Amulwe Kimün: A Community-Grounded Demo, Resource, and ASR Baseline for Mapuzugun

Cristian Ahumada Oliva, Fatiha Sadat
University of Quebec at Montreal
Montreal, QC. Canada
ahumada_oliva.cristian@courrier.uqam.ca, sadat.fatiha@uqam.ca

Abstract

This paper introduces Amulwe Kimün ("a means or path for knowledge" in Mapuzugun), a community-grounded multimodal quiz application co-created with Mapuche speakers to support the revitalization of Mapuzugun. Developed within a FACSO–CONADI collaboration during an intensive language course, the platform integrates multiple-choice, ordering and free-text exercises, as well as forums and chat functions to promote language practice, peer learning, and a sense of community. A pilot involving 32 learners produced 562 responses across 43 questions, with accuracies of 92.3% (multiple choice), 55.2% (ordering), and 7.1% (free-text), offering insights for refining item design and evaluation strategies. The low open-answer accuracy is related to a strict exact-match scoring and orthographic variation of the language. In addition, we present an initial Automatic Speech Recognition (ASR) prototype (Whisper-small + LoRA), establishing a fine-tuned baseline relative to zero-shot performance. The demo illustrates how community-grounded design, language resources, and lightweight evaluation can productively meet in a practical tool for an endangered language.

Keywords: Mapuzugun, Indigenous languages, endangered languages, educational technology, Natural Language Processing, multimodal learning, Automatic Speech Recognition.

1. Introduction

Mapuzugun (also written Mapudungun) has been classified as an endangered language since at least the early 20th century. As Pascual Coña stated in 1930, *"kalli rupape kiñe mufü tripantu, fey -tüfachi weche- epe kimwerpulayay ñi mapu dungun engün"* ("Let a few years pass—these young people will almost not know their *Mapuzugun*) (Coña, 2019). Eighty years later, UNESCO (2010) projected that 90% of the world's languages could disappear by the end of this century (UNESCO, 2010). A recent report by Chile's Ministry of Social Development and Family (Quiero, 2025) also highlighted the drastic decline in fluent Mapuzugun speakers(Ministry of Social Development and Family, 2024). Despite revitalization efforts, including language courses, publications, audiovisual projects, and online content, the number of high-level Mapuzugun speakers continues to decline. School policies related to the language seem to be insufficient, and the wisdom of *fütakeche* (Mapuche elders), which is crucial for preservation from a Mapuche point of view, is not always well documented. The recent passing of Rosendo Huisca, a respected *kimche* (elder and wise person) who contributed significantly to language revitalization, highlights this loss.

Several studies and initiatives across the Andes have explored Mapuzugun revitalization, educational materials, and its sociopolitical role (Mariano, 2021; Inostroza et al., 2021; Pell, 2023; Castillo et al., 2022; Rivas and Del Pino, 2023; Loncon et al., 2023). The University of Playa Ancha offers a formal translation program in Mapuzugun (Sarmiento, 2023), which provides some institutional support. At the same time, Mapuzugun is agglutinative and polysynthetic (Zúñiga, 2006), making it difficult to transfer tools and resources directly from other languages, while its minority status continues to limit its presence in digital environments.

This paper presents a web application designed to support Mapuzugun education through technology, developed collaboratively by the Faculty of Social Sciences (FACSO) of the University of Chile and the National Corporation for Indigenous Development (CONADI). The platform includes interactive forums, chat functions, and quizzes based on course content. In parallel, we report an initial ASR baseline as a preparatory step toward future oral-practice support; this ASR component was not yet deployed as an in-platform feature during the pilot reported here.

The paper is organized as follows: Section 2 reviews related work on computational and educational efforts for Mapuzugun. Section 3 briefly introduces the Mapuzugun language. Section 4 presents the project context. Section 5 details the methodology, including the study design and the intended role of ASR. Section 6 presents the evaluations and results. Section 7 concludes with perspectives for future work.

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 3360–3368
11–16 May 2026.

