

From Tone Sandhi to Multimodal Archives: Language, Oral Knowledge, and Digital Heritage among the Jino

Yuchen Li · Han Chen · Rui Zhang · Zihan Wang

Abstract

Research on the Jino (Jinuo) of Xishuangbanna has developed along several partly disconnected lines. Linguistic studies have described the grammatical and phonological structure of the language, ethnographic and ethnobiological work has documented oral traditions and ecological knowledge, and more recent digital-humanities research has explored computational approaches to cultural preservation. This review brings these strands into a single framework and argues that they are best understood as different levels of the same documentation problem. For a community whose traditional knowledge is transmitted primarily through speech, performance, objects, images, and place-based practice rather than through an inherited writing system, linguistic form and cultural content cannot be archived independently. The discussion first revisits the linguistic and sociocultural setting of Youle and Buyuan Jino, then examines the significance of acoustic work on tone sandhi in the Baka variety (Zhou 2024). It next considers oral and ecological knowledge as forms of situated heritage and evaluates the role of multimodal retrieval-augmented generation in organising text, audio, and image resources (Zhou and Lin 2026). The review concludes by proposing a research agenda centred on dialect-sensitive corpora, genre-aware annotation, provenance, and community governance. Taken together, the literature suggests that durable digital heritage for the Jino requires not simply more data, but a closer integration of phonological analysis, ethnographic context, and culturally accountable information infrastructure.

Keywords: Jino; Jinuo; tone sandhi; oral tradition; traditional knowledge; digital heritage; multimodal RAG; Tibeto-Burman

1. Introduction: one heritage problem, several research traditions

The Jino occupy a distinctive position in research on minority languages and cultural heritage in Southwest China. The community is concentrated primarily in the Jino Mountains (Youle) of Jinghong, Xishuangbanna, with a smaller Buyuan-speaking population in the Mengwang area. Jino varieties belong to the Lolo-Burmese (Yi-Burmese / Ngwi) branch of Tibeto-Burman and coexist with long-standing contact from Chinese and Tai Lue (Dai) (Gai 1986; Jiang 2010). Estimates of the speaker population vary considerably, in part because different sources count language competence and bilingual use differently. What is consistent across the literature, however, is that Jino continues to function in village life while Chinese has a much stronger presence in formal education and administration.

The absence of an inherited Jino writing system has shaped both cultural transmission and research practice. Histories of origin, ritual formulae, agricultural knowledge,

ecological classifications, and songs have circulated largely through spoken performance and embodied practice rather than through a local written canon. This makes Jino documentation a particularly revealing case for a broader question: what does it mean to preserve knowledge when the relevant evidence is distributed across speech, performance, landscape, material culture, photography, and video?

Earlier linguistic work established the descriptive basis for addressing that question. Gai (1986), Hayashi (2009), and Jiang (2010) provided major accounts of the language, while comparative and contact-oriented studies situated Jino within wider Lolo-Burmese history and the multilingual environment of Yunnan (Bradley 1979; Hayashi 2005, 2015). More recent research has added two new dimensions. Acoustic investigation of Baka Youle has shown that tone categories cannot be represented adequately by citation forms alone because regular sandhi patterns carry both synchronic and historical information (Zhou 2024). At the same time, multimodal digital-heritage work has explored how local text, audio, and image materials can be retrieved together rather than reduced to a single textual representation (Zhou and Lin 2026).

This review treats these developments not as separate topics but as parts of a shared problem of representation. The central argument is that the quality of a Jino digital archive depends on how accurately it represents linguistic variation, how carefully it preserves the social context of traditional knowledge, and how transparently it governs access to culturally sensitive materials. The following sections therefore move from language ecology and descriptive foundations to tone sandhi, oral knowledge, and finally digital infrastructure.

