



Uso de la transcripción fonética para lograr que avatares de inteligencia artificial pronuncien discurso en lengua quechua: caso Illariy

Using phonetic transcription to get artificial intelligence avatars to pronounce speech in Quechua: the Illariy case

Uso de transcrição fonética para fazer avatares de inteligência artificial pronunciarem fala na língua Quechua: caso Illariy

Óscar Huamán-Águila

Universidad Nacional Mayor de San Marcos, Lima, Perú
oscar.huaman3@unmsm.edu.pe
<https://orcid.org/0000-0003-2810-7611>

Carlos Fernandez-Garcia

Universidad Nacional Mayor de San Marcos, Lima, Perú
carlosenrique.fernandez@unmsm.edu.pe
<https://orcid.org/0000-0003-4864-9768>

Carlos Gonzales-García

Universidad Nacional Mayor de San Marcos, Lima, Perú
carlos.gonzales1@unmsm.edu.pe
<https://orcid.org/0000-0001-9408-4329>

Resumen

La inteligencia artificial (IA) ha generado noticia este año. En su quinta temporada, el noticiario Letras TV Willakun de la Facultad de Letras y Ciencias Humanas de la Universidad Nacional Mayor de San Marcos es conducido por Illariy (vocablo quechua que significa “Amanecer”), un avatar generado por inteligencia artificial, el único en su género que habla una lengua originaria. Este hecho nos lleva a plantearnos: ¿Cómo influye la transcripción fonética en la capacidad de la inteligencia artificial para generar discursos coherentes en lengua quechua? La investigación tiene como objetivo describir el flujo de trabajo concentrado en cómo la transcripción fonética puede lograr que la inteligencia artificial articule un discurso en lengua quechua. La metodología de la investigación tiene un enfoque cualitativo, basado en la observación y el análisis del flujo de trabajo para generar que los avatares pronuncien un discurso en quechua. Entre los resultados podemos observar la existencia de una pronunciación diferenciada entre el español y el quechua, limitaciones en la articulación de algunas consonantes de la lengua quechua y que resultan irreproducible en contextos tecnofonológicos, así como el uso de pausas y comas para la producción de una oralidad quechua en el Software D-ID.

Palabras clave: escritura fonética; quechua; avatar; inteligencia artificial; Illariy.

Abstract

Artificial Intelligence (AI) has been in the news this year. In its fifth season, the news program *Letras TV Willakun* of the Faculty of Letters and Human Sciences of the Universidad Nacional Mayor de San Marcos is hosted by Illariy (a Quechua word meaning “Dawn”), an avatar generated by Artificial Intelligence, the only one of its kind that speaks a native language. This fact leads us to pose the question: How does phonetic transcription influence the ability of Artificial Intelligence to generate coherent speech in the Quechua language? The aim of this research is to describe the workflow focusing on how phonetic transcription can enable Artificial Intelligence to articulate a speech in Quechua. The research methodology has a qualitative approach, based on the observation and analysis of the workflow to generate avatars to deliver a speech in Quechua. Among the results we can observe the existence of a differentiated pronunciation between Spanish and Quechua, limitations in the articulation of some consonants of the Quechua language that are irreproducible in techno-phonological contexts, as well as the use of pauses and commas for the production of a Quechua orality in the D-ID Software.

Keywords: phonetic writing; Quechua; avatar; artificial intelligence; Illariy.

Resumo

A Inteligência Artificial (IA) ganhou as manchetes este ano. Em sua quinta temporada, o programa de notícias *Letras TV Willakun*, da Faculdade de Letras e Ciências Humanas da Universidad Nacional Mayor de San Marcos, é apresentado por Illariy (palavra quíchua que significa “Amanhecer”), um avatar gerado pela Inteligência Artificial, o único do gênero que fala um idioma nativo. Esse fato nos leva a fazer a seguinte pergunta: Como a transcrição fonética influencia a capacidade da Inteligência Artificial de gerar uma fala coerente no idioma quíchua? A pesquisa tem como objetivo descrever o fluxo de trabalho de como a transcrição fonética pode ajudar a Inteligência Artificial a articular um discurso em quíchua. A metodologia da pesquisa tem uma abordagem qualitativa, baseada na observação e análise do fluxo de trabalho para gerar os avatares para pronunciar um discurso em quíchua. Entre os resultados, podemos observar a existência de uma pronúncia diferenciada entre o espanhol e o quíchua, limitações na articulação de algumas consoantes da língua quíchua que são irreproduzíveis em contextos tecnofonológicos, bem como o uso de pausas e vírgulas para a produção de uma oralidade quíchua no software D-ID.

Palavras-chave: escrita fonética; quíchua; avatar; inteligência artificial; Illariy.

Recibido: 17/01/2024

Aceptado: 14/07/2024

Publicado: 30/12/2024

1. Introducción

La creación de Illariy, un avatar generado por inteligencia artificial (IA) que presenta noticias en quechua, busca fortalecer la identidad cultural de los pueblos indígenas. Este innovador proyecto de la Universidad Nacional Mayor de San Marcos utiliza transcripción fonética para que la IA articule discursos en quechua, similar a tecnologías que recrean el habla de personajes de TV (Charles *et al.*, 2016). En el corazón de esta tecnología, se encuentra el Procesamiento del Lenguaje Natural (PLN), que permite a las computadoras entender y generar lenguaje humano (Morie *et al.*, 2012). Diversos estudios han contribuido a este campo, incluyendo sistemas de agentes para generar voz (Tebbi y Hamadouche, 2022), síntesis de voz multilingüe (Seong *et al.*, 2021), el conjunto de datos L2-ARCTIC (Zhao *et al.*, 2018), mejoras en reconocimiento de voz para árabe egipcio (Romanenko y Mendelev, 2016), y un transcriptor fonético para euskera (Barroso *et al.*, 2015). Ghazizadeh y Zhu (2020) han proporcionado una valiosa revisión del estado actual del PLN, subrayando la importancia de estos avances tecnológicos.

Este proyecto es particularmente relevante en el contexto peruano, donde existen 48 lenguas originarias. Sin embargo, estas lenguas enfrentan una situación de diglosia, siendo marginadas por el castellano (Carbajal *et al.*, 2018). La gravedad de esta situación se refleja en el hecho de que muchas de estas lenguas están en peligro de extinción (Unesco, 2010). El Ministerio de Cultura (s.f.) reconoce la diversidad lingüística del país, identificando 19 familias lingüísticas y 55 pueblos indígenas. Cavero (2012) ha señalado razones específicas para la extinción del quechua Ayacucho, incluyendo el avance

del castellano y la falta de políticas lingüísticas adecuadas. Para comprender mejor la magnitud de este desafío, García y Mere (2020) proporcionan datos detallados sobre los hablantes de lenguas originarias en Perú. En este contexto, Bruno-Seminario *et al.* (2020) enfatizan la importancia de preservar estas lenguas no solo como patrimonio cultural, sino también como medio para fomentar la coexistencia pacífica en una sociedad diversa.