2. Related Work

Several technological efforts have been made to support Mapuzugun through computational tools and language resources. Early work is focused on linguistic analysis, like morphological analyzers (Chandía, 2022), which aim to parse and generate word forms for educational and research purposes (Ahumada et al., 2022). Complementary tools include systems for detecting writing systems (or alphabets) and providing translation assistance, which facilitate literacy and crosslinguistic access for learners and researchers. The AVENUE project represents another line of development, producing a range of digital resources for Mapuzugun; although much of its early outputs were not publicly accessible (Alvarez et al., 2005), recent initiatives have made portions of these data available (Duan et al., 2020), providing a foundation for future research and application development. At the Pontifical Catholic University of Valparaíso, the MAPU software platform was designed as a language learning tool incorporating voice recognition, although its functionality remains limited and it is not currently publicly available (Alvarado, 2012). Similarly, CEDETI at the Pontifical Catholic University of Chile developed Mapudungun mew, a software solution aimed at integrating technological tools for language education (UC). At the national level, CENIA (National Center for Artificial Intelligence) has also worked on developing translation systems for minority languages, including Rapa Nui and Mapuzugun. An initial version of their Mapuzugun translation tool is accessible online (CENIA, 2023), but it is now restricted, highlighting the challenges of sustaining long term public access to digital resources for endangered languages. Collectively, these efforts illustrate a growing but fragmented ecosystem of computational tools, revealing both the potential and limitations of current approaches to supporting Mapuzugun language learning, documentation, and revitalization.

3. Mapuzugun Language

Mapuzugun, also known as Mapudungun, is the language of the Mapuche people, a First Nation from south and central Chile and southwestern Argentina. It is spoken in southern and central Chile and in southwestern Argentina, and it is generally classified as an Amerindian language. It is considered agglutinative and polysynthetic, that means that it forms words through the combination of multiple morphemes, often bringing complex grammatical relationships within a single word (Zúñiga, 2006).

Like many Indigenous languages, Mapuzugun has a rich oral tradition and is considered a low-resource language. Historically, it has also lacked a standardized written form, which has created challenges for language documentation, literacy, and the development of digital tools. Writing systems in use include Latin alphabets, which vary regionally and even within communities, and efforts have been made to adapt orthographic conventions to accurately represent phonological variants. Dialectal variation is significant, with differences in pronunciation, morphology, and vocabulary between regions, particularly between communities in Chile and Argentina. Mapuzugun is generally treated as a language isolate (Sadowsky et al., 2013), which makes collaboration and technological development more difficult than for languages that belong to larger language families. However, many of these challenges are shared with other Indigenous languages of the Americas (Mager et al., 2018; Arppe et al., 2016; Camacho and Zevallos, 2019). Documentation gaps and lack of standardization classify Mapuzugun as a low-resourced language (Iturrieta et al., 2024; Molineaux, 2023). See also the public corpus resources (Duan and contributors, 2019).

From a typological perspective, Mapuzugun shows agglutinative features (Zúñiga, 2006), extensive verb morphology and frequent use of suffixes and enclitics to encode tense, aspect, mood, and argument structure (Smeets, 1989).

4. Project Context

The project, a collaboration between FACSO-UChile and CONADI, was designed to support a Mapuzugun course following a participatory methodology (Maldonado et al., 2024). The underlying language intervention that motivated this work was implemented as an immersion-oriented teaching program (Programa de Enseñanza del Mapuzugun, PEM; Mapuzugun Teaching Program). The program combined regular evening sessions with intensive full-day immersion sessions, and incorporated experiential activities such as field visits, cultural crafts, and guided conversations with Mapuche speakers and knowledge holders to anchor language practice in culturally situated contexts (Maldonado et al., 2024).

Participants were recruited through an intentional (non-probabilistic) sampling strategy with inclusion criteria focusing on: (i) Mapuche youth aged 22–37, (ii) gender representativeness, (iii) residence in the Metropolitan Region of Chile, and (iv) official Indigenous status certification. The resulting cohort comprised 32 participants (23 women and 9 men; mean age 27), with many having migrated to Santiago from southern regions of Chile for work or study (Maldonado et al., 2024). This initial phase resulted in a Mapuzugun Teaching Program, which outlined course content and allocated time for each

topic. Sessions included common introductory topics such as Chalintukuwün ("introducing oneself"), as well as family, cultural practices, history, music, traditional crafts, and visits to spiritually significant sites.

