Let's Talk, Not Type: An Oral-First Multi-Agent Architecture for Guaraní

Anonymous authors
Paper under double-blind review

Abstract

Although artificial intelligence (AI) and Human-Computer Interaction (HCI) systems are often presented as universal solutions, their design remains predominantly text-first, underserving primarily oral languages and indigenous communities. This position paper uses Guaraní, an official and widely spoken language of Paraguay, as a case study to argue that language support in AI remains insufficient unless it aligns with lived oral practices. We propose an alternative to the standard "text-to-speech" pipeline, proposing instead an oral-first multi-agent architecture. By decoupling Guaraní natural language understanding from dedicated agents for conversation state and community-led governance, we demonstrate a technical framework that respects indigenous data sovereignty and diglossia. Our work moves beyond mere recognition to focus on turn-taking, repair, and shared context as the primary locus of interaction. We conclude that for AI to be truly culturally grounded, it must shift from adapting oral languages to text-centric systems to treating spoken conversation as a first-class design requirement, ensuring digital ecosystems empower rather than overlook diverse linguistic practices.

Conversational AI, Oral-first interaction, Low-resource languages, Diglossia, Guaraní, Culturally grounded AI, Multi-Agent Systems, Indigenous data governance

1 Introduction

Most AI systems and everyday interaction with machines remain oriented around text input: keyboards, menus, and form-like interfaces. Voice features are increasingly common, but in many deployments they function as a spoken front-end to a text-first pipeline (transcribe, parse, respond), rather than as conversation with turn-taking, clarification, and repair. Commercial assistants such as Amazon Alexa illustrate the dominant interaction style: wake word, short request, single response. This style does not match the interactional structure of human dialogue, and spoken systems frequently exhibit problems with interruptions, response timing, and turn coordination Skantze (2021). When misunderstandings occur, breakdown handling often depends on limited repair behaviors and places a substantial burden on the user to rephrase, repeat, or restart Alghamdi et al. (2024).

Conversation is a coordinated action. People manage understanding through grounding and shared context (common ground), using clarification and repair to maintain alignment Clark and Brennan (1991). Turn-taking is a core mechanism in this coordination, with strong cross-cultural regularities and measurable variation in timing across languages Stivers et al. (2009). These properties matter for interface design: if a system cannot manage turn-taking and repair, voice interaction tends to collapse into brittle command-and-control behavior, even when speech recognition and synthesis are available Skantze (2021); Alghamdi et al. (2024).

These limitations disproportionately affect low-resource and endangered languages. As Turin writes, "Up to half of the world's 6,500 languages spoken today may be extinct by the end of this century" Turin (2012). Many of these languages have limited written literature and are primarily oral in everyday use, yet language technology resources are often sparse

across speech and text modalities Turin (2012). Prior work has explored speech interfaces for information access in low-literacy settings Plauché et al. (2006), and reviews on interface design for illiterate and semi-literate users reinforce that interaction should align with real communication practices rather than assume literacy Islam et al. (2023). However, much of this line of work targets access barriers and usability for specific tasks; it does not fully address oral-first conversational interaction as a design goal, especially for low-resource languages where both speech and dialogue infrastructure are limited.

Rather than adapting oral languages to text-centric systems, oral interaction should be treated as a first-class design starting point. Orality-grounded HCI for the developing world (HCID) argues that design assumptions imported from literate settings can fail in oral cultures, where knowledge organization and recall often rely on narrative structure, repetition, and socially distributed memory rather than list-like externalization Sherwani et al. (2009). Recent projects suggest why this matters and what is feasible. For instance, Abena AI is presented as a Twi-capable voice assistant aimed at supporting local language interaction beyond English-dominant assistants Schwartz (2022). In Iceland, the Embla assistant was launched as an Icelandic-speaking voice interface, motivated by national efforts to integrate Icelandic into digital services **?**. In Australia, a partnership between Google and the University of Western Australia has focused on building a high-quality Aboriginal English speech dataset to reduce exclusion caused by poor recognition of First Nations voices and speech varieties Hutchinson (2025); The University of Western Australia (2025). These efforts demonstrate momentum toward speech-based systems for underrepresented languages and speech communities, but they also highlight a broader gap: many systems still inherit interaction assumptions from text-first design, and low-resource settings often lack conversational infrastructure (dialogue state tracking, repair strategies, safety and consent mechanisms) alongside the underlying models.

