

## Artículo original

### Inteligencia Artificial, propuestas de conservación y desafíos estructurales para las lenguas originarias del Perú

#### Artificial Intelligence, Preservation Initiatives, and Structural Challenges for the Indigenous Languages of Peru

#### Inteligência Artificial, propostas de preservação e desafios estruturais para as línguas indígenas do Peru

Víctor Gonzalo Ramírez Herrera

Universidad de Ingeniería y Tecnología-UTEC, Lima, Perú

[vramirez@utec.edu.pe](mailto:vramirez@utec.edu.pe)

<https://orcid.org/0000-0002-7274-010X>

### Resumen

#### Autor corresponsal:

Gonzalo Ramírez Herrera

[vramirez@utec.edu.pe](mailto:vramirez@utec.edu.pe)

#### Citar como:

Ramírez Herrera, V. G. (2026). Inteligencia Artificial, propuestas de conservación y desafíos estructurales para las lenguas originarias del Perú. *SYNTAGMAS* 5 (1), 01 – 16.

<https://doi.org/10.51343/syntagmas.v5i1.2035>

**Envío:** 30 de enero 2026

**Aceptado:** 18 de abril 2026

**Publicado:** 22 de junio 2026

#### Distribuido bajo:



**OPEN ACCESS**

Este artículo presenta una revisión crítica de iniciativas tecnológicas y mediáticas orientadas a la conservación de lenguas originarias en el Perú, desde el marco teórico de la revitalización lingüística (Fishman, 1991; UNESCO, 2003) y los estudios sobre tecnologías del lenguaje para lenguas de bajos recursos (Joshi et al., 2020; Kornai, 2013). Mediante una revisión de alcance (scoping review), se analizaron proyectos basados en inteligencia artificial (IA), corpus de voz, aplicaciones móviles, iniciativas estatales, revitalización mediática y centros de investigación desarrollados entre 2016 y 2024. Los casos examinados incluyen *Illariy*, *Huqariq*, *Yapaykuy*, *Uyariy*, el proyecto UNESCO-Grupo AJE y la Estación Científica Chana. El análisis identifica tensiones estructurales transversales: la escasez de datos digitales, la brecha tecnológica en comunidades indígenas, la sostenibilidad precaria de los proyectos, la insuficiente participación comunitaria en el diseño de herramientas, el riesgo de estandarización y uniformización lingüística, y la desigualdad educativa. Se concluye que la IA constituye un recurso promisorio pero insuficiente por sí solo para revertir la pérdida lingüística, y que su eficacia depende de condiciones estructurales, enfoques interculturales y marcos de colaboración no colonial con las comunidades hablantes.

**Palabras clave:** lenguas originarias, inteligencia artificial, tecnologías del lenguaje, revitalización lingüística, lenguas de bajos recursos, brecha digital, Perú, procesamiento del lenguaje natural

### Abstract

This article presents a critical review of technological initiatives aimed at the conservation of indigenous languages in Peru, framed within the theoretical

## **Artículo original**

---

traditions of language revitalization (Fishman, 1991; UNESCO, 2003) and research on language technologies for low-resource languages (Joshi et al., 2020; Kornai, 2013). Using a scoping review methodology, projects based on artificial intelligence (AI), voice corpora, mobile applications, state-led initiatives, media revitalization, and research centers developed between 2016 and 2024 were analyzed. Cases examined include Illariy, Huqariq, Yapaykuy, Uyariy, the UNESCO-Grupo AJE project, and the Chana Scientific Station. The analysis identifies cross-cutting structural tensions: scarcity of digital language data, the technological divide in indigenous communities, the precarious sustainability of projects, insufficient community participation in tool design, the risk of linguistic standardization, and educational inequality. It is concluded that AI represents a promising but insufficient resource on its own to reverse language loss, and that its effectiveness depends on structural conditions, intercultural approaches, and non-colonial collaboration frameworks with speaker communities.

**Keywords:** indigenous languages, artificial intelligence, language technologies, language revitalization, low-resource languages, digital divide, Peru, natural language processing

### **Resumo**

Este artigo apresenta uma revisão crítica das iniciativas tecnológicas e midiáticas voltadas para a conservação das línguas indígenas no Peru, a partir do marco teórico da revitalização linguística (Fishman, 1991; UNESCO, 2003) e dos estudos sobre tecnologias da linguagem para línguas com poucos recursos (Joshi et al., 2020; Kornai, 2013). Por meio de uma revisão exploratória (scoping review), foram analisados projetos baseados em inteligência artificial (IA), corpora de voz, aplicativos móveis, iniciativas governamentais, revitalização midiática e centros de pesquisa desenvolvidos entre 2016 e 2024. Os casos examinados incluem Illariy, Huqariq, Yapaykuy, Uyariy, o projeto UNESCO-Grupo AJE e a Estação Científica Chana. A análise identifica tensões estruturais transversais: a escassez de dados digitais, a lacuna tecnológica nas comunidades indígenas, a sustentabilidade precária dos projetos, a participação comunitária insuficiente no desenho de ferramentas, o risco de padronização e uniformização linguística e a desigualdade educacional. Conclui-se que a IA constitui um recurso promissor, mas insuficiente por si só para reverter a perda linguística, e que sua eficácia depende de condições estruturais, abordagens interculturais e marcos de colaboração não colonial com as comunidades falantes.