The program aimed to enhance both language skills and cultural knowledge through regular classes and intensive linguistic immersion, with orality playing a central role in revitalizing this "sleeping" language. Community building was also emphasized as a key outcome by participants (Maldonado et al., 2024).

As part of the project, a proof-of-concept (PoC) multimodal educational tool was developed to provide students with additional resources outside formal class sessions or field visits to culturally significant sites. The tool was co-created with a Mapuzugun-speaking developer, incorporating content from the Mapuzugun Teaching Program as well as materials provided by the native instructors. The primary goal was to reduce dependence on in-person instruction or existing transcriptions, enabling students to practice Mapuzugun independently—not as a replacement for live interaction, but as a supplementary resource to increase access to written and multimodal materials when speakers are unavailable.

5. Methodology

5.1. Study design and qualitative evaluation (program-level)

The following summarizes the evaluation design of the underlying teaching program (PEM) that motivated the platform (Maldonado et al., 2024). To evaluate implementation and outcomes, the program employed complementary qualitative methods, including interviews, observations, and audiovisual records (Maldonado et al., 2024). Data collection included 32 semi-structured interviews (one per participant), conducted individually at the university facilities, with a duration of approximately 45 minutes each. The program lasted four months and its evaluative phase was carried out in April 2024; interviews were recorded with participants' consent (Maldonado et al., 2024). In addition, 10 naturalistic observations were conducted in the context of Mapuzugun teaching between youth and Mapuche speakers and/or knowledge holders, complemented by audiovisual documentation (photos and videos). Audiovisual records covered multiple learning spaces used throughout the program (classrooms, library, patios, Mapuche homes, sanctuaries, and local gardens) (Maldonado et al., 2024).

Qualitative data were analyzed using a grounded-theory-oriented approach (Teorización Anclada), proceeding from inductive coding to identify emergent themes, followed by the construction of higher-level categories grouping codes by topic (Maldonado et al., 2024).

To align with the project goals and remain familiar to participants, the project leads, students, teachers, and the developer agreed to build a proof-of-concept quiz application. This tool enables Mapuzugun learners to practice reading and writing the language in their free time while interacting with other participants.

5.2. System overview and intended ASR integration

The platform was designed to extend course practice beyond scheduled sessions and to align with the program's emphasis on community building and orality (Maldonado et al., 2024). While ASR was not deployed as an in-platform feature during the pilot phase reported here, it is planned as an optional component to support oral practice outside the classroom. In future iterations, ASR-based transcriptions will be explored as lightweight feedback for short speaking and reading tasks, complementing teacher-led instruction. In this paper, we therefore report the current ASR baseline experiments as a preparatory step toward that integration.

5.3. Language quiz

A key feature of the application is its use of Mapuzugun as default. However, an option to switch to Spanish is included, as participants' level in Mapuzugun was initially insufficient for full immersion. This allows learners to begin in Spanish and gradually transition to Mapuzugun. Figure 1 illustrates the interface in both languages.

The quiz presents a randomly ordered set of questions in three formats: multiple choice, text completion, and sorting. In all formats, prompts are text-based, while multiple-choice answers can be text, audio, or images; sorting questions also support audio.

In the multiple-choice format, questions have 2 to 4 alternatives in Mapuzugun, except when the question involves translation. The text completion format requires students to fill in missing words in a given sentence, and the sorting format asks students to arrange numbered items in the correct order within a passage. Figure 1 illustrates examples of the questions.

Since the questions were provided by speakers without computer science training, standardized input tables were created for each question type to simplify data entry and automate database integration. Table 1 shows an example of the input format, which was parsed to implement the questions in the application. For multiple choice questions, for example, the order of the answers is not significant.

Feedback sessions were held in which project leaders, students, teachers, and the developer reviewed the questions to correct errors and discuss content. Each time a student completes a quiz, the system records their results, providing teachers with insights into overall class performance. Questions can also be tagged by difficulty level, which supports random presentation, targeted reinforcement of specific content, and progressive increases in challenge. Additionally, this data enables users to track their own progress within the application, displaying the total number of questions answered as well as correct and incorrect responses.