Lamentablemente, la discriminación hacia las lenguas originarias no se limita al ámbito social, sino que se extiende también a los entornos tecnológicos. Las plataformas de IA han sido acusadas de perpetuar estereotipos y sesgos (Ibañez, 2023), lo que plantea serios desafíos éticos y prácticos. Schaufler (2021) advierte sobre la vulnerabilidad lingüística desde la perspectiva de Judith Butler, destacando cómo el lenguaje puede ser una herramienta de poder y exclusión. En el mundo virtual, Agudelo (2020) analiza cómo los avatares reflejan y refuerzan modelos culturales establecidos, lo que puede exacerbar la marginación de las lenguas minoritarias. Plataformas como D-iD, Murf.ai, Lalamud y Synthesia son mencionadas como ejemplos de esta discriminación sistemática, evidenciando la necesidad de un enfoque más inclusivo en el desarrollo de tecnologías lingüísticas.

Frente a estos desafíos, la investigación se centra en cómo la transcripción fonética puede influir positivamente en la capacidad de la IA para generar discursos coherentes en quechua. El objetivo principal es describir un flujo de trabajo que permita a la IA articular un discurso en esta lengua originaria, utilizando la tecnología no como una barrera, sino como una herramienta poderosa para preservar y revitalizar la herencia cultural e idiomática de los pueblos indígenas del Perú. Este enfoque innovador promete no solo avanzar en el campo de la tecnología lingüística, sino también contribuir significativamente a la preservación y promoción de las lenguas originarias en la era digital.

2. Marco teórico-conceptual

2.1. Transcripción fonética del español al quechua

El sistema de transcripción fonética del español al quechua emplea símbolos gráficos, producto de un análisis fonético acústico, que permiten una categorización de estímulos (Bagudanch, 2013), los mismos que representan sonidos de cada articulación, según los principios ortográficos que rigen en la otra lengua (Bruña, 2000; Coloma, 2012), el quechua. Este repertorio de códigos fonéticos utilizados responde a sistemas léxicos referidos a la pronunciación (Ríos, 2002).

Rivera (2004) sostiene que es difícil proponer un solo patrón de transcripción fonética sin causar discrepancias por razones históricas y etimológicas, como ocurre con el quechua sureño y el quechua central. Estos símbolos se crean tomando en cuenta la pronunciación y las diversas variaciones geográficas, regionales y dialectales (Rivera, 2004), que pueden tener un mismo fonema o un mismo alófono (Coloma, 2012).

Es menester alcanzar un alto grado de precisión en las transcripciones fonéticas a partir del análisis de los corpus (Perea, 2023), que, incluso, deben considerar que el quechua Ayacucho-Chanka es trivocálico (a, i, u) y con un acento fijo en la penúltima sílaba, con patrones acentuales “en palabras polisilábicas no próxítonas” (Castro, 2008) y que contienen vocablos que ofrecen una doble transcripción y que poseen más de una pronunciación correcta (Rivera, 2004).

2.1.1. Diferencias entre una transcripción fonética y fonológica

Llisterri (2023) diferencia los procesos entre la transcripción fonética y la transcripción fonológica, al señalar que el primero corresponde a cada representación fonológica y la segunda escribe la representación ortográfica correspondiente. La transcripción fonética es la representación gráfica de los sonidos lingüísticos que se utilizan al hablar una lengua mediante un alfabeto (Alba, 2014). En esta representación, se busca formar segmentos fónicos que reproduzcan palabras o sílabas para superar peculiaridades alfabéticas empleadas en cada lengua escrita y lenguas sin tradición escrita. Suele usar más símbolos que una transcripción fonológica, dado que emplea signos distintos para diferentes alófonos que tienen los fonemas de un idioma. En cambio, una transcripción fonológica se basa en la representación de fonemas que se usan en dicha lengua (Coloma, 2012).

2.1.2. Sistema de transcripción fonética a una lengua nativa

Analizar un sistema de transcripción fonética de una lengua nativa genera la *forma fónica* que representa la pronunciación de elementos léxicos de una lengua de origen a otra de destino y que compromete el estudio de los siguientes componentes: fonética (articulatoria, acústica, perceptiva o auditiva), adaptación de la nativización y adaptación a la pronunciación.

a. Adaptación a la nativización (reglas entre el español y el quechua)

Polyákova y Bonafonte (2008) conceptualizan la nativización como la adaptación de la pronunciación en el idioma de origen a la pronunciación más cercana y apropiada en el idioma de destino, mediante el uso de tablas adecuadas con el fin de alcanzar “un alto nivel de inteligibilidad del habla sintetizada” (Polyákova y Bonafonte, 2008). Es crucial resaltar que el habla nativa y el habla no nativa difieren significativamente en muchos aspectos, por lo que se recomienda que cada vocablo tenga su propia búsqueda de transcripción fonética, palabra por palabra. Entre las reglas específicas de la transcripción nativizada, se encuentra la evasión de fonemas medio-nativos o mal pronunciados. “Si la palabra no se encuentra en el diccionario de la lengua origen, se procesa con el sistema automático de transcripción fonética de esa lengua y luego, igual que antes, se aplican las reglas de nativización” (Polyákova y Bonafonte, 2008, p. 208).

La pronunciación es uno de los elementos que caracterizan las variedades de registro, geográfica y social de la lengua, por lo que influye significativamente en las habilidades sociolingüísticas (Guerrero, 2007). Pronunciar es emitir en voz alta los fonemas de una variedad lingüística (Andreani, 2018).

2.1.3. Alfabeto fonético del quechua Ayacucho-Chanka

La variedad quechua ayacuchano se habla “[...] en los departamentos peruanos de Huancavelica y Ayacucho y la mitad occidental de Apurímac, con bastante homogeneidad” (Torero, 2007, p. 32). Cuenta con 16 consonantes (a, ch, h, i, k, l, ll, m, n, ñ, p, q, r, s, t, u, w, y) y con tres vocales (i, u, a).

Para Majerhwa (2008), el quechua de la familia Ayacucho-Chanka es inteligible respecto al resto de variedades peruanas. Su apuesta para estandarizar la escritura se basa en lo inteligible que resulta el dialecto Ayacucho-Chanka, comparándola con el resto de las variedades peruanas de esta lengua

originaria. El quechua Ayacucho-Chanka es el más simple y el más conservador respecto a sus pares y el que fonéticamente se aproxima más a su fonología (Majerhua, 2008).

Majerhua (2008) sostiene que es necesario estandarizar la escritura del quechua mediante técnicas de escritura que buscan avanzar hacia un alfabeto común sobre el quechua Ayacucho-Chanka. El alfabeto quechua Ayacucho-Chanka es autónomo basado en las propias normas del quechua y no en las normas del castellano como lengua dominante (Andreani, 2018).

Para desarrollar la transcripción fonética del castellano al quechua es necesario prestar atención a la fonética articulatoria, acústica, perceptiva o auditiva para investigar los procesos de comunicación lingüística y no solo a la pronunciación, tomando en cuenta que los alfabetos no sirven exclusivamente para pronunciar; solo sirven para leer y escribir (Andreani, 2018).

La descripción fonética de cada sonido, desde el punto de vista y modo de articulación nos permite estudiar estas manifestaciones lingüísticas desde la perspectiva del modo de articulación, análisis de las ondas acústicas y espectros de sonido para comprender los procesos de percepción y audición de los sonidos, tal y como aconseja Fernández de Molina Ortés (2018).

