservation in the Age of Artificial Intelligence. *Sustainable Multilingualism*, *21*(1), 1. https://doi.org/10.2478/sm-2022-0011

Moshagen, S. N., Antonsen, L., Wiechetek, L., & Trosterud, T. (2024). Indigenous language technology in the age of machine learning. *Acta Borealia*, *41*(2), 102–116. https://doi.org/10.1080/08003831.2024.2410124

Musaddat, S., Qodri, M. S., Atmaja, C., Murahim, M., & Mari'i, M. (2021). Pengembangan Program Literasi Bahasa Berbasis Kelas di Lombok Barat: Potret, Program, dan Kebutuhan Era Digital. *Jurnal Ilmiah Mandala Education*, *7*(1). https://doi.org/10.36312/jime.v7i1.1525

Nimb, S., Flörke, I., Olsen, S., Pedersen, B. S., Sørensen, N. C. H., Despot, K. Š., Anić, A. O., & Brač, I. (2024). COR.SEM, A NEW FORMAL SEMANTIC LEXICON FOR DANISH.

Research Portal Denmark, 131. https://local.forskningsportal.dk/local/dki-cgi/ws/cris-link?src=ku&id=ku-71db8ce9-c6e3-4f76-b590-a06297cc5166&ti=COR.SEM%2C%20A%20NEW%20FORMAL%20SEMANTIC%20LEXICON%20FOR%20DANISH

Nimb, S., Flörke, I., Olsen, S., Pedersen, B. S., Sørensen, N. C. H., Despot, K. Š., Anić, A. O., & Brač, I. (2024). COR.SEM, A NEW FORMAL SEMANTIC LEXICON FOR DANISH. *Research Portal Denmark*, 131. https://local.forskningsportal.dk/local/dki-cgi/ws/cris-link?src=ku&id=ku-71db8ce9-c6e3-4f76-b590-a06297cc5166&ti=COR.SEM%2C%20A%20NEW%20FORMAL%20SEMANTIC%20LEXICON%20FOR%20DANISH

Pradhan, U., & Dey, J. (2023). Language, artificial education, and future-making in indigenous language education. *Learning, Media and Technology*, *50*(2), 235–248. https://doi.org/10.1080/17439884.2023.2278111

Prof. Dr., I., Yani, A., Dr. Kurniawan, M. E. H., Wahyudin, N., Dr. Megawandi, Y., Dr. Wahyudi, S., Suhasri, Y., Lesmana, I., Prayoga, G. I., Kurnia, D., Arif, E., & Putra, A. L. (2023). *Fenomena Hidup Mewah (Hedonisme Dan Flexing) Para Aparatur Sipil Negara*.

Qulub, T., & Renhoat, S. F. (2020). *Penggunaan Media Padlet untuk Meningkatkan Keterampilan Menulis Teks Deskripsi*. https://jurnal.umj.ac.id/index.php/SAMASTA/article/download/7233/4461

Subroto, D. E., Supriandi, Wirawan, R., & Rukmana, A. Y. (2023). Implementasi Teknologi dalam Pembelajaran di Era Digital: Tantangan dan Peluang bagi Dunia Pendidikan di Indonesia. *Jurnal Pendidikan West Science*, *1*(7), 473. https://doi.org/10.58812/jpdws.v1i07.542

Sudana, O., Putra, I. K. G. D., Sudarma, M., Hartati, R. S., & Diyastama, I. P. P. (2017). E-Translator Kawi to Balinese. *International Journal of Engineering and Emerging Technology*, *2*(1), 46. https://doi.org/10.24843/ijeet.2017.v02.i01.p10

Wills, T., Jóhannsson, E. Þ., & Kosem, I. (2019). Reengineering an Online Historical Dictionary for Readers of Specific Texts. *Research Portal Denmark*, 116. https://local.forskningsportal.dk/local/dki-cgi/ws/cris-link?src=ku&id=ku-7939e0a9-c668-4e48-8964-4b3057785cb9&ti=Reengineering%20an%20Online%20Historical%20Dictionary%20for%20Readers%20of%20Specific%20Texts