We use Guaraní as a motivating case for oral-first conversational AI. Guaraní is one of Paraguay's two official languages alongside Spanish Organization of American States (OAS) (1992). It is actively used in daily life: according to Paraguay's Instituto Nacional de Estadística (INE), language use in the home among people age 5+ includes speakers who communicate mostly in Guaraní (30%), in Spanish (28.5%), or in both Guaraní and Spanish (38.7%) Instituto Nacional de Estadística (INE), Paraguay (2024). Guaraní has long been sustained through oral transmission, but written Guaraní has existed for centuries through early grammars and dictionaries produced in the colonial period (e.g., Ruiz de Montoya's works), with later national standardization efforts and orthographic governance Ruiz de Montoya (1639); Secretaría de Políticas Lingüísticas (Paraguay) (2010). Though widely spoken, it remains under-represented in the digital realm, which can push interaction toward Spanish even for bilingual speakers. This pressure reflects an institutional asymmetry: Spanish dominates many formal domains (official documents, schooling, administrative procedures), and the legal text itself establishes Spanish as the default for resolving interpretive doubt Organization of American States (OAS) (1992). For this setting, the challenge is not only recognition or translation. It is building interaction that can sustain multi-turn intent, resolve shared context across turns, and support clarification and repair in the language users actually speak.

To address this, we propose an oral-first assistant architecture that separates language understanding, conversation state, action execution, and governance into interacting agents. We argue for a Multi-Agent System (MAS) framework for low-resource conversational systems because it enables specialization (e.g., a Guaraní speech/intent agent) while keeping interfaces between components explicit and testable. Recent work on conversational task-solving demonstrates that decomposing complex interactions into specialized agents significantly improves task success compared to monolithic Large Language Models (LLMs) Becker (2024). Furthermore, surveys on multi-party agents highlight that maintaining a "participant's state of mind," essential for natural turn-taking and repair, requires dedicated state tracking distinct from generation or execution Chen et al. (2025).

In our proposal, we decompose oral-first interaction into six cooperating agents: (1) a Guaraní speech interface agent responsible for audio capture and turn-holding; (2) a Guaraní understanding agent that maps spoken input to abstract intents; (3) a conversation state and grounding agent that maintains shared context across turns and supports clarification and

repair; (4) domain-specific action agents that execute tasks such as media control, browsing, and file operations; (5) a response agent that generates conversational audio output based on dialogue context and action outcomes; and (6) a permission, privacy, and data governance agent that mediates consent, confirmations, and data retention.

This paper is structured as follows: Section 2 introduces Guaraní as a case and motivates oral-first requirements through the lens of language use and institutional diglossia in Paraguay. Section 3 presents the proposed multi-agent architecture and its governance mechanisms. Section 4 discusses limitations and open challenges, and Section 5 concludes.

2 Case Study: Guaraní in Paraguay

Diglossia as a design constraint. A key challenge for interface design in Paraguay is *diglossia*: a relatively stable sociolinguistic arrangement where two varieties (or languages) are functionally separated across domains, with a prestige hierarchy between a "High" (H) code used in formal, written, and institutional contexts and a "Low" (L) code used in everyday speech and informal interaction Ferguson (1959); Fishman (1967). In Paraguay, Guaraní is widely used in oral, intimate, and community domains, while Spanish is strongly associated with literacy, bureaucracy, and many public-facing written systems Ito (2012).