**Palavras-chave:** línguas indígenas, inteligência artificial, tecnologias da linguagem, revitalização linguística, línguas com poucos recursos, exclusão digital, Peru, processamento de linguagem natural

## **1. Introducción**

El Perú es uno de los países con mayor diversidad lingüística de América Latina: el Ministerio de Cultura (2023) reconoce 48 lenguas originarias pertenecientes a 20 familias lingüísticas, habladas por aproximadamente 4 millones de personas. Sin embargo, esta riqueza se encuentra bajo amenaza severa. Según el Atlas Interactivo

## **Artículo original**

---

de Lenguas en Peligro de la UNESCO, al menos 21 de estas lenguas se encuentran en situación de peligro o peligro crítico de desaparición, y algunas cuentan con menos de diez hablantes fluidos. Un ejemplo de esta situación es el taushiro, cuyo único hablante conocido reside en la región de Loreto.

La pérdida de una lengua supone mucho más que la extinción de un sistema de comunicación. Implica la desaparición de formas de conocimiento, cosmovisiones, prácticas culturales y memorias colectivas que no tienen equivalente en otras lenguas (Himmelman, 1998). Desde una perspectiva de derechos, la conservación lingüística es además una condición para garantizar el acceso equitativo a la educación, la salud, la justicia y la participación política de las comunidades indígenas (Leonard & Haynes, 2010).

En este contexto, la inteligencia artificial (IA) y las tecnologías digitales han emergido como herramientas con potencial para acelerar la documentación, revitalización y difusión de lenguas en riesgo. No obstante, como señala Bird (2020), el desarrollo de tecnologías del lenguaje ha reproducido históricamente sesgos coloniales, y ha privilegiado, en muchos casos de manera involuntaria, a las lenguas con mayor presencia digital y capital académico. Esto sucede mientras las lenguas indígenas permanecen subrepresentadas en los grandes modelos de lenguaje. En la misma línea, Joshi et al. (2020) documentan que el 88% de la investigación en procesamiento del lenguaje natural (PLN) se concentra en solo el 1% de las lenguas del mundo. Esto configura una jerarquía digital que replica y profundiza desigualdades preexistentes.

Frente a esta tensión, en el Perú han surgido en la última década iniciativas de diversas procedencias: universitarias, estatales, empresariales y comunitarias. Estas buscan aplicar tecnologías de IA a la conservación lingüística. El presente artículo se propone revisar sistemáticamente este conjunto de iniciativas, evaluar su potencial y sus limitaciones, e identificar los desafíos estructurales que condicionan su efectividad.

La revisión responde a tres justificaciones fundamentales. En primer lugar, la preservación de las lenguas originarias es una condición para garantizar derechos educativos, culturales y políticos de los pueblos indígenas. En segundo lugar, estas lenguas constituyen repositorios de conocimiento ancestral sobre ecosistemas, prácticas médicas, sistemas éticos y formas de organización social cuya pérdida es irreversible. En tercer lugar, la diversidad lingüística representa un patrimonio cognitivo colectivo cuya conservación enriquece la comprensión humana del lenguaje y del pensamiento.

## **2. Marco teórico**

### **2.1 Revitalización lingüística: enfoques y condiciones**

La revitalización lingüística es un campo interdisciplinario que estudia los procesos mediante los cuales

## **Artículo original**

---

comunidades de hablantes, junto con agentes externos, buscan revertir o detener el desplazamiento de una lengua amenazada. El modelo de referencia en este campo es la Escala de Disrupción Intergeneracional del Lenguaje (GIDS, por sus siglas en inglés) propuesta por Fishman (1991). Esta escala evalúa los grados de desarrollo o de extinción de las lenguas.

El grado de mayor estabilidad y seguridad para una lengua es su consideración de lengua internacional y el grado menor es el de extinta. Los grados intermedios pasan desde vulnerable a distinto grado de peligro de desaparición. Las lenguas originarias del Perú se encuentran entre los rangos de vulnerable y extinguida en su gran mayoría.

Un elemento central para la evaluación mencionada es la transmisión intergeneracional: considerada condición *sine qua non* para la supervivencia de una lengua. Sin esta base demográfica y comunitaria, ningún esfuerzo tecnológico puede garantizar la vitalidad lingüística a largo plazo.

La UNESCO (2003), en su marco para la evaluación de la vitalidad lingüística, establece nueve factores determinantes, entre los que destacan: la proporción de hablantes respecto a la población total de un país, la transmisión intergeneracional, la disponibilidad de materiales educativos, las políticas lingüísticas institucionales y las actitudes de la propia comunidad hacia su lengua. Este marco multifactorial es crucial para evaluar el alcance real de las iniciativas tecnológicas: una aplicación móvil puede contribuir a la presencia digital de una lengua, pero no puede sustituir la transmisión oral en el hogar ni modificar por sí sola las actitudes de depreciación lingüística internalizadas por hablantes en contextos de diglosia.

Rice (2011) y Leonard & Haynes (2010) insisten además en la dimensión ética de la documentación y revitalización: los proyectos que no involucran activamente a las comunidades en su diseño, implementación y gobernanza corren el riesgo de reproducir lógicas extractivistas. En estas, el conocimiento lingüístico, y todas las implicancias mencionadas, es capturado para fines académicos o comerciales sin retorno significativo para los hablantes.