2. Language ecology, cultural memory, and the problem of representation

Jino cultural knowledge is closely tied to the mountain environment in which it is produced and transmitted. Ethnobiological research has documented local systems of land classification, agricultural calendars, preserved forest categories, medicinal plants, edible insects, hunting practices, and other forms of ecological knowledge (Long et al. 1999). A later survey by Huang et al. (2024) recorded 490 items of traditional knowledge related to biodiversity and reported substantial diversity across domains, especially in agricultural genetic resources and practices of sustainable use. These findings matter beyond ethnobotany. The names of plants, animals, tools, ritual objects, and landscape features are simultaneously ecological categories and linguistic data.

Oral tradition provides another layer of the same knowledge system. Jino origin narratives, including traditions associated with Amo Yaobai and the flood story of Mahei and Maniu, connect collective memory to kinship, territory, and relations with neighbouring groups. Public ritual and festival practices, especially Temaoke, likewise combine narrative, music, objects, performance, and social authority. The sun drum, smith-related traditions, and agricultural rites are meaningful not as isolated symbols but as parts of recurrent events in which speech, gesture, sound, costume, and place interact.

This multimodality complicates documentation. A transcription of a ritual speech event may preserve lexical content but omit intonation, voice quality, rhythm, audience

participation, and visual context. A photograph may preserve an object while losing its local name, associated narrative, or ritual restrictions. A video may retain many channels at once but still provide little value if the metadata does not identify speaker, village, event, or language variety. The documentation problem is therefore not simply one of collecting more media. It is one of maintaining relationships among media and between media and context.

The distinction between Youle and Buyuan Jino further reinforces this point. These varieties are often treated as related but not fully mutually intelligible lects, making the label “Jino language” internally complex (Hayashi 2009, 2015). A digital collection that does not record variety and locality risks collapsing meaningful differences. In a similar way, the increasing standardisation and public presentation of festivals can create a gap between village-specific ritual practice and staged heritage performance. Archives that combine such materials without provenance may preserve content while obscuring the circumstances in which that content was produced.

3. Descriptive foundations: structure, contact, and historical comparison

The modern description of Jino rests on several complementary traditions. Gai (1986) provided an early comprehensive outline, Hayashi (2009) developed a detailed description of Youle grammar, and Jiang (2010) offered a reference grammar that remains central to later work. Together these studies establish the phonological and grammatical baseline against which more specialised research can be interpreted.

Classification within Lolo-Burmese also provides an important comparative frame. Bradley’s reconstruction of Proto-Loloish (1979) distinguished major historical tone categories and supplied a framework for tracing how modern tone systems reflect earlier segmental and tonal contrasts. For Jino, this historical perspective is especially useful because surface pitch values do not necessarily map one-to-one onto stable lexical categories. A tone heard in isolation may have a different realisation when the same morpheme enters a compound or other disyllabic structure.

Internal comparison is equally important. Hayashi (2015), for example, shows that the phonological relationship between Youle and Buyuan involves more than a simple difference in tonal values; Youle preserves a richer fricative inventory. Such findings caution against treating the two varieties as interchangeable sources for a single uniform phonological description. They also suggest that future digital corpora should preserve the lect-specific form of each item rather than normalising material into a single “Jino” layer.

Language contact adds another dimension. Chinese and Tai Lue influence is visible in the lexicon, and Hayashi (2005) discusses Chinese influence in grammatical domains such as classifiers, complex constructions, and sentence-final particles. The picture that emerges is not one of wholesale structural replacement, but of a language whose core grammar coexists with contact-driven change in a multilingual environment. This context is directly relevant to heritage documentation because younger speakers may produce forms that differ from those recorded from older speakers, while borrowed vocabulary may interact differently with native phonological patterns.

Seen from this perspective, descriptive linguistics is not merely preliminary background for cultural research. It supplies the categories needed to identify what counts as the same linguistic item across speakers, genres, and recording conditions. Without that foundation, digital heritage systems risk indexing orthographic labels or rough translations while failing to capture the structure of the spoken language itself.

4. Tone sandhi in Baka Youle: why citation forms are not enough

Tone is one of the clearest examples of the gap between a simple archive entry and actual speech. Descriptions of Youle Jino commonly identify a five-tone citation inventory, often represented as /55, 44, 33, 31, 35/, although labels and distributions vary across sources and varieties (Gai 1986; Jiang 2010). Such inventories are indispensable for lexical description, but citation tones capture only one context of use.