2.2. La inteligencia artificial

Según Ocaña-Fernández *et al.* (2019), el proceso de replicar las habilidades de inteligencia del cerebro humano se conoce como inteligencia artificial (IA). Como rama de las ciencias de la computación, la inteligencia artificial se ocupa del diseño de sistemas inteligentes asociados con la inteligencia en el comportamiento humano. El término fue acuñado por John McCarthy (Benito, 2023).

Mariño y Primorac (2016) profundizan en el tema al afirmar que la IA ofrece una amplia gama de técnicas, métodos y herramientas para modelizar y resolver problemas del mundo real a partir de la experimentación y aplicación simulada del comportamiento humano. Según Pérez (2018), el avance tecnológico está descubriendo una nueva forma de inteligencia que se une a la inteligencia humana: la “inteligencia artificial”, que se basa en las “máquinas que aprenden”.

Los formatos basados en IA prometen mejorar significativamente la educación en todos los niveles. Los procesos de enseñanza-aprendizaje con IA incrementan cualitativamente el desarrollo de capacidades y competencias, al brindar al estudiante una personalización segura de sus avances, integrando diversas maneras de interacción humana y tecnologías (Ocaña-Fernández *et al.*, 2019). La enseñanza de idiomas extranjeros tampoco puede exonerarse del uso de tecnologías actuales (Escobar, 2021).

Actualmente, Google Traductor, Deja Vu, DeepL y otros traductores automáticos son herramientas poderosas que cuentan, desde hace tiempo, con inteligencia artificial. La popularidad cada vez mayor de la IA, particularmente en el campo de la lingüística y los traductores automáticos, ha llevado a la aparición de “nuevas maneras de programación como el aprendizaje autónomo y el aprendizaje basado en redes neuronales” (Restrepo, 2019, p. 29).

Los actuales procesos de Google Traductor, por ejemplo, aplican un sistema de aprendizaje neuronal como parte de su “entrenamiento” para que funcione de forma similar al cerebro humano

que les permite analizar probabilísticamente de traducir “correctamente” o “incorrectamente” con base en el uso común de una palabra (Restrepo, 2019).

En el caso de los sistemas con Inteligencia Artificial aplicados a la educación quedan representadas por un ‘agente pedagógico’ que son avatares que ‘habitan’ el sistema y se presentan en todas las intervenciones posibles; tanto debido a evaluaciones de tipo cognitivo, como emocionales.

La inteligencia artificial y el desarrollo de su modalidad generativa destaca por su inteligencia artificial y proceso creativo muy similar al tradicional humano (Benito, 2023). Se caracterizan por emplear órdenes, instrucciones o comandos que traducen textos para generar texto, código, imágenes, vídeo o música como es el caso *ChatGPT*, *Midjourney*, *Káiber* o *Sound Drawn*, plataformas de IA generativa. *D-iD* es una plataforma generativa también.

2.2.1. El avatar

Según la RAE (s.f.), *avatar* significa “representación gráfica de la identidad virtual de un usuario en entornos digitales”. El vocablo tiene su origen en el sánscrito y se traduce como “encarnación terrestre de una deidad” (RAE, s. f.). En el ámbito educativo, el avatar es un *agente pedagógico* que facilita aprendizajes en entornos virtuales y que puede estar implicado en diferentes tareas en el proceso de enseñanza-aprendizaje (Pérez *et al.*, 2023).

2.2.2. Plataformas de voice generators (Text-to-Speech)

El panorama de los generadores de voz de inteligencia artificial está en pleno ascenso, gracias a plataformas de creación de videos AI que permiten crear videos de alta calidad, a partir de texto simple, de manera fácil y rápida, como *D-iD Studio*, *Synthesia*, *Murf.ai*, *Listnr*, *Speechelo*, *Describe Overdub*, *WellSaid Labs*, *Lalamu Studio*, *Play.ht*, *Veed.io*, *Clipchamp*, *Speechgen.io*, *Lovo.ai*, *Clipchamp*, *Resemble.ai*, *Designs.ai*, *Resemble.ai* o *Heygen.com*.

2.2.3. D-iD: Vídeos con voces neuronales

La Oficina de Comunicaciones e Imagen Institucional y la Cátedra Quechua para la transcripción fonética del quechua emplearon *D-iD* por su facilidad de crear vídeos con voces neuronales extensas en más de 120 idiomas y variantes, gracias a sus funciones para crear audiovisuales con un solo fotograma llamada *Live Portrait* y para trabajar una voz para el texto o el audio que se conoce como *Speaking Portrait*.

D-iD data del 2017 y fue creada como la primera solución de identificación de imágenes faciales y que pertenece a *Creative Reality™ Studio*, impulsada por *Stable Diffusion* y *GPT-3* y produce videos en más de 100 idiomas. Protege la privacidad de los personajes públicos y evita el contenido ofensivo o sexual gracias a la tecnología de *Microsoft Azure Content Moderation* y *Amazon Rekognition* para detectar imágenes ofensivas o celebridades (Fraga, 2023).

D-iD emplea la tecnología de *LipNet*, un modelo que mapea una secuencia de fotogramas de video de longitud variable a texto (Valqui *et al.*, 2023). *LipNet* ofrece una lectura de labios, de nivel de oración, de extremo a extremo, que utiliza la extracción de características espaciotemporales y

la agregación temporal eficiente para decodificar con precisión el texto del movimiento de la boca de un hablante. Supera los enfoques tradicionales de lectura de labios y tiene una tasa de precisión del 95,2 % en el corpus GRID frente al 55 % de la línea de base de lectura de labios humanos experimentados (Assael *et al.*, 2016).

En *D-iD*, las expresiones de los avatares presentan expresiones faciales lo más naturales posible y se mueven de manera natural y realista, se pueden clonar mediante voces de una persona específica y almacenar en la nube mediante el uso del SSE cifrados con un estándar AES-256 de 256 bits.

2.2.4. Transcriptor o fonetizador

Las tecnologías están cada vez más involucradas en la enseñanza de habilidades lingüísticas (lectura, escritura, escucha y habla) y de pronunciación (Orta *et al.*, 2018). Un transcriptor o fonetizador integra la forma fónica-sonidos, división silábica y acentuación para representar la pronunciación de los elementos léxicos (Ríos, 2002).

Plataformas como *D-iD*, *Syntensia* o *Murf* requieren de una herramienta tecnológica como un transcriptor, fonetizador o conversor de texto a voz, empleando una técnica de concatenación de segmentos acústicos básicos llamados “difonemas” y que incluyen la transición de dos fonos consecutivos, formado por una transición del primer con el segundo fonema para la síntesis de la voz (Sánchez, 2016).

El *prompt* del conversor de texto a voz que se emplea son las informaciones que aparecen en el guion que será interpretado por el avatar, configurado previamente por idioma de preferencia, personaje o emoción. En el caso del avatar Illariy de *Letras TV Willakun*, el conversor de texto a voz emplea un guion con un corpus de voz en quechua Ayacucho-chanca —caracterizado por contener un guion en alfabeto trivocálico— y la base de datos para dicha lengua originaria. La base de datos de difonemas del idioma quechua de la familia lingüística Ayacucho-Chanka se construyó a partir de la información recopilada en ese momento.