## **2.2 Tecnologías del lenguaje para lenguas de bajos recursos**

En el campo del procesamiento del lenguaje natural, se denomina "lenguas de bajos recursos" (*low-resource languages*) a aquellas que carecen de corpus digitales suficientemente extensos para entrenar modelos de aprendizaje automático con precisión aceptable. Kornai (2013) estimó que de las aproximadamente 7,000 lenguas del mundo, solo entre 200 y 300 tienen perspectivas reales de sobrevivir en el ecosistema digital, dado el umbral mínimo de datos requerido para desarrollar tecnologías funcionales.

Joshi et al. (2020) clasifican las lenguas en cinco categorías según su representación en recursos digitales

## **Artículo original**

---

y en la investigación de PLN, desde las "the winners" (inglés, chino, árabe) hasta las "the left-behinds" (la vasta mayoría de las lenguas del mundo, incluyendo las amazónicas). Las lenguas originarias del Perú pertenecen en su mayoría a esta última categoría, lo que implica que los modelos de lenguaje existentes, como los grandes modelos de IA generativa, tienen un desempeño deficiente en estas lenguas y reproducen sus ausencias digitales. A este punto, añadimos, los estudios de clasificación lingüística fomentan con ello poca confianza en estas tecnologías al equivocar respuestas por mezclar variedades dialectales sin criterio.

Bird (2020) propone un enfoque decolonial para el desarrollo de tecnologías del lenguaje: en lugar de adaptar herramientas diseñadas para lenguas dominantes, aboga por arquitecturas y metodologías construidas con las comunidades hablantes, en respeto con sus epistemologías y formas de transmisión del conocimiento. Este enfoque es coherente con la crítica de Henrich, Heine & Norenzayan (2010) a la sobrerrepresentación de poblaciones WEIRD (Western, Educated, Industrialized, Rich, Democratic) en la investigación científica es el que orienta los proyectos más sólidos identificados en este panorama, como la Estación Científica Chana, mencionada más adelante.

### **3. Metodología**

El presente trabajo adopta el enfoque de la revisión de alcance (*scoping review*), según la propuesta metodológica de Arksey & O'Malley (2005) adaptada al campo de las humanidades digitales y los estudios lingüísticos. Este tipo de revisión es adecuada cuando el objetivo es mapear la evidencia disponible sobre un tema emergente, identificar brechas en el conocimiento y caracterizar las iniciativas existentes sin pretensión de meta-análisis cuantitativo. Lo central de esta propuesta es evidenciar con claridad los criterios y rutas de acceso a la información comentada (p. 23). Dicho trabajo lo presentamos en los siguientes párrafos.

Los criterios de inclusión fueron los siguientes: (a) iniciativas desarrolladas o implementadas en el Perú; (b) con componente tecnológico digital explícito orientado a lenguas originarias; (c) documentadas en fuentes primarias o secundarias accesibles entre 2016 y 2024. Se excluyeron proyectos sin documentación pública verificable o cuya información institucional no podía ser contrastada con al menos una fuente independiente.

Las fuentes consultadas incluyeron publicaciones académicas indexadas en Scopus y Google Scholar, documentos institucionales del Ministerio de Cultura del Perú, repositorios de conferencias especializadas en PLN (LREC, ACL, COLING) y fuentes periodísticas de referencia. Las iniciativas identificadas fueron organizadas en tres categorías analíticas: (1) proyectos basados en IA y tecnologías del lenguaje; (2) proyectos institucionales y estatales; y (3) iniciativas de revitalización mediática y comunitaria. Para cada caso, el análisis sigue un esquema uniforme: descripción, evaluación crítica de su potencial y sus limitaciones, y posicionamiento respecto al marco

## **Artículo original**

---

teórico.

Se identificaron en total ocho iniciativas con documentación suficiente, además de la Estación Científica Chana, que se analiza separadamente por su carácter integral y multidimensional. La cobertura lingüística de las iniciativas revisadas abarca principalmente el quechua (en diversas variedades), el aimara, el shipibo-konibo, el awajun y el asháninka, siendo estas las lenguas con mayor representación digital en el Perú.

### **4. Proyectos de IA para la conservación de lenguas indígenas**

#### **4.1 Proyectos basados en inteligencia artificial y tecnologías del lenguaje**

##### **4.1.1 Illariy: avatar de IA para la presencia digital del quechua, aimara y awajun**

El proyecto Illariy, desarrollado por un equipo de investigación de la Universidad Nacional Mayor de San Marcos (UNMSM), constituye una de las iniciativas más innovadoras en el campo de la IA aplicada a lenguas originarias del Perú. Su nombre, que significa "amanecer" en quechua, refleja adecuadamente su objetivo central: incrementar la presencia digital de lenguas que han sido sistemáticamente excluidas de los ecosistemas tecnológicos dominantes.

La iniciativa responde a un diagnóstico preciso: los grandes modelos de lenguaje no incorporan las lenguas originarias peruanas porque existe escasa producción de contenido digital en estas lenguas. Esto ha generado un círculo vicioso de exclusión tecnológica. En palabras del investigador responsable, Carlos Fernández (Noticiasncc, 2024):

"Vimos en la inteligencia artificial que no estaba hecha para las lenguas [originarias] porque era discriminatoria frente a ellas, y la convertimos en un elemento importante para evitar la extinción de nuestras 48 lenguas originarias. Comenzamos a enseñar un camino de cómo podemos preservar en esencia el idioma a través de plantillas fonéticas que nos permitieran darle un corpus al lenguaje y la posibilidad de tener una persistencia en la memoria mayor."