Zhou (2024) addresses this limitation through acoustic analysis of monosyllables and disyllabic forms in the Baka variety of Youle Jino. The study shows that tone alternation in word formation is systematic rather than incidental. Importantly, the alternations are connected to historical tone classes: morphemes with comparable Proto-Loloish origins tend to pattern together in sandhi, even when the relevant historical categories are no longer transparent from citation pitch alone.

A second contribution concerns prosodic organisation. Reflexes associated with Proto-Loloish Tone *1 participate in changes that affect the relative prominence of syllables within disyllabic words. The resulting pattern is not well described as a mechanical replacement of one tone by another. Instead, tonal behaviour interacts with metrical structure, producing an unstressed-stressed, or iambic, organisation in relevant forms. This is significant because it links historical tonal development to synchronic word prosody.

Zhou (2024) also analyses restrictions on tonal combinations through the Obligatory Contour Principle. From this perspective, sandhi is not only a set of alternation rules attached to particular morphemes. It also reflects constraints on the tonal sequences that can surface inside the word. The study reports 105 alternating morphemes, of which 76 can be connected to Proto-Loloish reconstruction, and demonstrates that sandhi applies across many compounds rather than being confined to a single narrow morphological construction.

These findings have direct implications for documentation. A dictionary entry that records only a citation form may be formally correct and still fail to match the pronunciation that appears in compounds, songs, ritual expressions, or natural conversation. Likewise, a speech-retrieval system trained only on citation pronunciations may treat predictable sandhi variants as unrelated forms. For a language whose cultural record is overwhelmingly oral, this is not a marginal technical issue. It affects searchability, transcription, pedagogical representation, and the alignment of speech with lexical metadata.

Tone sandhi therefore illustrates a broader principle: linguistic archives should preserve alternation, not merely inventory. The value of Zhou’s acoustic work lies not only in refining the analysis of Jino phonology, but also in showing what a digitally useful linguistic record must contain if it is expected to support real speech rather than isolated wordlists.

5. Oral knowledge as situated and genre-dependent heritage

If tone sandhi shows why linguistic form changes with context, oral tradition shows why cultural meaning does as well. Jino verbal art includes narrative and social song traditions, ritual speech, origin narratives, and other forms of performance through which historical and practical knowledge is transmitted. In the absence of a traditional script, these genres have carried information that in literate traditions might be distributed across chronicles, manuals, genealogies, ritual texts, and specialised vocabularies.

The ethnobiological record demonstrates that this knowledge is also practical. Long et al. (1999) documented local categories and practices linked to land use, seasonal activity, forests, medicinal resources, and food. Huang et al. (2024) similarly emphasise both the richness of biodiversity-related knowledge and the pressures placed on its transmission by social and land-use change. The connection between these studies and language documentation is straightforward: specialised ecological knowledge is encoded in specialised vocabulary, classifications, narratives, and instructions, many of which may not appear in ordinary elicitation sessions.

Genre therefore matters. A botanical term elicited in isolation, the same term used in a farming explanation, and the same term sung in a narrative may differ in prosody, surrounding morphology, pragmatic meaning, and cultural significance. Treating all three instances as interchangeable “examples” risks losing information. A heritage corpus should instead record the relationship between linguistic form and event type: whether a recording is conversational, ritual, pedagogical, elicited, performed for tourists, or produced for documentation.

Authority and access are equally important. Traditional knowledge does not become context-free simply because it has been digitised. Village elders, ritual specialists, performers, and other knowledge holders may occupy different positions with respect to particular materials. Public festival performance may also differ from knowledge regarded as appropriate only within a village or restricted social setting. For this reason, provenance should be treated as part of the content itself rather than as optional administrative metadata.

At minimum, records should preserve information about speaker or knowledge holder, village, date or occasion, language variety, genre, recording conditions, and the intended status of the material. Such metadata does not solve questions of cultural authority, but it makes those questions visible and allows future users to distinguish between materials that might otherwise appear equivalent in a search interface.