3. Metodología

El enfoque es cualitativo. La metodología que seguiremos en este estudio buscará determinar mediante la observación y descripción de las diferentes unidades de análisis que realizaremos sobre la base de los guiones en español de *Letras TV Willakun* y que es la base para la traducción y transcripción fonética al quechua del presente texto. Para garantizar la calidad de los datos, es crucial que las transcripciones fonéticas al quechua Ayacucho-Chanka representen los sonidos de esa lengua originaria de manera sistemática y que se realicen de manera precisa y estandarizada.

Dado que la dependencia contextual de estas unidades es relevante para el presente estudio y que tienen como objetivo el observar y analizar el flujo de trabajo para que los avatares hablen en lengua quechua, para lo cual se pudo realizar de forma automática la identificación de nuestros parámetros de análisis. De esta manera, pudimos ver hasta qué punto dichos patrones se repetían con mayor frecuencia entre un subcorpus y otro.

3.1. Diseño del corpus de estudio

El corpus de estudio sobre el cual basamos el presente trabajo está compuesto de tres subcorpus: texto del guion de *Letras TV Willakun* (en español), representación ortográfica de entrada (texto en quechua Ayacucho-Chanka) y representación fonética de salida (transcripción fonética de salida en quechua Ayacucho-Chanka). Cada subcorpus ha sido diseñado con criterios relativos de unicidad como producto periodístico audiovisual bilingüe, en quechua con subtítulos en español, siguiendo los mismos criterios de composición: medio por el cual se difunde el mensaje, duración (aunque el texto traducido al quechua, suele ser más largo), lenguaje informativo y género textual periodístico.

Tabla 1

Corpus de estudio

CORPUS DE ESTUDIO		
Subcorpus de texto del guion de <i>Letras TV Willakun</i> (en español)	Subcorpus de representación ortográfica de entrada (texto en quechua Ayacucho-Chanka)	Subcorpus de representación fonética de salida (transcripción fonética de salida en quechua Ayacucho-Chanka)
14 guiones de texto informativo en español	14 guiones de texto informativo en quechua Ayacucho-Chanka	14 guiones de texto informativo de transcripción quechua Ayacucho-Chanka confrontados en el transcriptor del <i>D-iD</i>

Nota. La tabla corresponde al corpus y subcorpus de estudio

3.3.1. Selección de herramientas de análisis

Para la presente investigación, utilizaremos los guiones informativos de *Letras TV Willakun* (ediciones de la 1 a la 14 de la quinta temporada) como herramienta y técnicas de análisis por cada uno de los subcorpus, a fin de comprender las equivalencias entre el español, el quechua Ayacucho-Chanka y su correspondiente transcripción fonética, la misma que se integra como parte del procedimiento para permitirle a Illariy, avatar generado por Inteligencia Artificial hablar en idioma quechua. Para analizar el sonido recurriremos a la emisión de voz artificial del avatar en el programa *D-iD*.

3.3.2. Muestreo representativo

El tamaño y la representatividad del corpus son fundamentales para obtener conclusiones significativas, es por ello que proponemos como muestra los 14 guiones que forman parte de cada uno de los subcorpus de la quinta temporada, con el propósito de asegurar la representatividad muestral como objeto de estudio que se quiere analizar.

3.3.3. Revisión y validación

La presente investigación contará con la revisión y validación por parte de otros expertos en fonética o lingüística son esenciales para garantizar la precisión y profundidad de los resultados obtenidos.

4. Análisis

4.1. Flujo de trabajo

El flujo de trabajo se inicia con la cobertura total de las actividades de la Facultad de Letras y Ciencias Humanas de la Universidad Nacional Mayor de San Marcos. Seguidamente, se redactan las cuatro o cinco noticias más importantes en el ámbito universitario correspondiente a esa semana y se incluyen en el guion audiovisual. El guion se comparte en el *WhatsApp* del grupo de Comunicación e Imagen para que el responsable de la Cátedra Quechua realice la traducción a la lengua quechua. Una vez culminada la traducción, se procede a desarrollar la correspondiente transcripción fonética, a partir de las equivalencias establecidas entre las vocales y consonantes en lengua quechua, mediante un *prompt* que generamos desde el *D-iD*, configurando la voz de Martha, en “español” (Guatemala), donde se revisa acuciosamente la pronunciación que referirá el avatar. En caso la pronunciación no sea la acertada, se procederá a reemplazar por otro vocablo y otro hasta dar con la pronunciación precisa. En posproducción, se trabaja en *Photoshop*, primero, el *outfit* del avatar que se exporta a *png* con fondo verde; esta imagen se coloca en el *software D-iD* y en el *prompt* se coloca los textos de la transcripción fonética que corresponden al audio que es hablado por el avatar Illariy con el sincronismo pertinente, luego de configurar nuevamente en voz de Marta, en “español” (Guatemala). Una vez descargado el archivo de vídeo en *mp4* se sube al *After Effects* para su composición audiovisual, se colocan los fondos y en *Adobe Premiere* se compagina el resultado y se musicaliza. Finalmente, se colocan los textos de los titulares en quechua y la traducción al castellano (*vidifont*).

4.1.1. Sistematización de los corpus desarrollados a partir de la escritura fonética de los guiones del programa *Letras TV Willakun*

a. Uso de símbolos gráficos para representar las características fonéticas

Para generar una voz en la lengua quechua con Inteligencia Artificial se opta por el uso de la transcripción fonética. Esto comprende el proceso de adaptación de la escritura quechua al campo del repertorio gráfico de *D-iD*. Para lo cual se emplean los símbolos gráficos que reconoce el *prompt*. Este proceso es similar a lo que ocurre con los préstamos léxicos entre lenguas. Como señala Ccaccachahua (2022), “los préstamos léxicos son el resultado de las incorporaciones de una palabra a otra lengua, los cuales suelen encajar en la estructura gramatical de la lengua receptora” (p. 467). De manera análoga, la transcripción fonética para la IA busca adaptar los sonidos del quechua a un sistema que pueda ser reconocido y reproducido por la tecnología. Lo mencionado se puede percibir en la siguiente tabla.

La Inteligencia Artificial genera una voz inteligible con los símbolos alfabéticos representados en la columna “Repertorio fonético en el *D-iD*”, de la tabla 2. Con esta plantilla fonética quechua, generado durante el ensayo, se realiza la transcripción del micro noticiario de *Letras TV Willakun* de la Facultad de Letras y Ciencias Humanas de la Universidad Nacional Mayor de San Marcos. Pero no es suficiente este aspecto lingüístico para generar la voz quechua con Illariy, sino también se requiere otros aspectos más.

Tabla 2

Repertorio de símbolos

Alfabeto quechua Ayacucho-Chanka	Sonido	Repertorio fonético en el <i>D-iD</i>	Observaciones
a	/a/	/a/	
ch	/č/	/ch/	
h	/h/	/j, h/	
i	/i/	/i, e/	
k	/k/	/c, k/	
l	/l/	/l/	
ll	/ll/	/y/, /ll/	
m	/m/	/m/	
n	/n/	/n/	
ñ	/ñ/	/ñ/	
p	/p/	/p/	
q	/q/	/jj/	
r	/r/	/r/	
s	/s/	/s/	
t	/t/	/t/	
u	/u/	/u, o/	
w	/w/	/hu, w/	
y	/y/	/i, y/	

Nota. La tabla representa las equivalencias entre el alfabeto quechua Ayacucho-Chanka, el fonema y el repertorio fonético en el *D-iD*, elaborado por Óscar Huamán Águila.

