

Inteligencia Artificial para el análisis de la pronunciación de lenguas indígenas



Mtro. Eric Ramos Aguilar

Dr. J. Arturo Olvera López

Dr. Ivan Olmos Pineda

*Facultad de Ciencias de la Computación,
Benemérita Universidad Autónoma de Puebla*

Resumen

El análisis del lenguaje en la actualidad es considerado de importancia para diferentes investigadores, debido a que existen personas con el objetivo de aprender una segunda lengua, esto lleva a que científicos desarrollen aplicaciones que generen una retroalimentación de textos o pronunciaciones realizadas por los aprendices, proporcionando una puntuación y una recomendación dentro del proceso de enseñanza. Con el paso del tiempo las lenguas consideradas universales como el Inglés, Español o Mandarín se han tomado en cuenta para desarrollar dichas aplicaciones debido a la demanda de estas.

Sin embargo, en años recientes lenguas consideradas de bajos recursos, por la poca existencia de datos dentro de un entorno computacional y la escasez de hablantes, se han tomado en cuenta con distintos objetivos tales como: preservación, investigación semántica, aprendizaje, entre otros; considerando entre ellas lenguas indígenas de todo el mundo, siendo parte de una riqueza cultural dentro de las regiones que se hablan.

El objetivo de este artículo es presentar el lenguaje como una forma de comunicación y que puede considerar aspectos computacionales para este proceso, además, analizar algunas aplicaciones utilizadas para la enseñanza y aprendizaje de otra lengua, contemplando procesos de reconocimiento de voz automático y los retos que considera el análisis de lenguas indígenas de México.

Palabras clave

Pronunciación, evaluación, aplicaciones, lenguas indígenas

Abstract

Language analysis is currently considered important for different researchers because there are people with the objective of learning a second language, scientists develop applications that provide feedback on texts or pronunciations made by learners, generating a score and a recommendation within the teaching process. Over time, languages considered universal such as English, Spanish, or Mandarin are considered to develop these applications due to their demand.

However, in recent years other languages considered low resources due to their low appearance of data within a computational environment and the cases of speakers, have been considered with different objectives, preservation, semantic research, learning, among others, considering, among them, indigenous languages from around the world, being part of a cultural wealth within the regions that are spoken.

The objective of this article is to present language as a form of communication that can consider computational aspects for this process, in addition to analyzing some applications used for teaching and learning another language, considering automatic voice recognition processes and the challenges that consider the analysis of indigenous languages of Mexico.

Keywords

Pronunciation, evaluation, applications, indigenous languages

1. El lenguaje como una forma de comunicación

La expresión de ideas y forma de comunicación se han desarrollado comúnmente a través de lenguajes que forman parte de

un sistema estructurado y que ha servido para distintos fines. A través del lenguaje se pueden generar múltiples canales de diálogo para desarrollar negocios, educación, arte, ciencia y tecnología. Además, la lengua forma parte de una identidad comunitaria, debido a que influye en el pensamiento cultural y de tradiciones al relacionarse entre sí.

Los avances tecnológicos actualmente han desarrollado métodos para analizar el lenguaje de distintas familias lingüísticas con el fin de conocer la estructura, escritura y pronunciación. Sin embargo, el lenguaje es un tema complejo debido a las características lingüísticas (fonología, sintaxis, fonética, semántica y lexicografía) que conservan los idiomas, por lo que las investigaciones ponen en práctica diferentes métodos para crear herramientas que promuevan el análisis de lenguajes.

Las investigaciones recientes en procesamiento de lenguaje natural a través de inteligencia artificial se han enfocado y dividido en distintas áreas de investigación, tratando de resolver problemas concretos como reconocimiento de voz, análisis de pronunciación, detección de idiomas, retroalimentación de palabras, reconoci-

miento de errores de escritura en texto o voz, entre otros.

El objetivo de elaborar un sistema que sea similar a las capacidades de comunicación del ser humano toma como referencia el modelo presentado por Claude Shannon, donde una fuente de información (en este caso el habla) se transmite a través de un canal de comunicación, que es el medio físico por donde se transfieren los mensajes y que también mediante este un destinatario recibe dicha información, figura 1.