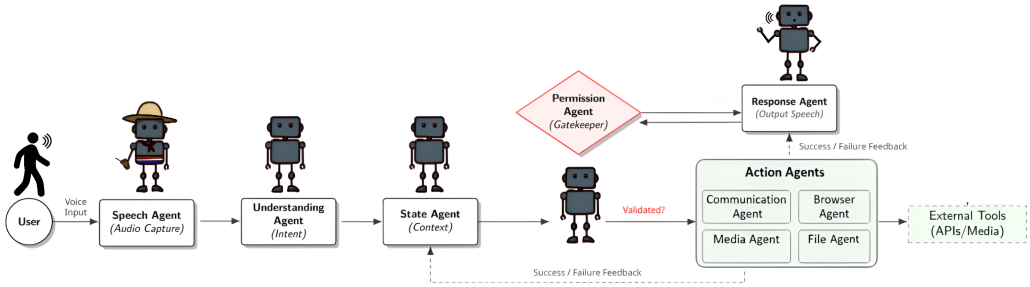
This division matters because most contemporary digital interaction is still mediated through *written* interfaces: menus, forms, onboarding flows, error messages, and search-oriented navigation. Even when a system offers "language support," the practical experience often remains Spanish-default at the points where users must read, type, verify, or resolve errors. Diglossia therefore creates a predictable adoption pattern: users may prefer Guaraní for everyday communication but accept Spanish as the "language of getting things done" in institutional and digital environments. In HCI terms, this is not simply a translation gap; it is a *domain mismatch* between the language of lived interaction and the language encoded into interface infrastructure Ferguson (1959); Ito (2012).

Literacy gap and interaction burden. Because Spanish dominates many formal domains (official documents, schooling, administrative procedures), written interaction frequently pulls users toward Spanish even in otherwise bilingual households Ito (2012). For bilingual speakers, this can impose an extra cognitive and social burden: code-switching is not just linguistic, but also a shift into a domain with different norms of correctness, authority, and accountability. As a result, text-based interfaces and text-first voice systems ("say it, then read/confirm it") can reinforce the H/L split rather than reduce it, especially when breakdown handling requires repeated re-phrasing, spelling, or form-filling.

Digital access vs. language choice. Internet access in Paraguay has expanded rapidly, with the share of the population aged 10+ using the internet rising from 61.1% (2017) to 81.6% (2024) Instituto Nacional de Estadística (INE), Paraguay (2025). However, increased connectivity does not guarantee increased Guaraní presence in digital services. Instead, diglossia can push bilingual users into Spanish because mainstream tools, defaults, and "official" channels are optimized for Spanish literacy rather than Guaraní conversational use Ito (2012). This amplifies a central motivation for oral-first design: if interaction remains text-centric at the moments that matter (setup, navigation, error recovery), then "supporting Guaraní" risks becoming symbolic rather than functional.

3 Proposed Architecture: An Oral-First Multi-Agent System

We propose an *oral-first* architecture that treats speech as the primary interaction modality, prioritizing multi-turn context, repair mechanisms, and community-led governance. Our Multi-Agent System orchestrates six specialized agents, each addressing a discrete aspect of oral language processing to transform a Guarani speaker's utterance into meaningful action. One can picture a conversation flowing through a series of expert interpreters, each contributing their specialized knowledge: one listens for when the speaker has finished talking, another interprets the cultural meaning behind the words, a third remembers what



System diagram showing Guaraní voice input routed through speech capture, intent understanding, conversation state tracking, and tool execution. A permission and governance agent mediates actions before execution.