El proyecto desarrolla plantillas fonéticas especializadas y técnicas de transcripción supervisadas por lingüistas expertos y permite que el sistema se articule con pronunciación apropiada en quechua, aimara y awajun. Su producto más visible es un avatar de IA utilizado en el noticiario Letras TV Willakun, que narra contenidos en lenguas originarias, y una aplicación generativa de historias —Illariy Willarisunki, disponible en la GPT Store— que produce relatos acompañados de imágenes generadas por herramientas como DALL-E. Adicionalmente, el proyecto ofrece microclases de quechua en TikTok bajo el programa Kuska Yachay (Aprendiendo Juntos).

## **Artículo original**

---

Desde una perspectiva crítica, Illariy enfrenta dos desafíos estructurales de consideración. El primero es la escasez de corpus suficientemente extensos y dialectalmente diversos: el sistema corre el riesgo de reforzar una variante estándar que invisibiliza la diversidad interna de las lenguas implicadas. El segundo es la asimetría entre variedades lingüísticas: las variedades con más hablantes y mayor producción digital consolidarán mejor su presencia tecnológica, mientras las variedades minoritarias dentro de las propias lenguas indígenas quedarán aún más rezagadas. Este fenómeno de estandarización intra-lingüística constituye un riesgo documentado en la literatura sobre PLN para lenguas pluricéntricas (Joshi et al., 2020).

### **4.1.2 Huqariq: corpus de voz multilingüe para lenguas originarias peruanas**

El proyecto Huqariq representa una apuesta metodológicamente robusta para resolver uno de los cuellos de botella más críticos en el desarrollo de tecnologías del lenguaje para lenguas indígenas: la ausencia de corpus de voz a escala suficiente. Huqariq es una iniciativa académica y tecnológica centrada en la construcción de un corpus de audio multilingüe con transcripción, orientado a habilitar el desarrollo de sistemas de reconocimiento automático del habla (ASR), identificación de idioma y síntesis de voz para lenguas de bajos recursos.

Zevallos et al. (2022) reportan que el corpus cuenta con más de 220 horas de audio transcrito, aportado por más de 500 voluntarios hablantes de quechua sureño, quechua central, aimara y shipibo-konibo. Esto lo convierte en el recurso de este tipo más extenso disponible para lenguas originarias del Perú. La metodología de crowdsourcing utilizada (participación voluntaria de hablantes nativos) tiene la ventaja de reducir costos y de involucrar activamente a la comunidad en la construcción del recurso. Sin embargo, esta misma metodología introduce sesgos de selección que deben reconocerse: los voluntarios tienden a ser hablantes urbanos, jóvenes y con mayor alfabetización digital. Tal situación puede no representar adecuadamente la variación dialectal ni los registros propios de hablantes rurales o mayores.

La cobertura del corpus, aun siendo la más amplia disponible para el contexto peruano, sigue siendo insuficiente para entrenar modelos de ASR con precisión comercial: umbral que en lenguas de altos recursos como el inglés requiere miles de horas. Los propios autores reconocen este límite y plantean la expansión hacia las demás lenguas originarias del Perú como objetivo de mediano plazo. Desde el marco de Kornai (2013), Huqariq representa un avance significativo hacia la viabilidad digital de estas lenguas, aunque aún se encuentra por debajo del umbral necesario para garantizar su supervivencia plena en el ecosistema tecnológico.

### **4.1.3 Yapaykuy: tecnología inclusiva e interoperabilidad lingüística**

Yapaykuy, cuyo nombre en quechua significa "incluir/incluyeme", es una startup peruana de tecnología inclusiva cuyo producto principal es una aplicación móvil que integra traducción en tiempo real de lenguas

## ***Artículo original***

---

originarias (quechua, aimara, awajun, ashaninka y shipibo-konibo) y de lengua de señas, con conversión bidireccional voz-texto, interpretación gestual por cámara y pictogramas para usuarios con dificultades de lectura o escritura. La aplicación incluye, además, videollamadas con intérpretes nativos certificados y frases preconfiguradas para contextos de atención al ciudadano en salud, banca y servicios municipales.

El valor de Yapaykuy reside en su orientación práctica hacia la interoperabilidad lingüística en servicios públicos, un ámbito donde las barreras idiomáticas tienen consecuencias directas sobre el acceso a derechos. En este sentido, complementa la perspectiva de revitalización con una dimensión de derechos lingüísticos aplicados.

Adicionalmente, la startup ofrece un programa de certificación y formación organizacional que ayuda a entidades a cumplir con la normativa peruana sobre atención intercultural. Es, como indica su página, una certificadora de inclusión.

Sin embargo, Yapaykuy enfrenta limitaciones técnicas y de sostenibilidad que merecen atención crítica. El procesamiento de lenguas con baja digitalización presenta complejidad técnica elevada, y la precisión de la traducción automática en lenguas de bajos recursos es significativamente inferior a la alcanzada en lenguas globales. Esto puede generar errores con consecuencias prácticas en contextos de atención médica o legal. Adicionalmente, la continuidad del proyecto depende del financiamiento privado: dicha situación introduce incertidumbre sobre su viabilidad a largo plazo, un problema transversal al ecosistema de startups de impacto social.