5.4. Communication features

Given the community-oriented, immersive, and intensive nature of the course, the platform includes communication features that allow participants to engage in public discussions and chat directly with other students or teachers. These functions were successfully implemented and tested by students, teachers, and project leaders. Communication is supported via email for account management and an in-app notification system that alerts users to new messages in chats or forums.

5.5. Automatic Speech Recognition

Although the first phase focused primarily on application testing, Automatic Speech Recognition (ASR) has been pursued as an enhancement, and its progress is reported here to update the project status. Using the CMU Mapudungun corpus (Duan et al., 2019, 2020)as a starting point, we fine-tuned pre-trained models. Given available resources, we used Whisper-small with LoRA (Song et al., 2024) (Dettmers et al., 2023) to establish a baseline ASR system for future comparisons, as it has been used in different languages, giving optimistic results (Radford et al., 2022) (Ntirenganya et al., 2024) (Daouad et al., 2024). The dataset was split into 27,696 training examples, 1,716 validation examples, and 1,619 test examples, with no overlap; longer audio files were segmented into clips of up to 30 seconds, which the model processes individually, we plan to improve the dataset adding more examples to vary domain and trying to include as many territorial variants as possible. Model performance was evaluated quantitatively and qualitatively by a Mapuche speaker born and raised in Alto Biobío, Biobío, Chile.

6. Evaluations and Results

There are two types of results in this study. The first type concerns application usage data, which allows assessment of overall user performance. This helps determine whether questions are too easy (very high correct answer rates) or too difficult (very low correct answer rates), which could indicate a mismatch between the application and students' course reality.

The second type of results relates to the integration of ASR. The ASR system is evaluated like a typical AI model using standard metrics—LOSS, Word Error Rate (WER), and Character Error Rate (CER)—complemented by qualitative validation from a Mapuzugun speaker.

6.1. Application results

With a group of 32 students, a total of 562 answers were collected; as participation was voluntary, not every question was answered by every student. The distribution and accuracy by question type were as follows:

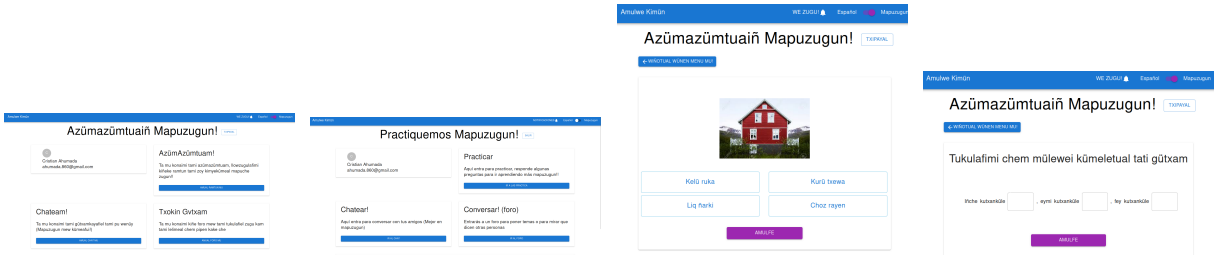
- Multiple choice (t1): 443 answers, 92.3% accuracy (33 questions)
- Sorting (t2): 105 answers, 55.2% accuracy (8 questions)
- Open answer (t3): 14 answers, 7.1% accuracy (2 questions)

Although the sample size is small due to the PoC nature of the study, it is representative of an actual Mapuzugun course where students were learning the language concurrently with the quiz. The results suggest that an analysis of question accuracy could inform a difficulty-based recommendation system, rather than presenting questions in random order. Future work could involve training a recommendation model with one group of students exposed to randomly ordered questions, and then evaluating its effectiveness with a new student group, using ordering by difficulty and category (Liu et al., 2024). Qualitative feedback from instructors and organizers indicated that the mixed classroom–application approach supported vocabulary growth, basic grammar practice, and listening/reading comprehension, while keeping cultural content central to both spaces.

6.2. ASR results

The first experiment was a zero-shot evaluation, showing that the pretrained Whisper-small model performs poorly on Mapuzugun, with a loss of 7.30%, as it was not fine-tuned for the language. The main experiment used 27,696 training examples, 1,716 validation examples, and 1,619 test examples, producing the results reported below.