Figure 1: Proposed Oral-First Multi-Agent Architecture. The Speech Agent (far left) is depicted in traditional Paraguayan attire to symbolize the system's grounding in local cultural identity.

was discussed a few moments ago, a fourth ensures the speaker's privacy is protected, and finally, specialist agents execute the requested tasks:

- **The Listener - Speech Interface Agent:** Every conversation begins with listening, knowing when to pay attention and when the speaker has finished their thought. The Speech Interface Agent serves as this conversational gatekeeper, employing Voice Activity Detection (VAD) to capture audio while using pause duration and timing cues to distinguish between brief pauses and actual turn completion. This design choice is particularly critical for Guaraní, where brief pauses or glottal stops (pedagogically described as *puso*) can occur within words and do not signal the speaker has finished Estigarribia (2020). Rather than processing meaning, this agent focuses solely on supporting floor-holding during silence, respecting the spoken turn-taking behaviors documented in conversational analysis research Skantze (2021); Stivers et al. (2009). By honoring Guaraní's natural oral pacing, the agent prevents premature interruption that would fragment the speaker's intended utterance.
- **The Cultural Interpreter - Guaraní Understanding Agent (NLU):** With the audio captured, the Understanding Agent interprets what the speaker's intentions are, mapping Guarani and Jopará speech to abstract intents like `PLAY_MUSIC`, `OPEN_TAB`. Rather than relying on synthetic translations, this agent requires training on authentic community-verified speech to recognize specific cultural references and loanwords (see section 3.2). This grounding ensures the agent recognizes lived language, complete with regional variations, borrowed terms, and the fluid code-switching patterns that define everyday jopará.
- **The Memory Keeper - Conversation State Agent:** Natural conversations build on shared history, and the Conversation State Agent maintains this dialogue memory. For example, if the user follows up with 'Nda che gustái' (No, I don't like [this]), this agent resolves the implicit reference understanding that "this" refers to the current one. Beyond anaphora resolution, it tracks conversation flow, what actions have been completed, what remains pending, and what context remains relevant for interpreting subsequent utterances. This persistent memory transforms isolated commands into coherent dialogue, enabling the fluid multi-turn exchanges that characterize natural Guaraní conversation.
- **The Guardian - Permission & Governance Agent (The "Gatekeeper"):** A sovereign agent separate from the execution layer. It checks every requested action against community-defined privacy norms (e.g., "Do not store this audio

clip") and user consent settings before allowing the Action Agents to proceed. This agent embodies data sovereignty, placing control over voice recordings, intimate captures of speakers' voices, homes, and conversations directly in the hands of Guaraní speakers themselves. By operating as a separate governance layer rather than an embedded permission check, it ensures that privacy enforcement cannot be bypassed or overridden by other system components.

- **The Conversationalist - Response Agent:** An agent that generates conversational responses grounded in the current dialogue state and action outcomes, including confirmations, denials, and repair prompts. All responses are checked by the Permission & Governance Agent before being delivered to the user. By treating system responses as conversation contributions rather than status messages, this agent sustains the oral-first interaction paradigm throughout the entire exchange.
- **The Specialists - Action Agents:** These are small, modular agents responsible for executing tasks in specific domains, such as a *Media Agent* (e.g., Spotify/YouTube) or a *Browser Agent*, with additional domain-specific agents introduced as needed. These small, focused agents operate independently within their narrow expertise, with additional domain-specific agents introduced as system capabilities expand. This modularity allows organic growth, adding weather information, smart home control, or calendar management without redesigning the core conversational architecture.

3.1 Training Data and Community Alignment

The question then comes: where does training data for the Guaraní Understanding Agent come from, and how do we ensure it reflects authentic, community-verified speech rather than synthetic translations? This is primarily because community data captures how Guaraní speakers actually communicate, complete with code-switching, regional variations, and the pragmatic nuances that define oral tradition. In this regard, as a design reference point, we identify two open initiatives that illustrate the types of data required for such a system: (a) **Mozilla Common Voice (Guaraní):** The Mozilla Common Voice Guaraní (`gn`) dataset, a massive crowdsourced initiative for open speech data Ardila et al. (2020), provides an example of the kind of open, multilingual speech resource that could support acoustic modeling in an oral-first system. (b) **The Aikuaa Initiative:** To ensure cultural relevance, a system of this kind would benefit from community-led sources such as *Aikuaa*, a project led by *El Surtidor* that organizes community "mingas" (collaborative gatherings) to collect and validate Guaraní voice data JournalismAI (2025). Initiatives like Aikuaa illustrate how conversational data can capture *Jopará* (mixed Guaraní–Spanish) usage and pragmatic variation that are often absent from more formal or archival datasets.