## **4.2 Proyectos institucionales y estatales**

### ***4.2.1 Uyariy: atención intercultural en servicios públicos***

Uyariy, "escuchar" en quechua, es una aplicación digital desarrollada por el Ministerio de Cultura del Perú, disponible para dispositivos Android e iOS, que proporciona a servidores públicos cadenas de texto y audios en castellano y en lenguas originarias (quechua variedad Cusco Collao, shipibo-konibo, ashaninka, awajun y aimara) con las expresiones más frecuentes en la atención ciudadana. El proyecto fue pilotado en 2019 y validado en regiones como Ucayali, Junín, Amazonas y Puno, con el objetivo de reducir las barreras de comunicación en instituciones como el RENIEC.

Uyariy constituye un avance concreto en la institucionalización de los derechos lingüísticos en el aparato estatal, y su diseño es funcionalmente adecuado para el objetivo declarado. No obstante, su impacto real se ve limitado por factores de implementación: el uso efectivo de la aplicación por parte de los funcionarios es heterogéneo y en muchos casos esporádico. Esto sugiere la necesidad de estrategias de capacitación y seguimiento más robustas. Adicionalmente, la cobertura lingüística de la herramienta es aún reducida en relación

## ***Artículo original***

---

con el total de lenguas reconocidas, y las actualizaciones del contenido no siguen una cadencia sistemática. Sin embargo, el trabajo es una excelente propuesta con potencial de mejora continua.

### ***4.2.2 Talleres de IA del Ministerio de Cultura***

El Ministerio de Cultura ha impulsado también talleres de formación en IA orientados a jóvenes y hablantes nativos, con el propósito de democratizar el acceso al conocimiento técnico y de capacitar a potenciales facilitadores comunitarios en el uso de herramientas de IA para la documentación y promoción de sus lenguas. Esta iniciativa responde a uno de los problemas estructurales más críticos identificados en la literatura citada en el marco teórico: la brecha entre los desarrolladores de tecnología y las comunidades a las que estas herramientas pretenden servir.

Sin embargo, la efectividad de estos talleres está condicionada por la brecha digital de base: muchas de las regiones donde se concentran los hablantes de lenguas en mayor riesgo carecen de conectividad estable o de dispositivos con capacidad suficiente para operar las herramientas enseñadas. Sin resolver esta condición previa, los talleres corren el riesgo de beneficiar principalmente a poblaciones urbanas ya digitalmente conectadas y ampliar en lugar de reducir la brecha intracomunitaria.

### **4.3 Iniciativas de revitalización mediática y comunitaria**

#### ***4.3.1 Proyecto UNESCO-Grupo AJE: revitalización de lenguas en peligro crítico***

El proyecto impulsado por la UNESCO Perú en alianza con el Grupo AJE (a través del movimiento Amarumayu) y con el apoyo del Ministerio de Cultura se centra en tres lenguas amazónicas en situación de peligro crítico: ikitu, kukama-kukamiria y taushiro, todas habladas en la región de Loreto. El taushiro, en particular, representa un caso extremo: cuenta con un único hablante fluido, lo que hace que toda estrategia de revitalización —en sentido estricto— sea técnicamente inviable y que el objetivo alcanzable sea la documentación y la memoria cultural, y no, lamentablemente, la recuperación de una comunidad hablante activa.

La iniciativa articula metodologías de participación comunitaria como talleres con jóvenes y sabios locales, formación de revitalizadores lingüísticos, señalética comunitaria en lengua originaria, y uso herramientas digitales como videos animados en stop motion que narran historias culturales. Este enfoque es coherente con los principios de colaboración no colonial propuestos por Leonard & Haynes (2010) y Rice (2011): el conocimiento se construye con la comunidad, no sobre ella. La sostenibilidad del modelo, sin embargo, sigue siendo su mayor

## **Artículo original**

---

vulnerabilidad: la formación de jóvenes revitalizadores requiere continuidad institucional y recursos que el financiamiento de proyecto no garantiza a largo plazo.

### **4.3.2 Ñuqanchik y Ama Llulla: presencia mediática de las lenguas originarias**

El noticiero Ñuqanchik, producido por el Instituto Nacional de Radio y Televisión del Perú (IRTP) y transmitido de lunes a viernes por TVPeru y Radio Nacional desde 2016, constituye el medio de comunicación masivo más significativo en lengua quechua del país. Su importancia reside en la información que transmite, pero sobre todo en la normalización cotidiana del quechua como lengua apta para el debate público y la comunicación de actualidad. Con ello contribuye a modificar las actitudes de depreciación lingüística identificadas como un factor crítico en el modelo de vitalidad de la UNESCO (2003).

Ama Llulla, "no mientas" en quechua, es una red colaborativa de *fact-checking* impulsada con apoyo del PNUD y coordinada por medios como Ojo Público, que produce verificaciones de información en castellano, quechua y asháninka. Su aporte específico reside en la extensión de las prácticas de ciudadanía digital e información verificada a comunidades indígenas, y conectar la revitalización lingüística con el ejercicio de derechos cívicos. Ambas iniciativas enfrentan la limitación estructural de la escasez de periodistas y comunicadores bilingües formados. Además de los ya mencionados financiamientos inestables.