It is important to note that two sets of WER and CER metrics were calculated. The first corresponds to the original text, while the second, "fixed" metrics, normalize spaces and word boundaries.



(a) UI in Mapuzugun (b) UI in Spanish (c) Multiple-choice example (d) Free-text example

Figure 1: Platform UI and question examples.

Question	Alternatives	English gloss / translation
chumten txipantu nieimi	– kümelekan – mai	*How old are you?*
chuchi kullin gei; (question.jpg)	– *alternative1.jpg* – *alternative2.jpg*	*What animal is it?*
question.mp3	– *alternative1.mp3* – *alternative2.mp3*	-
chuchi azentu mu müli ti mañke?	– *alternative1.jpg* – *alternative2.jpg*	*In which picture is there a condor?*
chuchi kullin gei? (question.jpg)	– mañke – txewa	*What animal is it?*
chumgechi pigekey luna mapuzugun mew?	– *alternative1.mp3* – *alternative2.mp3*	*How do you say "moon" in Mapuzugun?*
Yo tengo dos perros	– Inche nien epu txewa – Inche nien mari txewa	*I have two dogs*

Table 1: Examples of questions and alternatives (text, image, and audio modalities), with space for English glosses.

This adjustment accounts for the lack of a standardized orthography in Mapuzugun, where multiple representations of the same phrase can exist in Mapuzugun—for example, "Newengelai" versus "Newen gelai", both meaning "He/she/it does not have strength." Note that WER/CER can exceed 100% when insertions dominate under orthographic mismatch.

6.3. Human evaluation

The Mapuzugun speaker from Alto Biobío found very high agreement for some items and no recognizable match for others, indicating considerable item-level variability. Minimal preprocessing was applied to accommodate the scale and diversity of the dataset, so remaining issues—such as alignment inconsistencies, segmentation/orthography variation, and background noise—likely contributed to this variability. Future work will involve more thorough cleaning and normalization to improve robustness and reduce error inflation. Examples of prediction–manual transcription comparisons are shown in Table 5.

Setting	WER↓ (orig.)	WER↓ (no-space.)
Zero-shot (dev)	208.75%	99.88%
Zero-shot (test)	185.53%	100.00%
Fine-tuned (dev)	188.27%	97.96%
Fine-tuned (test)	168.32%	98.27%

Table 2: WER on Mapuzugun. "norm." removes spaces before scoring to handle non standard segmentation in Mapuzugun .

Setting	CER↓ (orig.)	CER↓ (no-space.)
Zero-shot (dev)	173.77%	161.38%
Zero-shot (test)	162.81%	152.07%
Fine-tuned (dev)	175.14%	163.36%
Fine-tuned (test)	163.80%	153.95%

Table 3: CER on Mapuzugun. "normalized" removes spaces before scoring.

6.4. Discussion

This work was originally designed to support learning in a course using multimodal data, but the application can also function as a standalone learning tool or be integrated into other courses or learning

Setting	Loss↓
Zero-shot (dev)	7.88
Zero-shot (test)	7.30
Fine-tuned (dev)	2.14
Fine-tuned (test)	1.89

Table 4: Loss on Mapuzugun dev and test splits.

contexts, thanks to its adaptability and culturally grounded content. Future improvements could include adding more multimodality with resources such as videos or an upgraded automatic speech recognition system, enabling more complex exercises that combine multiple data types and response formats.

A mixed classroom-application setup appears to translate into vocabulary gains and basic grammar practice (see qualitative observations), but open production likely requires more explicit guidance, or maybe a normalizer at the scoring stage. Reporting both strict exact-match and tolerance-aware metrics is pedagogically preferable: a non-binary scale surfaces *where* the error occurs (e.g., spacing, affix, or token choice) instead of labeling the whole answer as wrong. Concretely, we expose a strict score (correct /incorrect), a tolerant score based on character-level edit distance and no-space token matching, and (iii) a span-localized diff that highlights the erroneous segment in the student's answer.

Another potential enhancement is a system to present questions in an intelligent way, that is, dynamically select items based on prior student performance, reinforcing weak areas and gradually increasing difficulty to optimize learning results. This would personalize the learning experience and provide richer feedback for both students and teachers. Additionally, the application could evolve into a dedicated social platform for Mapuzugun learners, incorporating features such as video calls, user groups, and content posting to connect individuals interested in learning and teaching the language.