Las grafías mostradas en la tabla 2 en la etapa a ser utilizados en la conversión de la traducción quechua a la plantilla fonética. Esto se puede percibir en la tabla 3.

Tabla 3

Ejemplo de repertorio de símbolos utilizados en un guion de Letras TV Willakun

Texto del guion de <i>Letras TV Willakun</i> en español	Representación ortográfica de entrada (texto en quechua)	Representación fonética de salida (transcripción fonética de salida en quechua)
Letras y los premios Copé tienen una larga historia en común.	Letraswan Cope llallinakuyqa unayñam kanku	Letrashuan Cope yayinacuijja unaiñam cancu.

Nota. Letras TV Willakun (17 de junio de 2023).

En la columna “Representación fonética de salida” (tabla 4), se encuentra un texto con transcripción fonética. Este texto es facilitado al *prompt* de *D-iD* para que el avatar pronuncie un discurso quechua. Esta producción de voz quechua es factible por el reconocimiento de las grafías alfabéticas de *D-iD*. Si no reconociera sería imposible crear una noticia quechua con Inteligencia Artificial.

En esta fase de ensayo, se generaron las grafías que se muestran en la columna titulada “Representación fonética de salida”. Desde luego, algunas de las grafías alfabéticas que utiliza la lengua quechua en la representación fonológica se mantienen y en las otras hubo cambios en su representación simbólica de los sonidos. El ejemplo que ilustra este cambio son las grafías [h] > [j], [ll] > [y], [q] > [jj], [w] > [hu], [k] > [c], respectivamente. Cabe mencionar que este cambio

de caracteres, durante la transcripción fonética de texto quechua, se da de manera general en todos los contextos en los que se encuentra. Excepto el símbolo gráfico [ll] no se mantiene de manera permanente, sino hay una alternancia con [y] en algunos contextos estructurales de las palabras quechuas. Cuando va al inicio de las palabras o cuando antecede al símbolo gráfico [jj], se mantiene de manera constante como 'll'; en cambio, cuando se encuentra entre vocales y cuando va después de los consonantes, pasa a ser [y]. Por último, la semivocal /w/ alterna entre [hu, w] en la transcripción. Al respecto, [w] se mantiene cuando va después de una constante; por ejemplo, en la expresión [Letraswuan].

Las vocales también experimentan cambios cuando se pasa a la transcripción fonética. Por ejemplo, la vocal [i] pasa a ser [e] y [u] a [o]. Este cambio ocurre de manera constante cuando se encuentra antes y después de la consonante [jj]. En la siguiente tabla se visualiza lo mencionado.

Tabla 4
Sistema de transcripción fonética de vocales

Representación ortográfica de entrada (texto en quechua)	Representación fonética de salida (transcripción fonética de salida en quechua)
[...]qillqaman risqanta rimarqun. Allin chiqap rimaymi karqun.	[...]jjelljjaman risjjanta rimarjjon. Ayin chejjap rimaimi carjjon.

Nota. Letras TV Willakun (10 de junio de 2023).

Como se puede percibir en las palabras transcritas [jjelljjaman] y [chejjap], que se encuentran en la columna izquierda, la vocal [i] ha pasado a ser [e]. Lo mismo ocurre con la vocal [u], ya que en las expresiones transcritas [rimarjjon] y [carjjon] no se muestra la vocal [u].

Este suceso explica que los robots conversores de texto a voz solo admiten vocales y consonantes que se encuentran en el *prompt*.

b. Pronunciación diferenciada entre el español y el quechua

Entre la lengua española y el quechua ayacuchano existen algunas diferencias en la pronunciación de consonantes.

Las características del alfabeto del quechua ayacuchano cuentan con 3 vocales, 13 consonantes y 2 semiconsonantes (w, y). En cambio, el español cuenta con 22 consonantes y con 5 vocales. En este contraste de sonido de las lenguas disímiles, se puede ver la particularidad de la pronunciación de la quechua Ayacucho-Chanka que tiene sonidos que no existen en el español. Por esta diferencia de pronunciación que existe, no se logra articular los sonidos como: ll, q, h, w.

c. Criterios de segmentación / conmutación

El otro proceso que se hace para la producción del discurso inteligible en lengua quechua por medio de Inteligencia Artificial es la segmentación de la palabra forma de sílaba. Lo mencionado se hace para que la pronunciación sea comprensible a nivel auditivo y semántico para el receptor quechuahablante. Cabe señalar que no todas las palabras pasan por este proceso de distanciación

de los componentes de las palabras. Hasta la fecha solo se encontró en la expresión como *allinyay* ‘mejorar, sobre salir’, en la versión transcripción fonética [ayinyai]. Veamos en la siguiente figura.

Figura 1

Segmentación de la palabra

del Siglo de Oropi,

cajjpim apacunjja.

Jjamcunapajj, ayin yai cachun.

Nota. Letras TV Willakun (10 de junio de 2023).

La palabra [ayinyai], que se encuentra en la última línea, se ha segmentado en dos [ayin yai]. Esto se hace para que la pronunciación en *D-iD* salga inteligible.

La otra segmentación de las expresiones quechuas son frases cortas. Este es el otro mecanismo que se hace para obtener una voz comprensible en el noticiario. Esta segmentación se hace en el *prompt* de *D-iD* con el seguimiento de secuencia de la articulación de voz del avatar de IA. Esto se muestra en la siguiente figura.

Figura 2

Segmentación de la oración en frases Willakun

Literatura yachaipi cajj Raúl Morales,

Enrique Toledo, Lourdes Orihuela,

Carlos Daniel Ventura,

Franco Murgam

Nota. Letras TV Willakun (10 de junio de 2023).

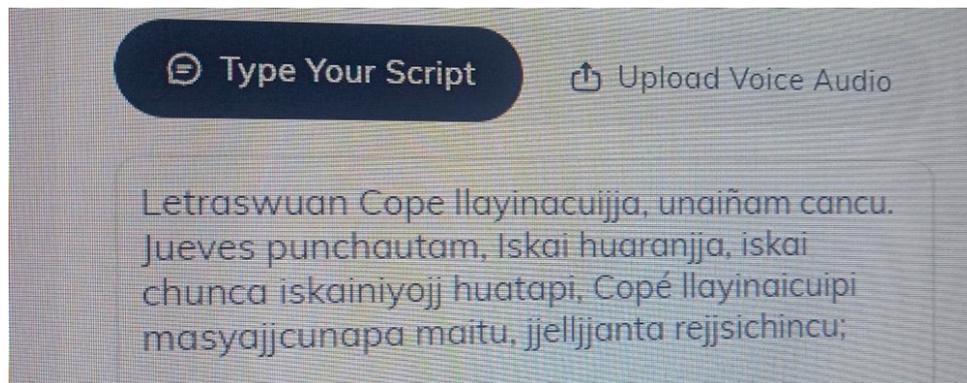
Las frases se encuentran en forma de versos. Esto le permite al avatar pronunciar con naturalidad y con una secuencialidad pertinente el discurso quechua. Sin esta segmentación en frases cortas no es posible obtener un discurso quechua.

En relación con el tema anterior, se encuentra la segmentación de la transcripción fonética quechua en bloques. Es decir, en el *prompt* de *D-iD*, solo se colocan veinte palabras quechuas. Las palabras que pasan la cantidad mencionada ya no son pronunciadas.