## **5. La Estación Científica Chana: un modelo integral**

La Estación Científica Chana para las Ciencias del Lenguaje y la Interculturalidad, inaugurada en octubre de 2023 en San José de Yarinacocha (Ucayali), representa la iniciativa de mayor profundidad epistémica y metodológica identificada en esta revisión. Desarrollada en colaboración entre la Pontificia Universidad Católica del Perú (PUCP), la Universidad de Zurich y el Instituto Max Planck para Antropología Evolutiva, el laboratorio, cuyo nombre remite a un ave amazónica capaz de imitar el canto de otras especies, simboliza la diversidad lingüística y la inteligencia cognitiva que orienta su trabajo.

Chana es dirigida por el lingüista Roberto Zariquiey, especialista en documentación y revitalización de lenguas indígenas amazónicas, y opera como un centro multidisciplinario que integra psicolingüística experimental, lingüística computacional, documentación lingüística y etnobiología. En el plano experimental, utiliza tecnología de *eye tracking* para investigar cómo los hablantes amazónicos procesan el lenguaje y las imágenes. En el plano computacional, desarrolla corpus, prototipos de traductores automáticos y correctores ortográficos para lenguas indígenas, entre ellos un prototipo de traductor shipibo-castellano.

El laboratorio impulsa además proyectos de aplicación directa en las comunidades, como la Escuelita

## **Artículo original**

---

Iskonawa, destinada a la enseñanza lúdica y tecnológica de la lengua iskonawa a niños de la comunidad, y estudios sobre cognición numérica en lenguas sin sistemas convencionales de número. Su importancia radica en su doble dimensión: ofrece infraestructura avanzada de investigación como laboratorios, salas de grabación, equipos especializados, ubicada en la Amazonia misma, en lugar de operar desde centros académicos lejanos; y promueve un enfoque científico comprometido con la devolución de conocimiento a las comunidades participantes.

Desde el marco teórico de este artículo, Chana encarna con mayor coherencia los principios de colaboración no colonial propuestos por Bird (2020) y Leonard & Haynes (2010): la investigación se produce desde y con las comunidades, en sus territorios, con metodologías validadas por los propios hablantes, y con un compromiso explícito de transferencia de capacidades a nuevas generaciones de investigadores amazónicos.

## **6. Desafíos estructurales transversales**

El análisis de las iniciativas presentadas permite identificar un conjunto de desafíos estructurales que trascienden las características específicas de cada proyecto y condicionan el impacto real de cualquier esfuerzo tecnológico en este campo. Estos desafíos no son meramente técnicos: reflejan condiciones sociales, económicas e institucionales que ningún desarrollo tecnológico puede resolver por sí solo. Es, en este sentido, que se reconocen como estructurales.

### **6.1 Escasez de digitalización lingüística**

La ausencia de corpus extensos, diccionarios digitales, bases de audio y herramientas adaptadas como teclados para sistemas ortográficos no latinos, es el límite técnico más inmediato para el desarrollo de tecnologías del lenguaje en lenguas originarias. Esta carencia es a la vez causa y efecto de la marginación digital: sin datos, no se pueden entrenar modelos; sin modelos funcionales, hay menos incentivos para producir datos. Superar este círculo requiere inversiones sistemáticas en infraestructura de datos, con participación activa de las comunidades hablantes y bajo marcos éticos de propiedad y gobernanza que reconozcan a estas comunidades como titulares y no solo como proveedores del conocimiento lingüístico.

### **6.2 Brecha tecnológica**

La desigualdad en el acceso a dispositivos, conectividad y alfabetización digital afecta tanto a los potenciales usuarios como a los posibles facilitadores comunitarios de estas herramientas. En el Perú, las regiones con mayor concentración de hablantes de lenguas en riesgo —la Amazonia y las zonas alto-andinas— son también las de menor conectividad. Según el Instituto Nacional de Estadística e Informática (INEI, 2022), solo el 30% de los hogares rurales del Perú tiene acceso a internet, frente al 78% de los hogares urbanos. En este contexto,

## ***Artículo original***

---

soluciones puramente digitales tienen un alcance estructuralmente limitado si no van acompañadas de políticas de infraestructura.

### **6.3 Participación comunitaria insuficiente**

Varios de los proyectos revisados fueron diseñados con escasa o nula participación de las comunidades hablantes en las fases de concepción y diseño. Esta comisión tiene consecuencias tanto éticas como prácticas: éticas, porque reproduce lógicas de intervención externa que no reconocen la agencia de los pueblos indígenas; prácticas, porque las herramientas diseñadas sin participación comunitaria tienden a no responder a las necesidades reales de los usuarios, a no adaptarse a sus contextos culturales y a no ser adoptadas de forma sostenida. La literatura es consistente en este punto: los proyectos con mayor impacto a largo plazo son aquellos en los que la comunidad no es objeto sino sujeto del proceso (Leonard & Haynes, 2010; Rice, 2011).