El análisis de la pronunciación de lenguas indígenas es un tema que ha captado la atención de distintos investigadores dedicados al estudio de idiomas, por la complejidad en la producción de sonidos y tonalidades que integran, enfocando el uso de inteligencia artificial aplicada al procesamiento de lenguaje natural con el fin de producir herramientas para un proceso de enseñanza-aprendizaje similar al de lenguas como el Inglés, Mandarín o Francés.

2. Aplicaciones para el análisis de la pronunciación

En la actualidad existen herramientas computacionales enfocadas a la pronunciación capaces de analizar distintas lenguas, ge-

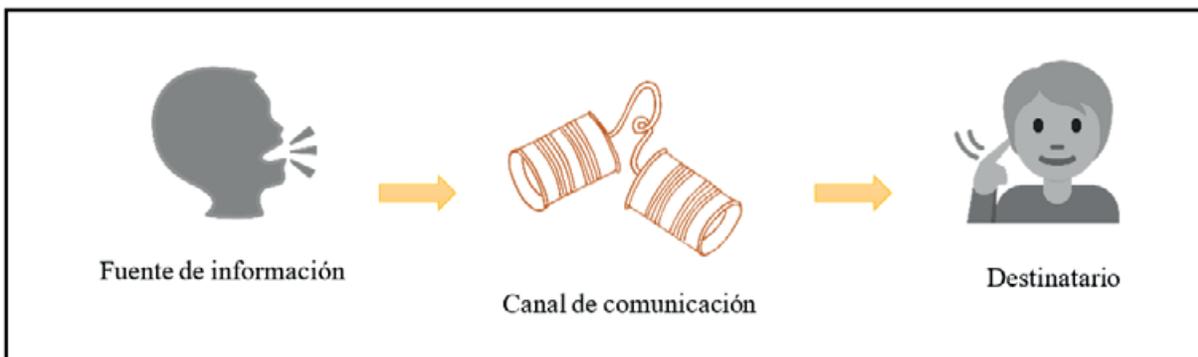


Figura 1. Representación del modelo de comunicación de Shannon.

Fuente: Elaboración propia con base en (Medina, 1999)

nerando una retroalimentación respecto a una mala o buena pronunciación por parte de un usuario. Con regularidad las lenguas utilizadas para generar una aplicación son aquellas que tienen un número de hablantes amplio y que son de gran demanda por los estudiantes, tal es el caso del idioma Inglés, Mandarín, Francés y Alemán.

Un ejemplo de una aplicación es Rosetta Stone, la cual es una herramienta de aprendizaje de lenguas a través de internet, la cual realiza una retroalimentación instantánea de los ejercicios y emite una evaluación sobre ellos, entre los idiomas que se pueden estudiar son Inglés, Italiano, Alemán, Español (Latino y de España), Árabe y Francés. Otra aplicación utilizada comúnmente para realizar este tipo de aprendizaje es Open English, la cual se considera como una escuela de Inglés en línea con un análisis y retroalimentación de la pronunciación, proporcionando los porcentajes de buena pronunciación dentro de una oración o de palabras.

Duolingo, es otra herramienta usada para el aprendizaje de Inglés, Francés, Alemán, Ruso, Italiano, entre otros, proporcionando una retroalimentación de las pronunciaciones similar a las de las aplicaciones anteriores. Las lenguas utilizadas para este tipo de herramientas son aquellas que consideran un gran número de hablantes y que son conocidas como un idioma universal por el hecho de ser utilizadas para poder comunicarse con otras personas de habla nativa diferente; por su parte los idiomas con un número de hablantes reducido en donde se contemplan las lenguas indígenas consideran aplicaciones en desarrollo para su aprendizaje.