El otro proceso de segmentación tiene que ver con la separación de la oración con comas. Esto se hace igual que del anterior párrafo para que el discurso sea entendible para el receptor. También, para que la pronunciación sea un poco pausada.

Figura 3

Segmentación de oraciones con comas



Nota. Imagen tomada por Óscar Huamán Águila (2023).

En la primera línea se muestra una coma después de la palabra [llayinacuijja], esto no tiene justificación a nivel gramatical, puesto que no es aceptable separar el sustantivo del verbo en la gramática quechua. Este mismo hecho se muestra en los números [Iskai huaranjja, iskai (...)] de la segunda línea.

Por último, cabe señalar el uso de mayúsculas en algunas palabras de la transcripción fonética para la inteligencia artificial. Mediante este proceso se logra una pronunciación quechua. Las palabras que el avatar pronuncia en inglés se escriben en minúscula.

Figura 4

Uso de mayúscula

[J]anchis Literatura AMIRINDIAS

Tincuijja chunca jucniyojj

agosto kiyapim achca

Nota. Letras TV Willakun (17 de junio de 2023).

Esta parte del análisis y descripción realizada ha sido estructurada para poder lograr un discurso óptimo, tanto en la pronunciación y como en la articulación de un discurso mayor; desde luego, para que la comunicación sea comprensible para el receptor. Por este factor (cadencia y ritmo de la voz), la primera línea termina en preposición y no tanto por la alineación que se hace en un texto. Esta parte se hace en el momento de generar la voz en el *prompt* de *D-iD*. Lo mencionado es el otro proceso que se hace en la creación de noticiario en la lengua quechua.

5. Conclusiones

Este estudio describe un flujo de trabajo que emplea la transcripción fonética para que la Inteligencia Artificial (IA) articule un discurso coherente en quechua. Se implementó un sistema de análisis auditivo para evaluar la precisión de la escritura fonética del quechua en D-Id, con el fin de identificar errores en la pronunciación de la variante Ayacucho-Chanka. La adaptación del quechua a la IA requiere un repertorio fonético especializado en el software D-iD, estableciendo correspondencias entre los fonemas quechuas y los símbolos gráficos reconocidos por el programa. Este proceso es fundamental para generar una voz quechua inteligible, permitiendo que la IA pronuncie adecuadamente el texto. Se enfatiza en las diferencias fonéticas entre el español y el quechua, particularmente en consonantes y vocales sin equivalentes directos en español. La segmentación de palabras y frases en la transcripción fonética facilita una pronunciación comprensible y natural. Esto incluye la división de palabras complejas en sílabas más simples y la estructuración de oraciones en frases breves. Estas técnicas optimizan la fluidez y comprensión del discurso quechua generado por la IA, tanto a nivel auditivo como semántico. El proceso implica una revisión meticulosa y ajustes continuos para garantizar la precisión fonética. Esto puede incluir la sustitución de palabras o modificaciones en la transcripción para mejorar la claridad y naturalidad del discurso. La investigación resalta la relevancia de un sistema de análisis auditivo para detectar y corregir errores de pronunciación, asegurando que la IA articule un discurso coherente y auténtico en quechua.

Este trabajo evidencia la relevancia de los estudios interdisciplinarios que integran lingüística, tecnología y preservación cultural. Para los lingüistas y las humanidades, esta investigación abre nuevas vías de exploración en la intersección entre lenguas originarias e inteligencia artificial. Es imperativo continuar esta línea de investigación, no solo para perfeccionar las capacidades tecnológicas, sino también para contribuir a la preservación y revitalización de lenguas en riesgo de extinción como el quechua. La colaboración entre lingüistas, expertos en IA y profesionales de la comunicación es esencial para abordar los desafíos que surgen al adaptar tecnologías avanzadas a las particularidades de las lenguas originarias. Futuras investigaciones podrían extender estas técnicas a otras lenguas indígenas, examinar el impacto de estas tecnologías en la percepción y uso de las lenguas originarias, y desarrollar herramientas pedagógicas basadas en IA para la enseñanza y difusión de estas lenguas. Este enfoque no solo enriquece el campo de la lingüística computacional, sino que también contribuye significativamente a la preservación del patrimonio lingüístico y cultural de los pueblos originarios.

Referencias

- Agudelo, C. (2020). *La sociedad de los avatares: Videojuegos, representación y discriminación* [Tesis de maestría, Pontificia Universidad Javeriana]. <https://doi.org/10.11144/Javeriana.10554.13464>
- Alba, O. (2014). *Manual de fonética Hispánica* (V Edición). Editorial Plaza Mayor. <https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=1008&context=books>
- Andreani, H. (2018, 6 de diciembre). Para qué un alfabeto quichua (en quichua). *Trazos Universitarios*. <http://revistatrazos.ucse.edu.ar/index.php/2018/12/06/alfabeto-quichua-quichua/>
- Assael, Y., Shillingford, B., Whiteson, S., y de Freitas, N. (2016). LipNet: End-to-End Sentence-level Lipreading. *arXiv*. <http://arxiv.org/abs/1611.01599>
- Bagudanch, A. (2013). La transcripción fonética en estudios dialectales: Propuestas en el caso del yeísmo. *Revista de Filología Española*, 93(1), 165-192.. <https://doi.org/10.3989/rfe.2013.06>
- Barroso, N., Lopez-de Ipina, K. y Calvo, P. M. (2015). An automatic and adaptive phonetic transcriber for the Basque language. En *2015 4th International Work Conference on Bioinspired Intelligence (IWOBI)* (pp. 163-168). San Sebastián, España.
- Benito, Z. (2023). *Aproximación a la nueva revolución tecnológica y creativa: Las inteligencias artificiales generativas* [Tesis de licenciatura, Universitat Jaume I]. <https://repositori.uji.es/xmlui/handle/10234/203214>
- Bruno-Seminario, Á., Córdova-ChuquiHuanga, A., y Santos-Arriola, J. (2020). Las lenguas originarias del Perú. Un análisis de su estado desde la multiculturalidad. *Pro Hominum*, 2(3), 92-104.
- Bruña, M. (2000). El primer diccionario francés-español con transcripción fonética (Cormon, 1800). En *La philologie française à la croisée de l'an 2000: Panorama linguistique et littéraire* (Vol. 2, pp. 165-178).
- Ccaccachahua, B. (2022). Préstamos léxicos en el campo de las tecnologías modernas del castellano en el quechua chanca del anexo de Puncuhuacca. *Lengua y Sociedad*, 21(2), 449-473. <https://doi.org/10.15381/lengsoc.v21i2.22799>
- Carbajal, V., García, F., Huamancayo, E., Mori, M., Rodríguez, M. y Verástegui, N. (2018). *Lenguas originarias del Perú*. Ministerio de Educación. https://centroderecursos.cultura.pe/sites/default/files/rb/pdf/Lenguas%20Originarias%20del%20Peru%20%282018%29_7_MB.pdf
- Castro, S. (2008). *La escritura fonémica del quechua ayacuchano* [Tesis de maestría, Facultad Latinoamericana de Ciencias Sociales]. <https://repositoriointerculturalidad.ec/jspui/bitstream/123456789/3367/2/TFLACSO-2008SMC.pdf>
- Cavero, M. (2012). *El quechua Ayacucho-Chanka en proceso de extinción: Un caso (provincia de Huamanga, Ayacucho)* [Tesis de maestría, Universidad Nacional Mayor de San Marcos]. https://cybertesis.unmsm.edu.pe/bitstream/handle/20.500.12672/14811/Cavero_am-Resumen.pdf?sequence=1&isAllowed=y