### **6.4 Riesgo de homogeneización lingüística**

Los modelos de IA tienden a privilegiar las variedades de mayor presencia digital y mayor número de hablantes, lo que puede generar una presión estandarizadora sobre lenguas pluricéntricas como el quechua que cuenta con numerosas variedades dialectales con diferencias fonológicas, léxicas y morfológicas significativas o el aimara. Esta estandarización tecnológica puede operar como una forma inadvertida de glotofagia interna: al promover una variante como la "correcta" para los sistemas de IA, se marginaliza implícitamente a los hablantes de otras variedades, y replica con ello, en el plano digital, las jerarquías lingüísticas preexistentes.

### **6.5 Sostenibilidad limitada**

La mayoría de los proyectos revisados depende de financiamiento temporal: fondos de cooperación internacional, subsidios estatales puntuales o inversión privada de corto plazo sin mecanismos claros de continuidad. Esta fragilidad financiera es especialmente crítica en el caso de las iniciativas de documentación: si un proyecto de grabación de voz o de producción de materiales educativos se interrumpe antes de completar su ciclo, el trabajo realizado puede quedar inaccesible o inutilizable para las comunidades. La sostenibilidad requiere no solo financiamiento estable, sino también modelos de transferencia de capacidades que permitan a las propias comunidades continuar y expandir los proyectos con autonomía creciente.

### **6.6 Desigualdad educativa**

La formación de docentes bilingües interculturales sigue siendo insuficiente en el Perú. El sistema de Educación Intercultural Bilingüe (EIB) enfrenta déficits de materiales pedagógicos actualizados, de formadores y

## ***Artículo original***

---

de políticas de contratación que aseguren la presencia de docentes hablantes de lenguas originarias en las comunidades que las necesitan. Sin esta base educativa, las herramientas tecnológicas carecen del sustrato humano necesario para integrarse significativamente en los procesos de transmisión lingüística intergeneracional.

### **7. Discusión**

Los resultados de esta revisión permiten articular tres conclusiones analíticas principales: Las tecnologías del lenguaje son necesarias pero no suficientes, hay riesgo de reproducir sesgos y jerarquías del mundo en el ecosistema digital y el proyecto Chana cumple, en nuestro análisis, con los criterios más prometedores. Las tres ideas se exponen y desarrollan a continuación en diálogo con el marco teórico.

En primer lugar, la IA y las tecnologías del lenguaje son condiciones necesarias pero no suficientes para la revitalización lingüística. Ninguno de los proyectos revisados, incluyendo los más sólidos metodológicamente, como Huqariq o Chana, puede por sí solo revertir el declive de una lengua si no existe una base demográfica de hablantes activos y un entorno sociocultural que favorezca su transmisión intergeneracional. Esto es coherente con el modelo de Fishman (1991): la tecnología puede amplificar y apoyar la revitalización, pero no puede sustituir la transmisión en el hogar y en la comunidad.

En segundo lugar, el riesgo de reproducción de jerarquías lingüísticas dentro del propio ecosistema de IA para lenguas indígenas es real y debe ser atendido proactivamente. Al priorizar las variedades con más datos disponibles, que tienden a ser las urbanas, las más estandarizadas y las de mayor número de hablantes, los sistemas de IA pueden operar como agentes involuntarios de uniformización, en contradicción con el objetivo declarado de preservar la diversidad lingüística. Este riesgo requiere políticas explícitas de representación dialectal en la construcción de corpus y en el diseño de modelos.

En tercer lugar, el enfoque más prometedor identificado en esta revisión es el que combina rigor tecnológico con colaboración no colonial y sostenibilidad comunitaria. La Estación Científica Chana es el ejemplo más desarrollado de este enfoque en el contexto peruano, aunque su escala sigue siendo limitada en relación con la magnitud del problema. La generalización de este modelo, que requiere infraestructura física en los territorios, formación de investigadores locales y marcos de gobernanza del conocimiento controlados por las comunidades, es un desafío de política pública que trasciende la capacidad individual de ninguna institución académica.

Finalmente, esta revisión tiene limitaciones que deben reconocerse. La información disponible sobre varios de los proyectos proviene principalmente de fuentes institucionales, lo que puede introducir un sesgo de presentación positiva. Los datos de impacto efectivo como el número de usuarios activos, cambios en actitudes lingüísticas, incremento en el uso cotidiano de las lenguas, son escasos o no están disponibles públicamente para

## ***Artículo original***

---

la mayoría de las iniciativas. Investigaciones futuras deberían orientarse a evaluaciones de impacto con metodologías participativas, incluyendo la perspectiva de los propios hablantes sobre la utilidad y pertinencia de estas herramientas. Esos datos permitirían obtener evaluaciones más detalladas y mostrar alcances con mayor claridad para los interesados.

### **8. Conclusiones**

Este artículo ha revisado crítica y sistemáticamente ocho iniciativas tecnológicas orientadas a la conservación de lenguas originarias en el Perú y ha buscado enmarcarlas en la literatura sobre revitalización lingüística y tecnologías del lenguaje para lenguas de bajos recursos. Los hallazgos permiten formular las siguientes conclusiones:

Primera: el Perú cuenta con un ecosistema emergente y diverso de iniciativas tecnológicas para la conservación lingüística, que incluye proyectos de IA generativa, corpus de voz, aplicaciones de traducción, herramientas de atención ciudadana, iniciativas mediáticas y centros de investigación. Esta diversidad es un activo valioso que no tiene equivalente en la mayoría de países de la región. En esta línea, este ecosistema debe seguir aumentando y aprendiendo de las propuestas que los anteceden.