6. From digitisation to retrieval: multimodal RAG and cultural heritage

The challenge of organising Jino materials becomes more apparent once documentation moves beyond conventional text archives. Text is searchable with relatively mature tools, but oral-heritage collections often contain audio, photographs, video, and descriptive notes whose relationships are difficult to query across media. A user may know the visual appearance of a ritual object but not its name, or may remember a spoken explanation without knowing how it was transcribed. In such cases, a text-only catalogue reproduces the very fragmentation that multimodal documentation is meant to overcome.

Zhou and Lin (2026) address this problem through a lightweight multimodal retrieval-augmented generation system focused on Jino traditional knowledge. Their architecture ingests text, audio, and image resources, represents content through multilingual embeddings generated with m3e-base, and uses FAISS for similarity-based retrieval. A Flask service supports cross-modal interaction, while large language models assist with query handling and grounded response generation. Docker, Cloudflare Tunnel, and a WordPress-facing interface are used to make the system deployable without a large institutional technical stack.

The importance of this approach lies less in any individual software component than in the retrieval model it embodies. RAG separates the generative interface from the evidential base: rather than relying on a language model to “know” Jino heritage from pretraining, the system attempts to retrieve locally curated resources and ground outputs in those materials. This is especially relevant for a small community whose language and cultural practices are unlikely to be represented adequately in general-purpose model training data.

The system also exposes several unresolved issues. Multilingual embedding models optimised for large languages may not encode fine-grained Jino phonetic distinctions, including predictable sandhi variants. An apparently successful semantic match may therefore overlook precisely the linguistic distinctions that matter for speech documentation. Conversely, a technically accurate retrieval can still be culturally inappropriate if the underlying item should not be publicly accessible. The problem is consequently both computational and institutional.

Zhou and Lin (2026) explicitly note ethical collection, intellectual property, customary ownership, and cultural authenticity as important constraints. Their work is best understood as a technical exploration rather than as evidence that the governance problem has been solved. The next stage of development should therefore evaluate not only retrieval accuracy but also source attribution, access control, speaker and village coverage, and the ability of community members to determine how knowledge is represented.

Placed beside Zhou (2024), the significance becomes clearer. Acoustic phonology addresses the internal form of spoken items; multimodal retrieval addresses the relationships among knowledge objects. A robust digital-heritage environment needs both.

Search infrastructure cannot compensate for inaccurate linguistic representation, while detailed phonological analysis has limited heritage value if recordings and annotations remain isolated and difficult to retrieve.

7. Toward an integrated research agenda

The literature reviewed here suggests that future Jino research would benefit from treating documentation as an integrated workflow rather than as a sequence of separate projects. Three priorities follow from the existing studies.

First, dialectal and speaker coverage should be expanded. Acoustic results from Baka Youle provide a strong basis for understanding sandhi, but they should be compared systematically with other Youle communities and with Buyuan. This would clarify which patterns are local, which are shared, and how historical correspondences are realised across present-day varieties. It would also make digital search more robust by preventing a single lect from becoming the implicit norm for the whole language.

Second, corpora should be genre-sensitive and multimodal from the outset. Wordlists, natural conversation, narrative song, ritual speech, ecological explanation, photographs, and video should not be collected as unrelated datasets if they refer to the same knowledge domains. Aligning these materials would make it possible to examine how lexical forms, sandhi patterns, performance style, and cultural meaning vary across contexts. For speech-related resources, annotations should include both citation-level lexical identity and context-specific surface forms.

Third, technical design should be coupled with community governance. The studies by Long et al. (1999) and Huang et al. (2024) demonstrate the richness of local ecological knowledge, while Zhou and Lin (2026) demonstrate the feasibility of multimodal retrieval. Yet greater technical accessibility is not automatically equivalent to better preservation. A sustainable system needs explicit rules for attribution, sensitivity, access level, correction, and removal, together with meaningful roles for Jino knowledge holders in deciding those rules.