Table 1 illustrates how the agents would collaborate to handle context and repair in a natural Guaraní dialogue:

But how do we measure whether this adaptation actually succeeds? How do we validate that our oral-first approach delivers on its promise to Guaraní speakers? This is discussed below:

3.2 Evaluation Criteria

Validating an oral-first architecture requires moving beyond traditional accuracy metrics. A system could achieve 95% word error rate on isolated commands yet completely fail at natural conversation, such as mishandling context, ignoring sentiment, or violating privacy norms. Our evaluation framework addresses four dimensions that reflect conversational success rather than mere technical performance. (i) **Task Success Rate (TSR):** measures the percentage of multi-turn goals, not just single commands, completed successfully. This metric captures whether the Conversation State Agent successfully maintains dialogue coherence and whether the Understanding Agent correctly interprets evolving intents across turns. (ii) **Repair Success Rate:** captures conversational resilience: how often the system recovers from failures (e.g., misheard words, misunderstood intents) without forcing the user to abandon their task. Oral communication is naturally disfluent; speakers pause,

User (Guaraní)	System Action & Agent Logic
"Che ahenduse purahei" (I want to listen to music)	1. **Speech Agent** captures audio. 2. **Understanding Agent** identifies intent: `PLAY_MUSIC`. 3. **Permission Agent** verifies "Music" is a safe category; approves. 4. **Media Agent** selects a popular playlist and starts playback.
Response Agent	*"Oĩ porã"* (Ok / It is good)
"Nda che gustái" (No, I don't like [this])	1. **Understanding Agent** detects negative sentiment/rejection. 2. **Conversation State Agent** resolves the implicit object: "this" = *current song*. Updates intent to `SKIP`. 3. **Media Agent** executes `NEXT_TRACK`.

Table 1: Multi-turn interaction demonstrating context resolution and repair.

restart, and code-switch mid-sentence. A robust oral-first system must handle these characteristics gracefully, using the Response Agent to clarify ambiguities and the Conversation State Agent to maintain context through repair sequences. (iii) **Perceived Sovereignty:** assesses whether users trust that their voice data remains under their control, a qualitative metric measuring confidence that audio is not being extracted without consent. This dimension directly evaluates the Permission Agent's effectiveness, but success is measured not in latency or throughput but in community trust. Do Guaraní speakers feel safe using the system? Do they believe their data sovereignty is respected? These questions require ethnographic evaluation methods that center community voices. (iv) **Latency:** ensures the system responds at Guaraní's conversational tempo, neither interrupting too quickly (violating the floor-holding principles encoded in the Speech Interface Agent) nor leaving awkward silences that break conversational flow. This metric recognizes that speed alone is insufficient—timing must match cultural expectations for turn-taking in Guaraní dialogue.

4 Discussions and Limitations

The architecture we've proposed represents a vision for oral-first language technology. Yet translating this vision into reality confronts three fundamental tensions that extend beyond technical implementation. Understanding these tensions is essential not just for building new technologies, but for ensuring that language technology serves rather than displaces the communities it claims to support.

4.1 Standardization vs. Lived Reality

A recurring challenge for Guaraní language technology is the mismatch between institutional standardization and everyday speech. National language policy in Paraguay explicitly treats Guaraní as an official language and establishes state structures for language planning and normative work, including the Academia de la Lengua Guaraní under the Ley de Lenguas (Léi 4251/10) Secretaría de Políticas Lingüísticas (Paraguay) (2010); Secretaría de Políticas Lingüísticas. In practice, however, Guaraní is routinely used in contact with Spanish through *jopará* and other forms of code-switching, especially in informal interaction and emerging digital contexts Mortimer (2006); Estigarribia (2015); Kellert and Tyagi (2025). This matters for interface design because the "right" form is often not the form users actually speak. Even when terminological resources exist or are actively developed for technical domains, uptake is uneven, and speakers may prefer loanwords, paraphrases, or mixed forms that preserve fluency and social meaning CONACYT Paraguay; Secretaría de Políticas Lingüísticas. An oral-first assistant should therefore treat variation as expected input rather than as er-

ror, prioritizing communicative intent and conversational continuity over enforcing a single "correct" register Mortimer (2006); Kellert and Tyagi (2025).