Segunda: ninguna de las iniciativas revisadas es suficiente, por sí sola, para revertir el desplazamiento de las lenguas originarias. Su eficacia está condicionada por factores estructurales (brecha digital, falta de corpus, sostenibilidad financiera, insuficiente participación comunitaria) que requieren respuestas de política pública de largo plazo, no sólo intervenciones tecnológicas puntuales. La tecnología es una excelente herramienta para atender el tema de las lenguas originarias, y por ello, las políticas públicas deben apoyar dicha ruta.

Tercera: el riesgo de homogeneización lingüística inducida por la IA es una amenaza real y sub documentada, que debe integrarse en el diseño de cualquier herramienta tecnológica orientada a lenguas pluricéntricas. La preservación de la diversidad dialectal debe ser un criterio explícito, no un efecto esperado, de las iniciativas tecnológicas. Este punto solicita la participación de especialistas en el tema de lenguas y la variación.

Cuarta: el modelo más prometedor para la revitalización lingüística apoyada en tecnología es aquel que combina rigor científico, colaboración no colonial con las comunidades hablantes, sostenibilidad comunitaria y enfoque de derechos. La Estación Científica Chana representa la expresión más coherente de este modelo en el contexto peruano actual.

Como agenda de investigación futura, se identifican tres prioridades: (1) realizar evaluaciones de impacto participativas sobre las iniciativas existentes, con inclusión activa de los hablantes como evaluadores; (2) desarrollar marcos éticos y jurídicos de gobernanza de datos lingüísticos que reconozcan los derechos de

## Artículo original

---

propiedad intelectual colectiva de las comunidades indígenas; y (3) explorar modelos de sostenibilidad financiera que reduzcan la dependencia de ciclos cortos de financiamiento externo mediante la transferencia de capacidades y la generación de economía local en torno a la producción de contenidos en lenguas originarias.

## Referencias

- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19-32. <https://doi.org/10.1080/1364557032000119616>
- Bird, S. (2020). Decolonising speech and language technology. *Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020)*, 3504-3519. <https://doi.org/10.18653/v1/2020.coling-main.313>
- Fernández, C. (2024, marzo). Entrevista sobre el proyecto Illariy [Video]. Noticias NCC. <https://www.noticiasncc.com/illariy>
- Fishman, J. A. (1991). Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages. *Multilingual Matters*.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3), 61-83. <https://doi.org/10.1017/S0140525X0999152X>
- Himmelman, N. P. (1998). Documentary and descriptive linguistics. *Linguistics*, 36(1), 161-195. <https://doi.org/10.1515/ling.1998.36.1.161>
- Instituto Nacional de Estadística e Informática [INEI]. (2022). Encuesta nacional de hogares sobre condiciones de vida y pobreza 2022. INEI.
- Joshi, P. R., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)*, 6282-6293. <https://doi.org/10.18653/v1/2020.acl-main.560>
- Kornai, A. (2013). Digital language death. *PLoS ONE*, 8(10), e77056. <https://doi.org/10.1371/journal.pone.0077056>
- Leonard, W. Y., & Haynes, E. (2010). Making "collaboration" collaborative: An examination of perspectives that frame linguistic field research. *Language Documentation & Conservation*, 4, 268-293.
- Ministerio de Cultura del Perú. (2023). Lenguas indígenas u originarias del Perú: Base de datos oficial. <https://bdpi.cultura.gob.pe>
- Ministerio de Cultura del Perú. (2024). Uyariy: Aplicativo para la atención intercultural en servicios públicos [Aplicación móvil]. <https://www.gob.pe/cultura/uyariy>
- Rice, K. (2011). Documentary linguistics and community relations. *Language Documentation & Conservation*, 5, 187-207.
- UNESCO. (2003). Language vitality and endangerment. UNESCO Ad Hoc Expert Group on Endangered Languages.
- UNESCO. (2022). International decade of indigenous languages 2022-2032: Global action plan. UNESCO Publishing.

## **Artículo original**

---

Zevallos, R., Camacho, R., Melgarejo, J., Rodríguez, P., Bel, N., & Aguirre, C. (2022). Huqariq: A large-scale speech corpus for Peruvian indigenous languages. Proceedings of the 13th Language Resources and Evaluation Conference (LREC 2022), 1775-1784. <https://aclanthology.org/2022.lrec-1.189>

### **Contribución de los autores**

Víctor Gonzalo Ramírez Herrera se ha hecho cargo del diseño, estructura y recolección de la información para la redacción de este artículo.

### **Reconocimientos**

Ninguno

### **Financiamiento**

La investigación se realizó sin financiamiento.

### **Conflicto de intereses**

No hay conflicto de intereses.

### **Trayectoria académica de los autores**

Licenciado en Lingüística por la PUCP y magíster en Filosofía por la misma universidad. En el primer campo, investiga temas de Lingüística Andina, Lingüística formal y Psicolingüística; en este marco, desarrolló una investigación sobre la percepción de sonidos vocálicos en el quechua huanca. En el segundo, trabajo en Epistemología y Filosofía de la Ciencia, área en la que realizó una investigación crítica sobre la inconmensurabilidad propuesta por Kuhn. Asimismo, cuento con producción en las áreas de poesía y teatro. Actualmente curso el doctorado en Lingüística Andina en el PEA-PUCP, con una investigación sobre “la lengua general de los Incas”