Actualmente las aplicaciones para aprender lenguas indígenas en México aún no

se han desarrollado como las mencionadas anteriormente, existen algunas técnicas de aprendizaje de la lengua que se basan en escuchar audio y realizar repeticiones de palabras, pero sin obtener una retroalimentación de estas, tal es caso de Fantre que ha desarrollado aplicaciones para dispositivos móviles que ayudan a conocer lenguas como Mixteco, Mazahua, Otomí y Aimara, este último perteneciente a la región sur del continente americano.

Las aplicaciones anteriores contienen audio, imágenes y videos que representan la lengua de la región, distribuidos en campos semánticos (animales, números, cuerpo, entorno, frutas y verduras) de ayuda para el aprendizaje de una segunda lengua. Existe una diferencia entre las aplicaciones utilizadas en las llamadas lenguas universales y estas, donde se considera un desarrollo computacional previo que ha utilizado el reconocimiento de voz automático para generar la retroalimentación de la pronunciación de la lengua y que logra una interacción entre los seres humanos y las computadoras.

3. ¿Qué hace el reconocimiento de voz automático?

El lenguaje es el recurso con el que las personas se comunican, generando interacciones y relaciones sociales, políticas, de negocios y más. Este tipo de comunicación se lleva a cabo en algunos casos a través de dispositivos electrónicos que funcionan como canal para este proceso, comúnmente mediante entradas de audio, lo cual es más cómodo para el usuario respecto al uso de un teclado.

La interacción humano-computadora ha sido posible con los desarrollos computacionales como aplicaciones en dispositivos

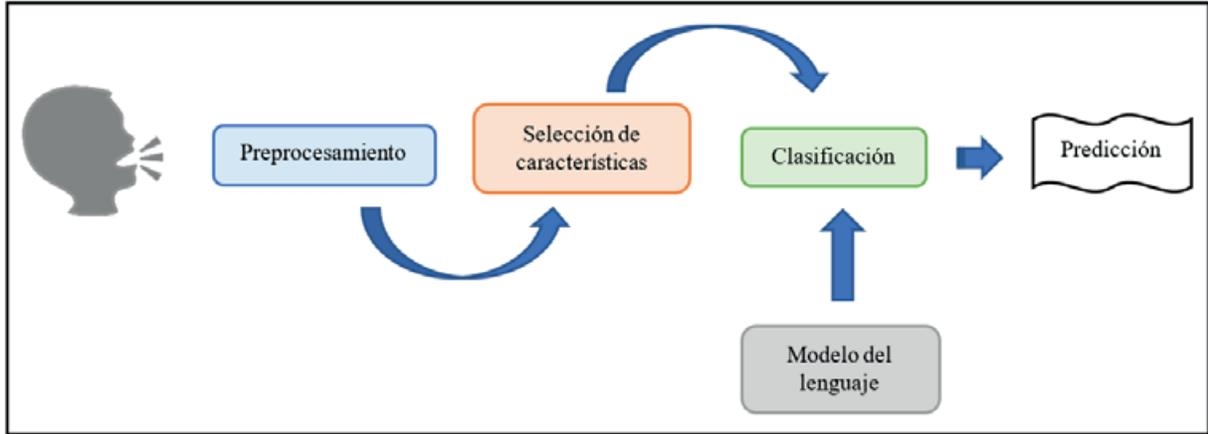


Figura 2. Esquema de las etapas del reconocimiento de voz automático.

Fuente: Elaboración propia con base en (Yu, 2016).

móviles o interacción a través de páginas web, acceso libre a los datos y el avance continuo del internet, con ello los dispositivos móviles e inteligentes han puesto en práctica procesos de reconocimiento de voz en la tecnología del habla y han mejorado la comunicación entre humanos y computadoras (Yu, 2016).

El reconocimiento de voz automático considera las etapas mostradas en la figura 2 que se describen a continuación:

- **Preprocesamiento de señal:** Se considera una señal de audio como entrada, en donde a través de procesos computacionales se realiza una reducción de ruido, distorsión y amplificación de la señal de audio (figura 3).