4.2 The Data Bottleneck is Specifically Conversational

Data scarcity is an obvious constraint, but in oral-first systems the core gap is not just "more data". It is the lack of *conversational* audio that captures turn boundaries, interruption behavior, repairs, and shared context. As Turin notes, many endangered languages are principally oral, often with limited written literature, and are at risk of disappearing without documentation Turin (2012). For Guaraní specifically, recent NLP work continues to describe the language as low-resource, and notes that Guaraní web content is scarce even on Paraguayan websites where Spanish dominates Chiruzzo et al. (2022); Góngora et al. (2021). Efforts such as parallel corpora for MT benchmarking and scraping-based corpus building are important, but they still primarily support text-based evaluation and do not resolve the lack of spontaneous, multi-turn spoken interaction data Chiruzzo et al. (2022); Góngora et al. (2021). Community-led collection and organization efforts are therefore particularly valuable, but future datasets should explicitly target multi-turn spoken interaction, not only read speech or isolated commands JournalismAI (2025).

4.3 Governance and Perceived Control in Oral Data Collection and Deployment

Oral interfaces also raise governance issues that are intensified in low-resource settings: audio is identifiable, culturally embedded, and easily repurposed. Indigenous data governance frameworks emphasize community benefit, authority to control, and accountability, which are directly relevant to speech data collection, storage, and downstream model use Carroll et al. (2020). In practical system terms, this supports separating action execution from a dedicated permission and privacy layer that mediates consent, confirmations, and retention rules, including explicit "do not store audio" defaults. This separation is not only an ethical commitment; it is a design strategy that can increase perceived control and trust in multi-turn interaction Carroll et al. (2020); Alghamdi et al. (2024).

5 Conclusion

This work contributes to broader efforts in culturally grounded AI by treating conversation itself as a design principle. When users cannot interact with systems in ways that align with how their language is lived and practiced, language support risks remaining symbolic rather than meaningful. Our proposed multi-agent architecture for Guaraní moves beyond "universal" design assumptions by treating spoken conversation, context, and repair as central requirements. By decoupling the *Understanding Agent* from a dedicated *Permission and Governance Agent*, we provide a technical blueprint for AI that respects Indigenous data sovereignty and community-led norms. This design ensures that the system is not merely "deployed for" Guaraní speakers, but is structured to align with their specific sociolinguistic reality of diglossia and oral tradition. Ultimately, this work argues that for AI to be truly equitable, it must move beyond text-first assumptions and adapt interaction models to the communicative practices of the communities it serves. In the context of Paraguay, this approach bridges the gap between high digital connectivity and low linguistic representation, ensuring that AI strengthens cultural identity and empowers communities to interact with technology on their own terms.

References

Essam Alghamdi, Martin Halvey, and Emma Nicol. System and user strategies to repair conversational breakdowns of spoken dialogue systems: A scoping review. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces*, CUI '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400705113. doi: 10.1145/3640794.3665558. URL https://doi.org/10.1145/3640794.3665558.

Rosana Ardila, Megan Branson, Kelly Davis, Michael Kohler, Josh Meyer, Michael Henretty, Reuben Morais, Lindsay Saunders, Francis Tyers, and Gregor Weber. Common voice: A massively-multilingual speech corpus. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4218–4222, Marseille, France, May 2020. European Language Resources Association. ISBN 979-10-95546-34-4. URL https://aclanthology.org/2020.lrec-1.520/.

