

LANGUAGE DOCUMENTATION AND REVITALISATION OF ENDANGERED AFRICAN LANGUAGES: STRATEGIES FOR SUSTAINABLE PRESERVATION

Onyinye Constance Amamgbo (PhD)

Department of Igbo, African and Communication Studies,

Nnamdi Azikiwe University, Awka.

oc.amamgbo@unizik.edu.ng

08033086688

Abstract

Language documentation and revitalisation of African endangered languages in so many ways and then, distinguishing this technical and methodological undertaking from the broader questions of cultural heritage and identity addressed elsewhere in this set of papers. The researcher noticed that whenever language disappears, the meaning of and the means of communication loses. This study uses a qualitative documentary design, the study draws on documentary linguistics scholarship, language endangerment assessment frameworks and case material on revitalisation efforts across several African language communities, particularly the Khoisan languages of southern Africa and small, undocumented languages of Nigeria and Cameroon. Revitalisation strategies that have shown real promise elsewhere in the world, particularly language nests and master-apprentice pairing, have been tried only sparingly in Africa, held back less by the strategies' unsuitability than by a severe shortage of trained documentary linguists, sustained funding and community-controlled infrastructure for archiving and teaching. The aims of this study is for proper documentation of some linguistic features and also examine some factors responsible for language endangerment. The findings show that a meaningful share of Africa's most endangered languages now have fewer than a dozen fluent first-language speakers, most of them elderly, and that existing documentation, where it exists at all, is frequently incomplete, held in inaccessible formats, or produced without lasting benefit to the community it describes. The paper argues that documentation and revitalisation must be treated as two connected but distinct tasks, and that community ownership of both is what ultimately determines whether documentation work leaves behind a living resource or simply an academic record of a language no longer spoken.

Keywords: Endangerment, Documentation, Revitalisation, Languages.

Introduction

Somewhere in southern Africa, in the years around the turn of this century, a small research team sat with the last handful of people who still spoke N|uu as a first language, running a tape recorder and trying to capture as much of the language's sound system, vocabulary and grammar as they could before it was too late. This scene, repeated with variations across the continent, N|uu, various Khoisan languages, dozens of small, undocumented languages spread across the linguistic patchwork of Nigeria, Cameroon and elsewhere, is the specific, urgent, often quietly heartbreaking work this paper is about.

Africa is home to somewhere between one thousand five hundred and two thousand living languages, depending on how closely related varieties are counted, and a meaningful share of them are spoken by populations small enough, and ageing enough, that serious linguists now classify them as endangered or even moribund, meaning no children are currently learning them as a first language. This paper is concerned specifically with what happens, or fails to happen, once a language reaches that point: how

its endangerment is measured, what documenting it properly actually involves, and what, if anything, can realistically be done to slow or reverse its decline.

It is worth pausing on why this matters beyond the purely linguistic interest such languages hold for researchers. Each of these languages represents a distinct, irreplaceable record of human cognitive and cultural achievement, encoding sound systems, grammatical structures and ways of categorising the world that took many centuries to develop and that, once the last fluent speaker is gone, cannot be reconstructed from outside no matter how skilled the researcher attempting it. The Khoisan click languages in particular are frequently cited in the linguistics literature as containing sound systems and grammatical features found in few or no other language families on earth, meaning their loss would represent not simply the disappearance of one more language among many but the closing off of an entire, distinctive branch of what human language is capable of doing.

This is a deliberately narrower and more technical question than the broader cultural preservation theme addressed elsewhere in this set of papers. Documenting a language, producing a dictionary, a grammar, a body of recorded and transcribed speech, is a distinct undertaking from preserving the cultural practices a language carries, and revitalising a language, actually increasing the number of speakers or restoring intergenerational transmission, is a further, distinct undertaking again, one that documentation alone does not achieve and was never designed to achieve on its own. Conflating the three has, if anything, made the underlying problem harder to see clearly and harder to fund appropriately.

This paper asks three linked questions: how is language endangerment in Africa actually measured and how severe is it for the continent's most at-risk languages; what does rigorous documentation of an endangered language involve in practice, and how well is that work currently being done; and what revitalisation strategies, tested elsewhere in the world or attempted within Africa itself, offer realistic paths back from the edge of language loss.