Any development like this must be proposed and discussed with Mapuche communities and people , respecting their data sovereignty and cultural priorities, even in the absence of explicit legal frameworks in Chile for community work and field work in digital projects.

7. Conclusion and perspectives

In this research, we presented an application designed and developed by, with, and for Mapuche people and communities to support the Mapuzugun language. The application helps connect learners with the same goals and enables multimodal practice of Mapuche language, contributing to revitalization and supporting the development of new speakers to preserve the culture.

For future work, we propose evaluations with Mapuche users like students and teachers, because those are necessary. Beyond educational quizzes and ASR, there is a need to explore additional NLP tools such as intelligent learning and educational assistants, machine translation, assisted translation, and text to speech. These are areas where Mapuzugun currently lacks robust educational and technological support. These tools could help engage younger generations with their language and culture through technology.

Finally, we found that a multidisciplinary workflow with community speakers, combined with a classroom–application setup, has strong potential going forward. The approach and tooling are not Mapuzugun-specific and can be replicated for other languages; each advance in one Indigenous language yields reusable methods and resources that can benefit others. The ultimate goal of this study is to contribute to the revitalization and preservation of the Mapuzugun language.

Ethics Statement

This project was co-designed by, with and for Mapuzugun speakers and course organizers, following community timelines and decision-making. Participation in the course platform was voluntary; no personally identifying information was collected in usage analytics, which are reported only in aggregate. The application is intended to supplement, not replace, live instruction by Mapuche teachers.

For the ASR experiments, we exclusively used the publicly released AVENUE/Mapudungun corpus (Available online Mapudungun Corpus (2020)). The data were collected under a human-subjects protocol, with written consent, and are anonymized; culturally proprietary knowledge disclosed by *machi* was removed at the time of transcription. The original release reports participants aged 16–100. We abide by the dataset's conditions, do not attempt re-identification, and do not redistribute raw audio. Our models are for research/educational use and are not intended for clinical or other high-stakes settings. We align our data-governance choices with Indigenous data-sovereignty principles (e.g., CARE), and any future release involving new data will follow community approval and an appropriate licence (Duan et al., 2019, 2020)

Reference	Hypothesis	Match
Good		
¿chumimi fey kat**rütual** chafo **notual ti?**	**¿chumimi fey k**ü**trütual** chawün **notual ti?**	Good
fey tami <uhm> notumun fey**chi kutranfel**	**fey tami** he notummu **chi kutran fel**	Good
feyta külonengü llemay **make**	**feyta küll**ünne koni **may make**	Good
külon make	**külon make**	Good
feyta pichike re**len inche tati** müleputun tañi kampumu **inche** anta **mülenolu** tañi **mapumu ta fanten ta dunguken tati**	**feyta pichike r**ü**len inche tati** mülefutukelu taiñ ka pomeu **inche** anta **mülenulu** taiñ **mapumu ta fanten**mu **ta dunguken tati**	Good
Medium		
rüf kümelkalekayngün **kom** weluengün	**kom welu** engün	Medium
ngelay dungu tati ngelay dungu chew felen ta faw . . .	**ngelay dungu tati ngelay dungu** chew felen ta pau rukelay **dungu** . . .	Medium
chafon fey famechi küpatun **welu**, wentuküley-mukan	feymu tachi küpatuken **welu**, **welu** tukulelngi ka	Medium
<%> **feyta** amukatuy ta chafo ka	**feyta** amuka tati chaw rume	Medium
<%> kimimi feychi **l'awen'**	ka kümewi veychi **l'awen'**	Medium
Bad		
petu ta umawküley kiñeke tañi pu che . . .	fey tüfi, fey, fey	Bad
femi tati femi	may	Bad
feleyal anchi feleyal anchi	feyti	Bad
<*SPA>si%	may	Bad
¿mapuche l'awen'?	feyti	Bad

Table 5: Fifteen reference–hypothesis pairs grouped by qualitative match. Exact word overlaps are highlighted in **bold**.