- Charles, J., Magee, D., y Hogg, D. (2016). Virtual immortality: Reanimating characters from TV shows. En G. Hua y H. Jégou (Eds.), *Computer Vision – ECCV 2016 Workshops* (pp. 879–886). Springer International Publishing.
- Coloma, G. (2012). Caracterización fonética de las variedades regionales del español y propuesta de transcripción simplificada. *Revista de Filología Románica*, 28(0), 11-27. https://doi.org/10.5209/rev_RFRM.2011.v28.37217
- Escobar, J. (2021). La Inteligencia Artificial y la Enseñanza de lenguas: Una aproximación al tema. *Decires*, 21(25), 29-44. <https://doi.org/10.22201/cepe.14059134e.2021.21.25.3>
- Fernández de Molina Ortés, E. (2018). La transcripción fonética en Twitter: Tradición y modernidad en la educación. En REDINE (Ed.), *Innovative strategies for Higher Education in Spain* (pp. 141-150). Adaya Press. <https://doi.org/10.58909/ad18757372>
- Fraga, E. A. (2023, febrero 13). *Gil Perry: «Los avatares sintéticos facilitarán experiencias digitales más personales y humanas con los clientes»*. ConVerCom. <https://convercom.info/comunidades/disrupcion/gil-perry-los-avatares-sinteticos-facilitaran-experiencias-digitales-mas-personales-humanas-los-clientes/>
- García, G. y Mere, M.(2020). Las lenguas indígenas u originarias del Perú en el internet. *LinguaTec*, 5(1), 109-136. <https://doi.org/10.35819/linguatec.v5.n1.4112>
- Ghazizadeh, E. y Zhu, P. (2021). A systematic literature review of natural language processing: Current state, challenges and risks. En K. Arai, S. Kapoor y R. Bhatia (Eds.), *Proceedings of the Future Technologies Conference (FTC) 2020, Volume 1. FTC 2020. Advances in Intelligent Systems and Computing* (Vol. 1288). Springer, Cham. https://doi.org/10.1007/978-3-030-63128-4_49
- Guerrero, A. (2007). ¿Qué es la pronunciación? *Revista Electrónica de Didáctica*, 9, 1-16.
- Ibañez, M. (2023, abril 11). El algoritmo es machista y racista | La inteligencia artificial también discrimina—XL Semanal [ABC]. *XL Semanal*. <https://www.abc.es/xlsemanal/ciencia/inteligencia-artificial-discriminacion-tecnologia-chat-gpt-dalle.html>
- Llisterri, J. (2023). *La descripción fonética y fonológica del español: Los procesos fonológicos*. Joaquim Llisterri.cat/. https://joaquimllisterri.cat/phonetics/fon_esp/procesos_fonologicos_espanol_ejercicio.html
- Majerhúa, S. (2008). *La escritura fonémica del quechua ayacuchano* [Tesis de maestría, Facultad Latinoamericana de Ciencias Sociales]. <https://repositorio.flacsoandes.edu.ec/bitstream/10469/804/3/TFLACSO-2008SMC.pdf>
- Mariño, S. y Primorac, C.(2016). Propuesta metodológica para desarrollo de modelos de redes neuronales artificiales supervisadas. *IJERI: International Journal of Educational Research and Innovation*, 6, 231-245. https://repositorio.unne.edu.ar/bitstream/handle/123456789/33803/RIUNNE_FACENA_AR_Mari%c3%b1o-Primorac.pdf?sequence=1&isAllowed=y
- Ministerio de Cultura. (s. f.). *Lista de lenguas indígenas u originarias* | BDPI. Base de Datos de Pueblos Indígenas u Originarios BDPI del Ministerio de Cultura. Recuperado 7 de julio de 2023, de <https://bdpi.cultura.gob.pe/lenguas>

- Morie, J., Chance, E., Haynes, K., y Rajpurohit, D. (2012). Embodied conversational agent avatars in virtual worlds: Making today's immersive environments more responsive to participants. In P. Hingston (Ed.), *Believable bots* (pp. 99-118). Springer.
- Ocaña-Fernández, Y., Valenzuela-Fernández, L. A., y Garro-Aburto, L. L. (2019). Inteligencia artificial y sus implicaciones en la educación superior. *Propósitos y Representaciones*, 7(2), 536-568. <https://doi.org/10.20511/pyr2019.v7n2.274>
- Orta, M, Santos, R. y Cardozo, C. (2018). Creación colaborativa de recursos educativos abiertos con Voicethread y Genial.Ly para la práctica de la transcripción fonética a partir del dictado en la Universidad. *Tercer Congreso de la Asociación Argentina de Humanidades Digitales. La Cultura de los Datos*. <https://www.aacademica.org/aahd2018/11>
- Perea, J. (2023). Las primeras transcripciones de la lengua española en Le Maître Phonétique (1888-1901), *Loquens*, 9(1-2), e092. *Loquens*, 9, e092.
- Pérez Orozco, B. (2018). *La inteligencia artificial y sus aplicaciones* (M. E. Rentería Rodríguez, Responsable). INCyTU. https://www.foroconsultivo.org.mx/INCyTU/documentos/Completa/INCYTU_18-012.pdf
- Pérez, E., Fernández, C. y Gonzalez García, C. (2023). El avatar como recurso tecnológico para promover una educación innovadora e inclusiva. *Revista Tecnología, Ciencia y Educación*, 7-32. <https://doi.org/10.51302/tce.2023.3397>
- Polyáková, T., y Bonafonte, A. (2008). Transcripción fonética en un entorno plurilingüe. *Proceedings of V Jornadas en Tecnología del Habla*, 207-210.
- RAE. (s. f.). *Avatar | Diccionario de la lengua española*. Edición del Tricentenario. <https://dle.rae.es/avatar>
- Restrepo, S. (2019). *Google Translate vs. Traducción Humana: Percepciones de ocho traductores en torno al papel de este traductor automático en su labor* [Tesis de licenciatura, Pontificia Universidad Javeriana]. <https://repository.javeriana.edu.co/bitstream/handle/10554/43243/Trabajo%20de%20Grado%20SANTIAGO%20RESTREPO%20KLINGE.pdf?sequence=3&isAllowed=y>
- Ríos, A. (2002). Un alfabeto fonético del español para usos informáticos. *Estudios de lingüística del español*, 16. <https://raco.cat/index.php/Elies/article/view/195497>
- Rivera, M. (2004). ¿Se puede enseñar fonética con un diccionario? *ASELE, Actas XV*, 733-740. https://idus.us.es/bitstream/handle/11441/42649/15_0731%20Rivera%20Gonz%C3%A1lez.pdf?sequence=1
- Romanenko, A. y Mendeleev, V. (2016). Speaker-Dependent Bottleneck Features for Egyptian Arabic Speech Recognition. En A. Ronzhin, R. Potapova y G. Németh (Eds.), *Speech and Computer. SPECOM 2016. Lecture Notes in Computer Science (vol. 9811)*. Springer. https://doi.org/10.1007/978-3-319-43958-7_75
- Schaufler, M. L. (2021). El mapa de la discriminación. Reflexiones sobre la insistencia del sexismo y el racismo en la construcción de encuestas. *Avatares de la Comunicación y la Cultura*, 22. <https://publicaciones.sociales.uba.ar/index.php/avatares/article/view/7131>