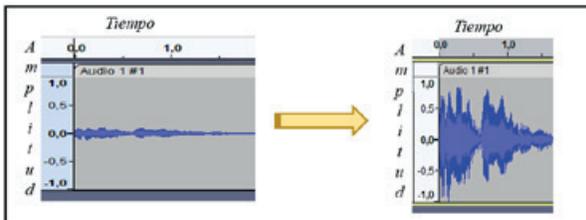


Figura 3. Representación de preprocesamiento de una señal de audio.

Fuente: Elaboración propia con base en (Yu, 2016)

- **Extracción de características:** Las representaciones en el dominio del tiempo y la frecuencia proporcionan valores numéricos que son manipulados computacionalmente, como la altura y ancho de una señal, el tiempo que dura un audio, las repeticiones de una señal dentro de un espacio de tiempo que son nombradas frecuencias, a partir de los cuales es posible obtener valores estadísticos como el promedio y la desviación estándar con lo que se crea una vector llamado vector de características (figura 4).

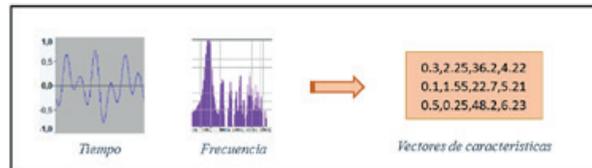


Figura 4. Modelo de selección de características.

Fuente: Elaboración propia con base en (Yu, 2016).

- **Clasificación:** Con las características obtenidas previamente y a través de un proceso computacional (que se vuelve un analizador, seleccionador y buscador de coincidencias entre caracteres) se encuentran patrones que describen palabras de manera general (figura 5).

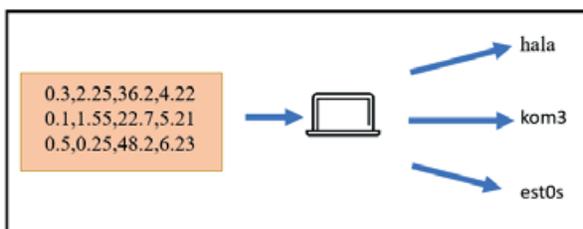


Figura 5. Representación de la clasificación con respecto a un vector de características.

Fuente: Elaboración propia con base en (Yu, 2016).

- **Modelo del lenguaje:** Captura las reglas gramaticales o información semántica de la lengua objetivo, considerando palabras individuales, fonemas o sílabas a las que se les llama tokens de salida, ayudando a realizar correcciones, facilitando la comprensión debido a que permiten la representación del habla en forma de texto (figura 6).

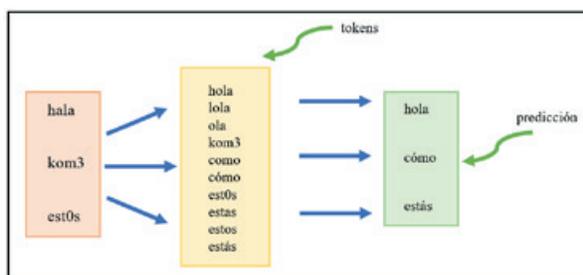


Figura 6. Predicción de palabras a través de tokens en un modelo del lenguaje.

Fuente: Elaboración propia con base en (Yu, 2016)

Con el proceso descrito anteriormente es posible realizar un reconocimiento de voz de alguna lengua, sin embargo, a veces este proceso no se puede aplicar a todas las lenguas, debido a factores que afectan el rendimiento del sistema para realizar un posible reconocimiento, uno de esos factores es la cantidad de datos utilizados, ya que entre mayor sea el número de ejemplos de palabras u oraciones, un sistema computacional puede tener más represen-

tación de la lengua y con ello se obtienen mejores resultados en la predicción.

En la literatura referida a análisis de lenguas, es posible encontrar corpus como TIMIT, VoxForge, LibriSpeech, EpaDB; los cuales contienen más de ocho mil instancias, con descriptores léxicos y fonéticos, grabaciones de audio de personas de diferentes edades y géneros, con conocimientos acústicos.