Literature Review and Theoretical Framework

Measuring Endangerment: From Fishman's Graded Intergenerational Disruption Scale to the Expanded Scale

Fishman's Graded Intergenerational Disruption Scale, developed originally to guide language revitalisation planning, ranks a language's vitality according to how far intergenerational transmission has broken down, from a language still used in advanced education and government at one end, to a language remembered only by a small number of elderly people with no one left to teach at the other. Lewis and Simons later expanded this into the Expanded Graded Intergenerational Disruption Scale, a thirteen-level framework now widely used by linguists and organisations such as Ethnologue to classify language vitality with more precision, distinguishing, for instance, between a language still spoken widely at home but absent from education, and one spoken only by a shrinking number of grandparents with no children learning it at all. This framework matters for the present discussion because it gives documentation and revitalisation planning a shared, reasonably precise vocabulary for describing how urgent a given language's situation actually is, rather than relying on vaguer terms like endangered that can mean very different things depending on who is using them.

Documentation Linguistics: Recording a Language Is Not the Same as Describing It

A second key theoretical contribution comes from Himmelmann's foundational distinction between language documentation and language description. Documentation, in this framework, refers to the systematic recording and archiving of a language as it is actually used, natural speech, conversation, narrative, ritual language, ideally captured on audio or video and annotated in enough detail that future researchers, or future members of the speech community itself, can access and understand it without needing the original speakers present. Description, by contrast, refers to the analytical work of writing grammars and dictionaries that summarise and explain the patterns found in that documented material. Himmelmann's point, influential across the field since, is that documentation deserves to be

treated as a legitimate goal in its own right, not merely as raw material collected on the way to writing a grammar, because a well-archived body of documented speech remains useful to a community and to future researchers even if a formal grammatical description is never completed, whereas a grammar written without adequate underlying documentation cannot later be checked, expanded or corrected once the last speakers are gone.

Community-Based Language Planning and the Question of Ownership

A third relevant framework comes from Grenoble and Whaley's work on language revitalisation planning, which argues that revitalisation efforts succeed or fail largely according to how much genuine control and ownership the speech community itself holds over the process, as opposed to having documentation and revitalisation designed and run primarily by outside academic researchers with their own institutional priorities and timelines. This framework is particularly relevant to the African context given the continent's long and, in places, still-continuing history of documentation conducted by foreign researchers who extracted linguistic data for academic publication with limited benefit returned to the community that provided it, an extractive pattern that community-based language planning explicitly positions itself against.

Taken together, these three frameworks structure the analysis that follows. The Graded Intergenerational Disruption Scale and Expanded Graded Intergenerational Disruption Scales provide a way of measuring how urgent a given language's situation is. Himmelmann's documentation-description distinction clarifies what documentation work actually needs to achieve and why it should not be judged solely by whether it produces a finished grammar. Grenoble and Whaley's community ownership framework provides the standard against which the sustainability of any given documentation or revitalisation effort should ultimately be judged.

It is worth noting briefly what this paper is not attempting. It does not offer a comprehensive census of every endangered African language, a task well beyond the scope of a single paper and one already pursued, with varying completeness, by resources such as Ethnologue and UNESCO's own endangerment atlas. Instead, the aim is to draw out, through the three theoretical lenses above, the recurring methodological and structural patterns that determine whether documentation and revitalisation efforts for any given endangered African language succeed, stall or arrive too late, patterns that appear to hold with reasonable consistency across the specific cases examined here despite the very different languages and communities involved.

Methodology

This study used a qualitative documentary design. Sources included documentary and descriptive linguistics scholarship, language endangerment databases and assessment reports, and case material on documentation and revitalisation projects involving specific African language communities, with particular attention to the Khoisan languages of southern Africa, including N|uu and various San languages, and to small, poorly documented languages of the Nigeria-Cameroon borderland region, an area recognised in the linguistic literature as containing an unusually high concentration of small, under-researched languages. Materials were selected for relevance to three themes: the measurement and current severity of language endangerment in Africa, the methodology and practical challenges of documentation work, and the design and outcomes of revitalisation strategies attempted within or adapted to African contexts. Sources were coded thematically and compared across cases to identify patterns holding broadly across the continent as distinct from patterns specific to a particular language community. Where sources disagreed on classification or on the number of remaining speakers for a given language, both figures are noted rather than silently resolved, since speaker counts for the most severely endangered languages are themselves often contested and change quickly. Attribution followed standard scholarly convention throughout.