Evan Becker. Multi-agent large language models for conversational task-solving. arXiv preprint arXiv:2410.22932v1, 2024. URL https://arxiv.org/abs/2410.22932.

Stephanie Russo Carroll et al. The care principles for indigenous data governance. *Data Science Journal*, 2020.

Siyuan Chen et al. Multi-party conversational agents: A survey. arXiv preprint arXiv:2505.18845v1, 2025. URL https://arxiv.org/abs/2505.18845.

Luis Chiruzzo, Santiago Góngora, Aldo Alvarez, Gustavo Giménez-Lugo, Marvin Agüero-Torales, and Yliana Rodríguez. Jojajovai: A parallel Guarani-Spanish corpus for MT benchmarking. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 2098–2107, Marseille, France, June 2022. European Language Resources Association. URL https://aclanthology.org/2022.lrec-1.226/.

Herbert H. Clark and Susan E. Brennan. Grounding in communication. In *Perspectives on Socially Shared Cognition*. American Psychological Association, 1991. URL https://www.cs.cmu.edu/~illah/CLASSDOCS/Clark91.pdf.

CONACYT Paraguay. Terminología guaraní en dominios técnicos (material institucional). Use the PDF you linked as the canonical reference.

Bruno Estigarribia. Jopará and guaraní in paraguay (discussion of contact and mixed speech), 2015. Cited in Guaraní NLP work as evidence for contact-driven variation.

Bruno Estigarribia. *A Grammar of Paraguayan Guaraní*. UCL Press, 2020. URL https://uclpress.co.uk/book/a-grammar-of-paraguayan-guarani/.

Charles A. Ferguson. Diglossia. *Word*, 15(2):325–340, 1959. doi: 10.1080/00437956.1959.11659702. URL https://www.tandfonline.com/doi/pdf/10.1080/00437956.1959.11659702.

Joshua A. Fishman. Bilingualism with and without diglossia; diglossia with and without bilingualism. *Journal of Social Issues*, 23(2):29–38, 1967. doi: 10.1111/j.1540-4560.1967.tb00573.x. URL https://spssi.onlinelibrary.wiley.com/doi/10.1111/j.1540-4560.1967.tb00573.x.

Santiago Góngora, Nicolás Giossa, and Luis Chiruzzo. Experiments on a Guarani corpus of news and social media. In Manuel Mager, Arturo Oncevay, Annette Rios, Ivan Vladimir Meza Ruiz, Alexis Palmer, Graham Neubig, and Katharina Kann, editors, *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*, pages 153–158, Online, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.americasnlp-1.16. URL https://aclanthology.org/2021.americasnlp-1.16/.

Ben Hutchinson. A partnership with the university of western australia to improve speech technology for aboriginal and torres strait islander people's voices. Google Australia Blog, 02 2025. URL https://blog.google/intl/en-au/company-news/technology/a-partnership-to-improve-speech-technology-for-first-nations-voices/.

Instituto Nacional de Estadística (INE), Paraguay. Día internacional de la lengua materna: Diversidad lingüística en paraguay, 2024. URL https://www.ine.gov.py/noticias/2298/dia-internacional-de-la-lengua-materna-diversidad-linguistica-en-paraguay. Household language-use reporting based on EPHC 2023. Accessed 2026-02-02.

Instituto Nacional de Estadística (INE), Paraguay. 8 de cada 10 personas utiliza internet en paraguay (eph 2017–2024), 2025. Reports 81.6% internet use among population aged 10+ in 2024, up from 61.1% in 2017; includes noted exclusions. Accessed 2026-02-02.

Muhammad Nazrul Islam, Nafiz Imtiaz Khan, Toki Tahmid Inan, and Iqbal H. Sarker. Designing user interfaces for illiterate and semi-literate users: A systematic review and future research agenda. *SAGE Open*, 2023. doi: 10.1177/21582440231172741. URL https://journals.sagepub.com/doi/10.1177/21582440231172741.