4. El uso de datos en el análisis del lenguaje

Dentro del proceso de aprendizaje automático o inteligencia artificial el uso de datos es indispensable para realizar tareas específicas, una de estas es el procesamiento del lenguaje natural, considerando distintos campos de estudio, tal es el caso referente al reconocimiento de voz o evaluación del lenguaje, pero, cuál es la importancia de utilizar un conjunto amplio de datos.

Se pueden considerar tres razones importantes del uso de grandes cantidades de datos:

- La variedad de estos consiste en representar la precisión y variabilidad del mundo real, ya que con ello una aplicación computacional se puede desarrollar para realizar modelos representativos y con ello será capaz de generalizar nuevas situaciones.
- La complejidad de las tareas de inteligencia artificial requiere un gran número de datos para producir una mejora en los procesos computacionales como el tiempo, ayudando a capturar uno o varios patrones y características que mejoren una aplicación computacional.

- Finalmente, se ha demostrado que un mejor rendimiento como tiempo, exactitud y precisión de las aplicaciones computacionales considera un número amplio de datos.

Actualmente la cantidad de datos utilizados en las lenguas universales consideran cerca de ocho mil renglones de datos como se ha mencionado, tal es el caso de un generador de fonemas llamado Allosaurus (Li, 2020) o bien un número alto de horas de grabación de sesenta a ciento veintiocho horas (Mukhamadiyev, 2022) que se proporcionan como entrada para los modelos de inteligencia artificial y pueden

llevar a cabo el reconocimiento de voz o análisis de pronunciación con un 95% a 99% de exactitud en la evaluación.

En la figura 7 se muestra una representación de análisis de lenguaje a través de inteligencia artificial, donde datos etiquetados con buena y mala pronunciación son ingresados a un modelo de entrenamiento (que tiene como objetivo lograr una generalización de conocimientos a través de los datos de entrada y aplicarlos a datos no conocidos). Generalmente, para el proceso de entrenamiento del sistema se considera un porcentaje del conjunto de datos etiquetados y se reserva otra porción (conjunto de prueba