Findings

A Significant Share of Africa's Languages Sit at the Most Severe End of the Endangerment Scale

Applying EGIDS-style classification across the continent's language inventory, the literature identifies a substantial number of African languages, concentrated particularly among the Khoisan languages of southern Africa and smaller language communities in the Nigeria-Cameroon borderland, the Nuba Mountains and parts of the Rift Valley, that sit at the most severe end of the scale, spoken only by a small, ageing population of first-language speakers with no children currently acquiring the language at home. Several specific languages, N|uu prominent among them, are documented in the literature as having reached single digits of fully fluent first-language speakers, placing them at genuine risk of moving from severely endangered to extinct within a single human generation.

The Last-Speaker Problem Creates Extreme Time Pressure on Documentation Work

Case material on N|uu and comparable Khoisan languages illustrates a pattern that recurs across the most severely endangered languages reviewed here: documentation work conducted under acute time pressure, with researchers aware that any given recording session might be the last opportunity to capture a particular speaker's knowledge before illness, further ageing or death removes it permanently. This pressure has real methodological consequences, since documentation carried out hurriedly, often with limited funding and a small research team working against a closing window, tends to be less systematic and less complete than documentation of a healthier, more stable language community would allow, meaning the historical record of some of Africa's most linguistically distinctive languages, including some of the click consonant systems found nowhere else in the world, is inevitably thinner than researchers would prefer.

Existing Documentation Is Often Incomplete, Fragmented or Hard to Access

Beyond the last-speaker cases specifically, the findings show a broader pattern of fragmented documentation across many African languages that are endangered but not yet at the most extreme end of the scale. Older documentation, some dating from the colonial and early postcolonial period, frequently exists only in wordlists or partial grammatical sketches rather than the fuller, richly annotated recordings that modern documentary linguistics considers standard, and much of it is held in physical archives, university collections or individual researchers' personal files rather than in digital repositories accessible to either later researchers or the communities the material describes. This fragmentation means that even where documentation technically exists, it frequently cannot be readily located, accessed or built upon, effectively repeating some of the same information-loss risk that motivates documentation work in the first place.

Digital Tools Have Improved Documentation Quality but Not Solved the Access Problem

Advances in digital audio and video recording, automated transcription assistance and online linguistic archiving platforms have measurably improved what documentation projects can capture and how durably they can store it, compared to the paper-based wordlists and analogue tape recordings of earlier decades. However, the findings suggest that digital archiving has not, on its own, solved the access and ownership problem identified above. Digital archives housed at foreign universities, however well organised technically, do not automatically translate into meaningful access for the community whose language was recorded, particularly where that community has limited internet connectivity, no institutional relationship with the archiving university, or was never given copies of the recordings in a usable format in the first place.

Language Nests Have Worked Well Elsewhere but Remain Rare in Africa

Language nest programmes, immersive early childhood settings where young children are cared for and taught entirely in an endangered language by fluent elder speakers, originated in New Zealand's Māori revitalisation movement and have since been adapted successfully in a range of other contexts, generally showing strong results in restoring at least some intergenerational transmission where implemented with adequate funding and community support. The case material reviewed here finds only a small number of documented attempts at language nest-style programmes for African languages, and most of these remain small in scale, pilot-level and dependent on external funding

rather than integrated into national early childhood education systems, suggesting the approach's applicability to the African context has been demonstrated only narrowly rather than tested at anything like the scale the scope of the problem would require.

Master-Apprentice Pairing Suits Africa's Last-Speaker Cases Particularly Well

The master-apprentice model, in which a committed adult learner is paired one-on-one with a fluent elder speaker for sustained, immersive language learning outside a formal classroom setting, is identified in the literature as particularly well suited to exactly the kind of extremely low-speaker-count situation found among languages like N|uu, where there are simply not enough fluent speakers remaining to staff a full language nest or classroom-based programme. Case material describing master-apprentice or closely comparable pairing efforts among Khoisan language communities suggests the model can produce genuinely new speakers even from a starting population of only a handful of elders, though the number of new speakers produced this way remains small relative to what would be needed to restore a language to broader community use rather than simply preserving a narrow thread of continued transmission.