These priorities point toward a model of digital heritage in which language analysis, ethnography, and computation are mutually dependent. Linguistic analysis identifies distinctions that must be represented accurately. Ethnographic context explains when those distinctions matter and who has authority over the material. Computational systems organise and retrieve the resulting records, but should remain accountable to both kinds of knowledge.

8. Conclusion

Jino scholarship has progressed from foundational grammatical description and comparative reconstruction toward acoustic measurement and multimodal digital experimentation. This trajectory creates an opportunity to rethink what counts as adequate documentation. For Jino, a durable record cannot consist only of transcribed words, only of ethnographic description, or only of digitised media files. The language is tonal and context-sensitive; traditional knowledge is oral, practical, and genre-dependent;

and cultural materials are distributed across sound, image, performance, objects, and place.

The existing literature already provides the components of a more integrated approach. Descriptive grammars establish the linguistic baseline (Gai 1986; Hayashi 2009; Jiang 2010); comparative work provides historical structure (Bradley 1979; Hayashi 2015); contact research situates ongoing change (Hayashi 2005); ethnobiological studies document ecological knowledge and its transmission (Long et al. 1999; Huang et al. 2024); acoustic analysis demonstrates the importance of context-dependent phonological form (Zhou 2024); and multimodal RAG offers a prototype for cross-media retrieval (Zhou and Lin 2026).

The next challenge is to connect these components without flattening their differences. The strongest future archive would be dialect-sensitive, acoustically informed, genre-aware, provenance-rich, multimodal, and governed in collaboration with the communities whose knowledge it represents. Under that model, digital heritage is not simply the conversion of cultural materials into files. It is the construction of relationships among linguistic forms, knowledge holders, media, contexts, and permissions so that those materials remain intelligible and meaningful over time.

References

- Bradley, David. 1979. *Proto-Loloish*. London / Malmö: Curzon Press.
- Gai, Xingzhi [盖兴之]. 1986. *Jinuoyu jianzhi* [Outline of the Jino language]. Beijing: Nationalities Publishing House.
- Hayashi, Norihiko [林範彦]. 2005. Hanyu dui Jinuoyu de yingxiang: yufa fangmian [Chinese influence on Jino: grammatical aspects]. *Kyoto University Linguistic Research* 24: 211-220. <https://doi.org/10.14989/87850>
- Hayashi, Norihiko. 2009. *A descriptive study on the grammar of the Jino language (Youle dialect)* [チノ語文法（悠樂方言）の記述研究]. Kobe: Research Institute of Foreign Studies, Kobe City University of Foreign Studies. (Research Monographs No. 43).
- Hayashi, Norihiko. 2015. Origins of Jino Fricatives. *Bulletin of Chinese Linguistics* 8(1): 61-77. <https://doi.org/10.1163/2405478X-00801004>
- Huang, Qing, Yinzhi Kuang, Hao Zhou, Xunqi Li, and Lun Yin. 2024. Biodiversity conservation in Xishuangbanna, China: diversity analysis of traditional knowledge related to biodiversity and conservation progress and achievement evaluation. *Diversity* 16(5): 260. <https://doi.org/10.3390/d16050260>
- Jiang, Guangyou [蒋光友]. 2010. *Jinuoyu cankao yufa* [A reference grammar of the Jino language]. Beijing: China Social Sciences Press.
- Long, Chunlin, Taku Abe, Hong Wang, Meilan Li, Hengmei Yan, and Yilan Zhou. 1999. Biodiversity management and utilization in the context of traditional culture of Jinuo society in S Yunnan, China. *Acta Botanica Yunnanica* 21(2).
- Zhou, Xiaoyu. 2024. Acoustic experimental study on the tone sandhi of the Jino language. *Linguistics of the Tibeto-Burman Area* 47(2): 254–283. <https://doi.org/10.1075/lbta.00021.zho>

Zhou, Xiaoyu, and Jing Lin. 2026. Multimodal RAG for cultural heritage: a technical exploration based on Jino traditional knowledge. *Digital Scholarship in the Humanities* 41(2): 1138–1153.
<https://doi.org/10.1093/llc/fqag042>