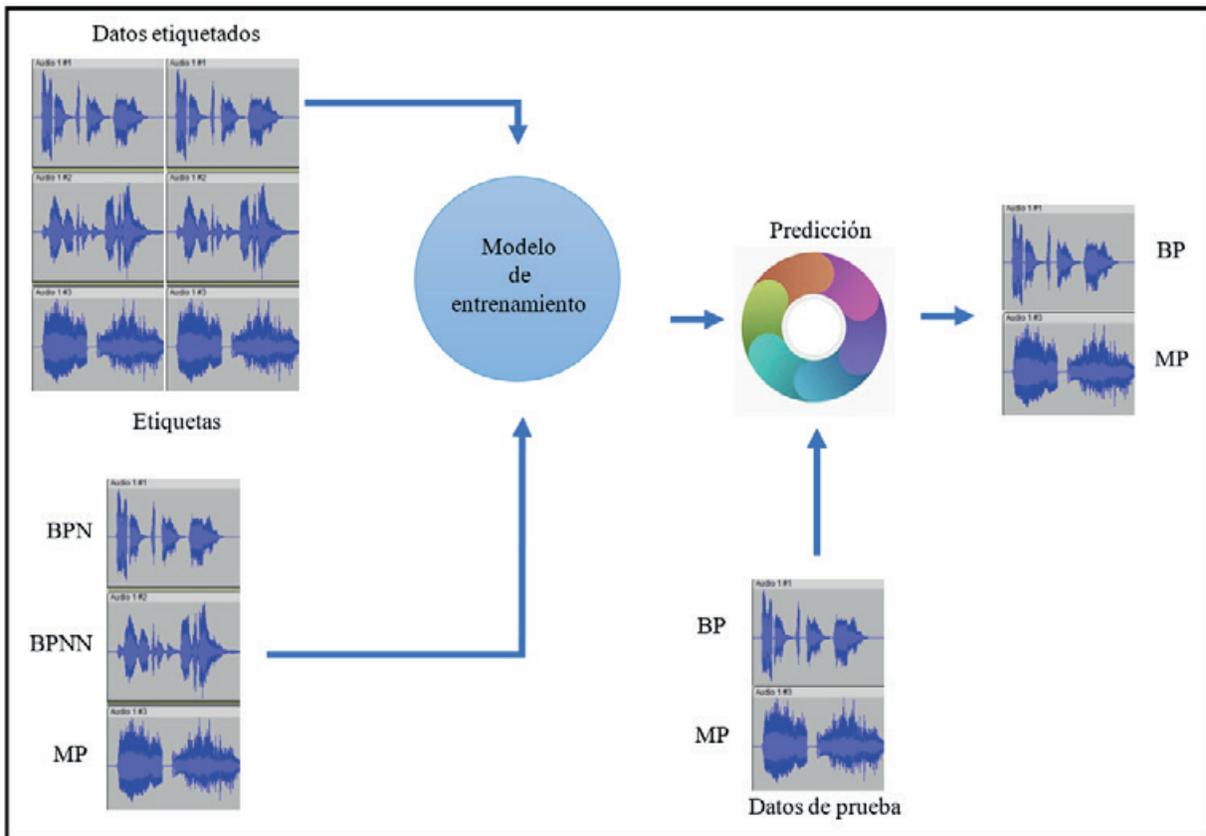


Figura 7. Proceso de análisis de buena y mala pronunciación de una lengua a través de un modelo de aprendizaje automático. BPN (Buena Pronunciación Nativo), BPNN (Buena Pronunciación No Nativo), BP (Buena Pronunciación), MP (Mala pronunciación).

Fuente: Elaboración propia con base en (López, 2006)

ba), para finalmente evaluar el desempeño en términos de la cantidad de ejemplos que se predicen (de entre el conjunto de prueba) como buena o mala pronunciación.

Los modelos de inteligencia artificial como el explicado anteriormente aún no se han aplicado lenguas indígenas del mundo en su totalidad ya que estas integran un número reducido de datos (audio o texto digital), sin embargo, algunas de ellas han utilizado otras lenguas con características similares para poder realizar un análisis y elaborar procesos computacionales similares a las lenguas universales, no obstante, aun con esto los resultados han quedado por debajo de los propuestos por las lenguas universales.

Por ejemplo, el corpus LibriSpeech se utiliza en (Rajendran, 2021) para crear reconocimiento de voz automático en el idioma tamil de India y parte de Sri Lanka. Otro caso es el análisis para identificar lenguajes africanos como Afrikáans, IsiNdebele, IsiXhosa, IsiZulu, Sepedi, Sesotho, Setswana, Siswati, Tshivenda y Xitsonga, donde se utiliza el corpus VoxForge para entrenar una red neuronal (Feng, 2019).

En la actualidad, no es común encontrar desarrollos tecnológicos en inteligencia artificial que consideren análisis de audio en las lenguas indígenas de México. El enfoque propuesto para un estudio de estas lenguas es desarrollado con una extracción de representaciones fonéticas escritas, llevado a cabo con una herramienta computacional como Praat a la cual se le introducen audios unitarios cortos de 30 segundos como máximo y se realizan recortes de manera manual para analizar fonemas y tonos de las palabras.

En Praat es posible apreciar una representación en frecuencia del audio (espectrograma) donde se identifican formantes que son una representación en frecuencia de los fonemas, con ello, muchos lingüistas han reconocido prefijos, sufijos y la recurrencia fonética, logrando interpretar la semántica que tienen palabras o frases de algunas lenguas.

Este tipo de análisis ha ayudado al estudio de las lenguas y su comprensión, además se han realizado con las representaciones fonéticas un alfabeto para algunos lenguajes, tal es el caso del otomí de Ixtenco (Yuhmu), donde nunca se ha establecido una ortografía y la transmisión de conocimiento de la lengua se ha desarrollado a través del habla, por lo que se ha tornado importante este tipo de representación escrita.