Limitations

Our study is a small educational pilot (32 learners) with limited duration, so generalization beyond this cohort is uncertain. The current ASR baseline demonstrates feasibility rather than production readiness; it was not evaluated in-situ with students and only underwent a limited native-speaker qualitative check. The training data are health-domain conversations from legacy recordings, which introduces domain and register bias; dialectal coverage (e.g., Lafkenche, Pewenche) remains uneven, and acoustic conditions include background noise and segmentation artifacts. Non-standard orthography and spacing inflate WER/CER; our "no-space" variants partially address this but remain imperfect proxies of communicative adequacy. Finally, the project is currently ongoing, however there is a limit to iteration, deployment and full accessibility testing.

Acknowledgements

The authors are grateful to the Mapuche community members for initiating this project, sharing their knowledge and experience, and engaging in meaningful collaboration. We also thank our partners and collaborators at the Faculty of Social Sciences (FACSO) of the University of Chile and the National Corporation for Indigenous Development (CONADI), as well as the anonymous reviewers for their thoughtful and valuable feedback. This project was undertaken thanks to funding from IVADO (R^3AI) and the Canada First Research Excellence Fund.

8. Bibliographical References

C. Ahumada, C. Gutierrez, and A. Anastasopoulos. 2022. Educational tools for mapuzugun. In *Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022)*, pages 183–196, Seattle, Washington. Association for Computational Linguistics.

M. Alvarado. 2012. Sistema para el aprendizaje del mapudungun, incluyendo características de reconocimiento de voz y bot conversacional. Master's thesis, Pontificia Universidad Católica de Valparaíso.

A. Alvarez, R. Aranovic, R. Brown, J. Carbonell, C. Fasola, A. Lavie, L. Levin, A. Font-

Llitjos, C. Monson, E. Peterson, K. Probst, and R.M. Vega. 2005. Informe final proyecto avenue/mapudungún: Desarrollo de herramientas informáticas para el mapudungún que se habla en chile. Technical report, Instituto de Tecnologías del Lenguaje, Universidad Carnegie Mellon.

Antti Arppe, Jordan Lachler, Trond Trosterud, Lene Antonsen, and Sjur N. Moshagen. 2016. Basic language resource kits for endangered languages: A case study of plains cree. In *CCURL 2016—Collaboration and Computing for Under-Resourced Languages: Towards an Alliance for Digital Language Diversity (LREC 2016 Workshop)*, Portorož, Slovenia. European Language Resources Association (ELRA).

Luis Camacho and Rodrigo Zevallos. 2019. Siminchikkunarayku. lingüística computacional para la revitalización y el poliglotismo: Hoja de ruta. Technical report, Fundación Siminchikkunarayku and Pontificia Universidad Católica del Perú, Lima, Perú. Roadmap document.

S. Castillo, S. Mayo, and J. Soto. 2022. Production of educational materials for mapuzugun teaching: An educators' perspective and experience approach. In *Addressing linguistic inequality in Latin America*. Unknown.

CENIA. 2023. Chile a la vanguardia de la ia: Los avances que nos deja el centro nacional de inteligencia artificial este 2023. Accessed: 2025-05-08.

A. Chandía. 2022. A mapudüngun fst morphological analyser and its web interface. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 6540–6547, Marseille, France. European Language Resources Association.

Pascual Coña. 2019. *Kuyfi mapuche chumgechi ñi azmogekeel egün. Genlol*.

Mohamed Daouad, Fadoua Ataa Allah, and El Wardani Dadi. 2024. Optimizing whisper models for amazigh asr: a comparative analysis. *International Journal of Speech Technology*, 28(1):27–37.

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2023. Qlora: Efficient finetuning of quantized llms. Relevant for memory-/compute-efficient parameter-efficient finetuning.

M. Duan, C. Fasola, S. Rallabandi, R. Vega, A. Anastasopoulos, L. Levin, and A. Black. 2020. A resource for computational experiments on mapudungun. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 2872–2877, Marseille, France. European Language Resources Association.

M. Duan, C. Fasola, S.K. Rallabandi, R. M. Vega, A. Anastasopoulos, L. Levin, and A.W. Black. 2019. A resource for computational experiments on mapudungun.

Mingjun Duan and contributors. 2019. Mapudungun speech/text corpus (github repository). https://github.com/mingjund/mapudungun-corpus. Accessed 2025-10-15.