Hiroshi Ito. With spanish, guaraní lives: a sociolinguistic analysis of bilingual education in paraguay. *Multilingual Education*, 2(1):6, 2012. doi: 10.1186/2191-5059-2-6. URL https://link.springer.com/article/10.1186/2191-5059-2-6.

JournalismAI. Guarani ai: When building language tech means building community, 2025. URL https://www.journalismai.info/blog/5fcm6ayykhqq7564kbvt9nw92wwmy9. Documents community "mingas" and Guaraní audio dataset efforts. Accessed 2026-02-02.

Olga Kellert and Nemika Tyagi. Where and how do languages mix? a study of spanish-guaraní code-switching in paraguay. In *Proceedings of the Workshop on Computational Approaches to Linguistic Code-Switching*. Association for Computational Linguistics, 2025.

Katherine Mortimer. Guaraní académico or jopará? educator perspectives and ideologies on language in paraguay, 2006. Often cited for discussions of standardization, literacy, and ideologies around Guaraní/Spanish mixing.

Organization of American States (OAS). Paraguay's constitution of 1992 with amendments through 2011. PDF, 1992. URL https://www.oas.org/ext/Portals/33/Files/Member-States/Parag_intro_textfun_eng_1.pdf. English translation; consolidated text with amendments through 2011.

Madelaine Plauché, Udhyakumar Nallasamy, Joyojeet Pal, Chuck Wooters, and Divya Ramachandran. Speech recognition for illiterate access to information and technology. In *Proceedings of the International Conference on Information and Communication Technologies and Development (ICTD)*, pages 83–92. IEEE, 2006. doi: 10.1109/ICTD.2006.301842.

Antonio Ruiz de Montoya. *Tesoro de la lengua guarani*. Juan Sanchez, Madrid, 1639. URL https://archive.org/details/tesorodelalengua00ruiz. Facsimile digitized by the John Carter Brown Library via Internet Archive.

Eric Hal Schwartz. Abena ai's voice assistants speak african languages untouched by google assistant and alexa. Voicebot.ai, 06 2022. URL https://voicebot.ai/2022/06/30/abena-ais-voice-assistants-speak-african-languages-untouched-by-google-assistant-and-alexa/.

Secretaría de Políticas Lingüísticas. Academia de la lengua guaraní. https://spl.gov.py/es/academia-de-la-lengua-guarani/.

Secretaría de Políticas Lingüísticas (Paraguay). Ley nº 4251/2010: Ley de lenguas (texto bilingüe). PDF, 2010. URL https://spl.gov.py/files/legal/Ley%204251%20-%20bilingue.pdf.

Jahanzeb Sherwani, Nosheen Ali, Carolyn Penstein Rosé, and Roni Rosenfeld. Orality-grounded hcid: Understanding the oral user. *Information Technologies & International Development*, 5(4):37–49, 2009. URL https://itidjournal.org/index.php/itid/article/download/422/422-1096-2-PB.pdf.

Gabriel Skantze. Turn-taking in conversational systems and human-robot interaction: A review. *Computer Speech & Language*, 67:101178, 2021. doi: 10.1016/j.csl.2020.101178.

Tanya Stivers, N. J. Enfield, Penelope Brown, Christina Englert, Makoto Hayashi, Trine Heinemann, G. Hoymann, Federico Rossano, Jan P. de Ruiter, Kyung-Eun Yoon, and Stephen C. Levinson. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26):10587–10592, 2009. doi: 10.1073/pnas.0903616106. URL https://www.pnas.org/doi/10.1073/pnas.0903616106.

The University of Western Australia. First nations people to benefit from inclusive technology partnership. UWA News, 02 2025. URL https://www.uwa.edu.au/news/article/2025/february/first-nations-people-to-benefit-from-inclusive-technology-partnership.

Mark Turin. Voices of vanishing worlds: Endangered languages, orality, and cognition. *Análise Social*, 205(47), 2012. URL https://www.researchgate.net/publication/262778986_Voices_of_vanishing_worlds_Endangered_languages_orality_and_cognition.