A Severe Shortage of Trained Documentary Linguists Constrains Everything Else

A recurring and, in some ways, foundational finding across the literature concerns capacity. There are simply too few trained documentary linguists, African or foreign, working on the continent's most endangered languages relative to the number of languages that need urgent attention, and the specific technical skills documentation work requires, phonetic transcription, corpus annotation, archiving standards, are taught in relatively few African university linguistics programmes at present. This shortage means that documentation priorities are often set less by which languages are most urgently endangered and more by which languages happen to have an available, funded researcher interested in them, a somewhat arbitrary basis for allocating attention to what is, in most cases, an irreversible and time-limited opportunity.

Funding Remains Short-Term, Project-Based and Poorly Matched to the Timescale of the Problem

The findings also point to a persistent mismatch between how documentation and revitalisation work is typically funded, through relatively short academic grants tied to specific research projects, and the actual timescale the underlying problem demands, since meaningful revitalisation, in particular, generally requires sustained effort across years or decades rather than the two or three years a typical grant cycle covers. Several sources note that promising documentation or early revitalisation initiatives have stalled or been abandoned partway through, not because the underlying approach failed, but because project funding ran out before the work could reach a stage where it might become self-sustaining within the community itself.

Community Attitudes Toward Documentation Are Not Always Straightforwardly Positive

A further finding, easy to overlook if documentation is assumed to be an unambiguous good from the community's perspective, concerns mixed or cautious community attitudes toward outside researchers seeking to document an endangered language. Case material describes instances where community members, particularly among groups with a long history of being studied by outside researchers without much visible benefit in return, expressed scepticism about the motives and eventual usefulness of yet another documentation project, especially where earlier projects had extracted recordings and data for academic publication without leaving the community anything usable. This finding reinforces the relevance of Grenoble and Whaley's ownership framework directly, since it shows that community wariness is often not resistance to documentation as such, but a reasonable response to a documentation practice that has, in many past cases, treated the community as a source of data rather than a partner in the work.

Multilingual Repertoires Complicate the Idea of a Single Endangered Language to Be Saved

A final theme in the findings concerns the complexity of many African speakers' actual linguistic repertoires. In several of the smaller language communities reviewed, particularly in the Nigeria-Cameroon borderland, individual speakers are frequently multilingual across several small local

languages plus a regional lingua franca, and the boundaries between neighbouring related varieties are not always as sharp in practice as language classification systems suggest. This complicates the sometimes-implicit assumption behind documentation and revitalisation work that each endangered language is a single, clearly bounded object to be separately recorded and, ideally, separately revived. Several sources argue that documentation projects designed around a single named language in isolation risk missing important parts of the actual linguistic picture, including how speakers move between related varieties and what that movement itself reveals about the region's linguistic history and current dynamics.

Discussion

Read through Himmelmann's documentation-description distinction, several of these findings come into sharper focus. The fragmentation and access problems identified in the literature suggest that a great deal of African language documentation to date has prioritised description, the production of grammars and academic publications, over documentation in Himmelmann's stricter sense, the creation of a durable, accessible, richly annotated record of the language as actually spoken. This ordering of priorities is understandable given how academic incentives are usually structured, publications count toward a researcher's career in a way that raw archived recordings generally do not, but it has left many African languages with published academic descriptions sitting alongside a comparatively thin and hard-to-access underlying documentary record, precisely the situation Himmelmann's framework warns against, since a description cannot later be corrected or expanded without a solid documentary base to return to.