M. Inostroza, E. Quero, and A. Berrios. 2021. Profile of participants in mapuche language immersion programs organised by mapuzuguletuaiñ. *Nueva Revista del Pacífico*, (74).

Eduardo Iturrieta, Cristian Lagos, and Elizabeth Ávila. 2024. Los grafemarios de la lengua mapuche como herramienta de revitalización lingüística: una revisión bibliográfica.

Yuxuan Liu, Haipeng Liu, and Ting Long. 2024. Hierllm: Hierarchical large language model for question recommendation.

E. Loncon, B. Villena, and S. Fernández-Silva. 2023. Chumleafel chi anültuzugu chile mapu mew: The role of mapuzugun in the chilean constituent process (2019–2022). *Revista Signos*, 56(113):582–609.

Manuel Mager, Ximena Gutiérrez-Vásquez, Gerardo Sierra, and Iván Meza. 2018. Challenges of language technologies for the indigenous languages of the americas. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018)*, Santa Fe, New Mexico, USA. Association for Computational Linguistics. Resource link associated with the paper.

F. Maldonado, A. Arjona, and D. Johnson. 2024. Revitalización lingüística: Diseño, implementación y evaluación de un programa de intervención para jóvenes mapuche. *Interciencia: Revista de ciencia y tecnología de América*, 49(8):471–480.

A. Mariano. 2021. Mapuche language revitalization through the mapuzuguletuaiñ wallmapu mew project. *Documentos Lingüísticos y Literarios*, (40).

Ministry of Social Development and Family. 2024. *Informe final. Estado de la lengua mapuzugun*. Accessed: 2024-04.

Benjamin Molineaux. 2023. The corpus of historical mapudungun: morpho-phonological parsing and the history of a native american language. *Corpora*, 18(2).

Jean de Dieu Ntirenganya, Didier Habineza, Yvan Niyigena, et al. 2024. Kinyawhisper: Enhancing asr for kinyarwanda using whisper models. *arXiv preprint arXiv:2405.08072*. Demonstrates fine-tuning of Whisper models for Kinyarwanda ASR as a low-resource case.

M. Pell. 2023. La región que traza el proceso de revitalización del mapuzugun. In C. Hammerschmidt, L. Anapios, C. Tomadoni, F. Oliveira de Souza, and S. Espul, editors, *América Latina en discusión: una apuesta por las metodologías horizontales*, pages 136–146. Editorial Universidad de Guadalajara, México.

N. Quiero. 2025. *Estudio confirma que la continuidad de la lengua mapuche está en riesgo*. Accessed: 2025-01-02.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2022. Robust speech recognition via large-scale weak supervision.

M. Rivas and M. Del Pino. 2023. Enseñanza del mapuzugun en chile: ecosonoridad como apresto a la adquisición de la lengua indígena. *Educación*, 32(62):185.

Scott Sadowsky, Héctor Painequeo, Gastón Salamanca, and Heriberto Avelino. 2013. Mapudungun. *Journal of the International Phonetic Association*, 43(1):87–96.

V. Sarmiento. 2023. Traducción español-mapuzugun en la universidad de playa ancha: un proyecto de innovación educativa con impacto social, cultural y político. *Revista digital de políticas lingüísticas*, (19).

I. Smeets. 1989. A mapuche grammar. *Leiden: Rijksuniversiteit te Leiden*.

Zheshu Song, Jianheng Zhuo, Yifan Yang, Ziyang Ma, Shixiong Zhang, and Xie Chen. 2024. Lora-whisper: Parameter-efficient and extensible multilingual asr.

CEDETI UC. Mapudungun mew. `http://www.cedeti.cl/tecnologias-inclusivas/software-educativo/mapudungun-mew`. Accessed: 2025-05-08.

UNESCO. 2010. *Atlas of the World's Languages in Danger*, 3rd edition. UNESCO, Paris.

F. Zúñiga. 2006. *Mapudungun. El habla mapuche*. Centro de Estudios Públicos, Santiago de Chile.

9. Language Resource References

Mapudungun Corpus. 2020. *Mapudungun Corpus*. LREC 2020. Freely available speech/written corpus. PID https://lremap.elra.info/search_by_map_responsive.php?query=Mapudungun+Corpus.