- Seong, J., Lee, W. y Lee, S. (2021). Multilingual speech synthesis for voice cloning. En *2021 IEEE International Conference on Big Data and Smart Computing (BigComp)* (pp. 313-316). IEEE. <https://doi.org/10.1109/BigComp51126.2021.00067>
- Tebbi, H. y Hamadouche, M. (2022). Multi-agent based Arabic speech synthesis. *Int J Speech Technol.* <https://doi.org/10.1007/s10772-022-09975-8>
- Torero, A. (2007). *El quechua y la historia social andina*. Fondo Editorial del Pedagógico de San Marcos.
- Zhao, G., Sonsaat, S., Silpachai, A., Lucic, I., Chukharev-Hudilainen, E., Levis, J. y Gutierrez-Osuna, R. (2018). L2-ARCTIC: A Non-native English Speech Corpus. *Proc. Interspeech*, 2783-2787.
- Unesco. (2010, febrero 23). *SPDA Actualidad Ambiental*. <https://www.actualidadambiental.pe/unesco-29-lenguas-originarias-peruanas-permanecen-en-peligro-de-extincion/>
- Valqui, J., Ziemendorff, M., Ziemendorff, S. y Oisel, G. (2023). Consideraciones histórico-lingüísticas acerca del topónimo Kuélap. *Indiana*, 40(1), 131-154. <https://doi.org/10.18441/ind.v40i1.131-154>

Contribución de los autores

Óscar Huamán-Águila fue responsable y creador del sistema que permite la escritura fonética. Carlos Fernandez-García participó en la elaboración, recolección de datos, diseño de la investigación y, junto a Carlos Gonzales-García, en la redacción del artículo. Además, Carlos Gonzales-García realizó la revisión crítica del artículo.

Agradecimientos

Los autores desean expresar su gratitud a todas las personas y entidades que, de manera directa o indirecta, contribuyeron al desarrollo de este estudio. Agradecemos especialmente a la Universidad Nacional Mayor de San Marcos por proporcionar los recursos necesarios y el ambiente académico propicio para llevar a cabo esta investigación. También extendemos nuestro agradecimiento a nuestros colegas y amigos por su constante apoyo y motivación.

Financiamiento

La investigación se realizó sin financiamiento.

Conflicto de intereses

Los autores no presentan conflicto de interés.

Correspondencia: carlosenrique.fernandez@unmsm.edu.pe

Trayectoria académica de los autores

Óscar Huamán Águila es Licenciado en Literatura y Magíster en Lingüística por la Universidad Nacional Mayor de San Marcos (UNMSM). Sus áreas de investigación se centran en la literatura de pueblos originarios de América Latina, la lingüística andina y el contacto de la lengua quechua con la lengua arawak de la Amazonía. Se desempeña como docente de pregrado y de idioma quechua en el Centro de Idiomas de la Facultad de Letras y Ciencias Humanas de la UNMSM. Además, trabaja en la traducción de los noticieros quechuas de Letras TV Willakun de la Facultad de Letras. En 2020, obtuvo una mención de reconocimiento del “Premio Casa de las Américas 2020” en la categoría de estudios sobre Culturas Originarias de América. Ha publicado artículos en revistas especializadas y es miembro colaborador del grupo de investigación “Discursos, representaciones y estudios interculturales (EILA)” de la UNMSM.

Carlos Fernandez-García es un comunicador experto en tecnologías emergentes como la realidad aumentada e inteligencia artificial aplicadas al periodismo, la comunicación audiovisual, organizacional y el diseño profesional. Es docente investigador RENACYT, doctorando en Educación de la Universidad Nacional Mayor de San Marcos, y magíster en Educación con diplomados en Realidad Aumentada y Realidad Virtual (Universidad Tecnológica Tech, México) y en Acompañamiento y Evaluación de la Tesis Universitaria (Universidad Nacional Mayor de San Marcos, Perú). Además, tiene una especialización en Immersive Journalism: Virtual Reality & 360 Video del Knight Center for Journalism in the Americas. Se desempeña como docente asociado del pregrado de Comunicación Social y de la maestría en Comunicaciones en el posgrado de la Facultad de Letras de la UNMSM. Pertenece a MEDIALAB, un grupo de investigación de la UNMSM reconocido en 2022 como el mejor grupo de investigación de la Facultad de Letras y Ciencias Humanas. Actualmente, es jefe encargado de la Oficina de Comunicación e Imagen Institucional de la misma facultad. Entre sus logros más destacados se incluyen la fundación del canal de eventos y conferencias Letras TV, la creación y producción del noticiero universitario en quechua Letras TV Willakun, la realización de la campaña de realidad aumentada para la prevención del COVID en quechua, y la creación de Illariy, la primera presentadora de noticias generada por Inteligencia Artificial en lengua quechua. Es autor del libro “7 Ensayos de la Interpretación de la Realidad Aumentada” (Universidad Privada del Norte, Perú, 2020) y coautor de siete capítulos en libros como “Artificial Intelligence in Higher Education and Scientific Research Future” (Singapur, 2023), entre otros. Ha publicado artículos científicos en revistas como The Electronic Journal of Information Systems in Developing Countries [Scopus] (China, 2020), Actas de Diseño [Dialnet] (Argentina, 2021 y 2022), EVUlation (Colombia, 2019), EducaUMCH [Latindex] (Perú, 2023), y otras.

Carlos Ricardo Gonzales-García es comunicador y especialista en Periodismo y Relaciones Públicas. Es licenciado en Ciencias de la Comunicación y Magíster en Periodismo y Comunicación Multimedia (2016) por la Universidad de San Martín de Porres (1998) y Magíster en Investigación de la Comunicación Social por la Universidad Nacional Mayor de San Marcos (2015). Se ha desempeñado en Relaciones Públicas y Periodismo en gremios empresariales como la Cámara de Comercio de Lima, ONG como Fovida, y en medios periodísticos de Lima y Huancayo. Es autor de artículos en revistas científicas indexadas en WoS y Scopus en Ecuador, Chile, Argentina y España. Actualmente, coordina MediaLab en la Facultad de Letras y Ciencias Humanas de la UNMSM. Ha sido profesor invitado en la Pontificia Universidad Católica del Perú (2009), la Universidad Particular de Chiclayo (2010) y el Instituto de Arquitectura, Diseño y Arte de la Universidad de Ciudad Juárez, México (2018). También ha sido profesor en la Universidad Nacional del Centro del Perú (2002-2013) y en la Universidad de San Martín de Porres (2016-2018). Actualmente, coordina la maestría en Comunicaciones de la Facultad de Letras y Ciencias Humanas de la UNMSM y es miembro del consejo directivo de las revistas ‘Tesis’ y ‘Socialium’. En noviembre de 2015, recibió un reconocimiento como investigador con publicación científica otorgado por el Vicerrectorado de Investigación de la UNMSM.