The last-speaker cases, N|uu chief among them, illustrate both the best and the most sobering aspects of current documentary practice. The urgency of these situations has produced some genuinely careful, well-resourced documentation work, precisely because the stakes are so visible and so final. But this same urgency reveals an uncomfortable truth about how documentation priorities get set across the continent more broadly: a language typically attracts serious, well-funded attention once it is already down to a handful of speakers and the loss is imminent and dramatic, rather than earlier, when a language has, say, a few thousand ageing speakers and a slower but still real trajectory toward the same endpoint. A genuinely proactive approach to endangerment, informed by EGIDS-style classification, would prioritise documentation before a language reaches crisis point, when there is more time, more speakers to draw on, and more chance that documentation work can feed directly into revitalisation rather than functioning mainly as a historical record after the fact.

The findings on language nests and master-apprentice pairing suggest these two strategies are not simply interchangeable options but suited to different stages of endangerment. Language nests depend on having enough fluent speakers, and enough community infrastructure, to staff an ongoing immersive childcare setting, conditions that a language with only a handful of elderly speakers left, like N|uu, simply cannot meet. Master-apprentice pairing, by contrast, can function even at that extreme end of the scale, since it requires only one fluent speaker willing and able to work closely with one committed learner. This suggests that revitalisation strategy should be matched explicitly to a language's position on the EGIDS scale rather than treating one internationally successful model, language nests in particular, as a universal template to be applied regardless of how many speakers actually remain.

The capacity and funding findings, taken together, point to what is arguably the deepest structural problem running through this entire area of work: documentation and revitalisation for Africa's most endangered languages are currently organised around the availability of individual researchers and individual grants, rather than around a coordinated, continent-wide assessment of which languages are most urgently at risk and a matched, sustained allocation of funding and trained personnel to address that risk in order of urgency. Fixing this would not necessarily require dramatically more total funding so much as a more deliberate, better-coordinated way of directing the funding and expertise that already exists, a planning failure as much as a resource failure.

Read through Grenoble and Whaley's community ownership framework, the access and archiving findings carry a clear implication: documentation that is not returned to, and controlled by, the

community it describes has only partially succeeded, even where it is technically excellent by academic standards. A beautifully annotated digital archive sitting on a foreign university server, inaccessible to the community whose grandparents' voices it contains, represents documentation without the ownership component that community-based language planning identifies as decisive for whether that documentation ever translates into revitalisation rather than remaining an academic artifact. This suggests that community access and control should be built into documentation project design from the outset, not added as an afterthought once the academic work is complete.

The finding on community wariness deserves to be taken seriously rather than treated as an obstacle to be managed around. A community that has previously been documented, published about and then left with nothing usable has good reason to be cautious about a new research team arriving with recording equipment and good intentions, and rebuilding that trust generally takes longer, and requires more visible reciprocity, than a single funded project cycle allows for. This is a further, human argument, alongside the funding-timescale argument made earlier, for treating documentation as long-term relationship building rather than a short, extractive research trip, since the quality of a community's cooperation, and therefore the quality and completeness of what gets documented, depends substantially on whether that trust has actually been earned.

The multilingual repertoire finding raises a genuinely difficult methodological question that this paper does not claim to resolve fully: what, exactly, is being documented and revitalised when the object in question is not a single, cleanly bounded language but a fluid, overlapping set of related varieties that speakers themselves move between without much difficulty. This does not undermine the case for documentation, since even a fluid, overlapping linguistic situation is still worth recording before it changes further, but it does suggest that documentation projects should be designed with enough flexibility to capture that fluidity itself, rather than forcing every recording session into a rigid single-language framework that may not match how speakers actually experience their own linguistic repertoire.

Conclusion and Implications

This paper examined the documentation and revitalisation of Africa's most endangered languages as a specific, technical undertaking distinct from broader cultural preservation. It found that a significant number of African languages, particularly among the Khoisan languages of southern Africa and small languages of the Nigeria-Cameroon borderland, sit at the most severe end of recognised endangerment scales, that existing documentation is often incomplete, fragmented or difficult to access even where it exists, that digital tools have improved documentation quality without solving the underlying access and ownership problem, and that revitalisation strategies such as language nests and master-apprentice pairing show real promise but have been applied only narrowly across the continent, held back chiefly by a shortage of trained linguists and by funding poorly matched to the timescale the problem actually demands. It further found that community attitudes toward documentation are shaped, reasonably, by past experience of extractive research practice, and that many small African language communities live within multilingual repertoires that do not always map neatly onto the single-language framework much documentation methodology still assumes.

For universities and research funders, the clearest implication is to expand training specifically in documentary linguistics within African institutions, rather than continuing to rely predominantly on foreign researchers working on short project cycles, and to fund documentation work earlier in a language's decline rather than waiting until it reaches crisis point. For archiving institutions, the findings support a clear priority on building accessible, community-controlled digital repositories rather than archives designed primarily around academic researcher access, including practical measures such as returning full copies of recordings directly to the communities that provided them.

For governments and regional bodies, a coordinated, continent-wide endangerment assessment, using EGIDS-style classification consistently across countries, would allow documentation funding and trained personnel to be directed according to actual urgency rather than according to which languages happen to have an interested researcher already working on them. For communities themselves, particularly those closest to the most severe end of the endangerment scale, the master-apprentice

model in particular offers a realistic, low-infrastructure starting point even where a full language nest or classroom programme is not yet feasible, and community organisations may find real value in pursuing this option directly rather than waiting for external researchers to initiate it.

For research ethics boards and funding bodies specifically, the community wariness finding points to a concrete, actionable change: requiring documentation projects to include a clear, binding plan for returning materials to the community in accessible form as a condition of funding, rather than treating this as a discretionary courtesy left to individual researchers' goodwill. For linguists designing documentation methodology, the multilingual repertoire finding suggests building enough flexibility into project design to capture code-switching and movement between related varieties as data in its own right, rather than treating it as noise to be filtered out in pursuit of a single, cleanly bounded target language. For scholarly associations and conferences such as this one, there is a further, smaller but tangible role to play in simply keeping this work visible, since sustained public and academic attention is itself one of the more reliable predictors of whether funding for a given endangered language community continues past an initial project's end.

In closing, the story of N|uu and languages like it is not really a story about the past, even though it is frequently told that way. It is a live, ongoing test of whether the field, and the funders, institutions and communities around it, can move fast enough and coordinate well enough to reach the continent's most endangered languages before their last fluent speakers are gone. Some of that work is already happening, carefully and well, in specific pockets. The task now is making it the rule rather than the exception, and starting earlier, for the languages that have not yet reached the edge but are visibly heading there. Every year that coordinated, adequately resourced attention is delayed is a year in which more of the continent's linguistic and cognitive inheritance quietly and irreversibly narrows, one ageing speaker at a time.

References

- Austin, P. K., & Sallabank, J. (Eds.). (2011). *The Cambridge handbook of endangered languages*. Cambridge: Cambridge University Press.
- Crystal, D. (2000). *Language death*. Cambridge: Cambridge University Press.
- Fishman, J. A. (1991). *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages*. Multilingual Matters.
- Grenoble, L. A., & Whaley, L. J. (2006). *Saving languages: An introduction to language revitalization*. Cambridge: Cambridge University Press.
- Güldemann, T., & Vossen, R. (2000). Khoisan. In B. Heine & D. Nurse (Eds.), *African languages: An introduction* (pp. 99–122). Cambridge University Press.
- Himmelman, N. P. (1998). Documentary and descriptive linguistics. *Linguistics*, 36(1), 161–195.
- Hinton, L., & Hale, K. (Eds.). (2001). *The green book of language revitalization in practice*. Academic Press.
- Lewis, M. P., & Simons, G. F. (2010). Assessing endangerment: Expanding Fishman's GIDS. *Revue Roumaine de Linguistique*, 55(2), 103–120.
- Sands, B. (2010). Jul'hoan documentation and the encoding of endangered languages. *Language Documentation and Conservation*, 4, 1–17.
- Simons, G. F., & Fennig, C. D. (Eds.). (2023). *Ethnologue: Languages of the world*. SIL International.
- Skutnabb-Kangas, T. (2000). *Linguistic genocide in education—or worldwide diversity and human rights?* Lawrence Erlbaum Associates.
- UNESCO. (2003). *Language vitality and endangerment*. UNESCO Ad Hoc Expert Group on Endangered Languages.
- Woodbury, A. C. (2003). Defining documentary linguistics. In P. Austin (Ed.), *Language documentation and description* (Vol. 1, pp. 35–51). SOAS.