5. Retos computacionales actuales de las lenguas indígenas de México

Como se ha mencionado, el proceso de análisis de las lenguas indígenas se ha llevado a cabo mediante software que ayuda a identificar fonemas, tonalidades y segmentos de palabras de una lengua nativa, sin embargo, estos procesos realizados son costosos en tiempo y subjetivos, por lo que es necesario utilizar un proceso que ayude a realizar un análisis de las lenguas que reduzca los tiempos de ejecución y que ofrezca una exactitud en las evaluaciones.

Un ejemplo de evaluación de pronunciación en lenguas indígenas de México es el realizado por (Penner, 2019), en donde se grabaron palabras y frases de persona, incorporando un corpus de seis hablantes nativos de mixteco los cuales fueron analizados a través del software Praat, identificando semántica y características de la lengua.

La automatización de procesos y resultados de análisis con una tasa de error cercanos a los de las lenguas universales pueden ser obtenidos mediante técnicas de inteligencia artificial, sin embargo, esto no quiere decir que se reemplace a un lingüista que analiza la lengua, pero si puede ayudar a su evaluación, considerando que este especialista es quien definirá los procesos de recolección de datos para definir cuáles son funcionales para poder ser evaluados a través de inteligencia artificial.

La adquisición y resguardo de datos son relevantes dentro de un análisis de lenguaje indígena debido a que la preservación de agrupaciones lingüísticas es importante, ya que con ello se conserva la identidad y cultura de un pueblo, se mantiene la forma de comunicación original de este y la lingüística nativa permanece, considerando una forma diferente de comunicarse del ser humano, generando un rescate de estas lenguas en decadencia.

Lo anterior y la decadencia de las lenguas indígenas representa una problemática para las comunidades en México, debido al número de hablantes nativos, los cuales son los únicos que preservan distintas variantes a nivel nacional, siendo en su mayoría personas de la tercera edad; además de desconocer el grado de dominio lingüístico que se tiene de la lengua. De tal forma que los modelos creados mediante inteligencia artificial han tomado relevancia para continuar con la preservación y el análisis de lenguas indígenas mexicanas con datos textuales, de audio y video.

Tomando en cuenta lo anterior es posible enumerar retos importantes de análisis de lenguas indígenas de México:

1. Escasez de datos digitales etiquetados: La mayoría de las lenguas indígenas tienen una cantidad limitada de datos disponibles, especialmente en lo que respecta a datos etiquetados necesarios para el entrenamiento de modelos de inteligencia artificial. Dificultando los desarrollos de modelos efectivos y precisos.
2. Variedad lingüística: México es conocido por su diversidad lingüística, con una gran cantidad de lenguas indígenas habladas en diferentes regiones del país incorporadas a 62 familias. Cada una de estas lenguas presenta sus propias características lingüísticas únicas, dificultando el desarrollo de modelos de inteligencia artificial que sean generalizables a todas ellas.
3. Ausencia de recursos tecnológicos y acceso limitado a la tecnología: Muchas comunidades indígenas en México pueden carecer de acceso a la tecnología y recursos necesarios para recopilar, procesar y analizar datos de manera efectiva utilizando inteligencia artificial, por lo que es necesario la incorporación de ingenieros y lingüistas para recabar datos para su preprocesamiento, análisis y retroalimentación.
4. Adaptación de modelos preentrenados: Los modelos de inteligencia artificial preentrenados generalmente están disponibles en idiomas como el Inglés y el Mandarín. Adaptar estos modelos para trabajar con lenguas indígenas de México puede ser un desafío debido a las diferencias con respecto a las familias lingüísticas que las componen, debido a que no todas coinciden con los mismo sonidos fonéticos.

5. Preservación cultural y ética: Es importante considerar la preservación cultural y los aspectos éticos al trabajar con comunidades indígenas y sus idiomas. Esto incluye el consentimiento informado, la protección de datos y la participación de las comunidades en el proceso de desarrollo de tecnología, trabajando en conjunto con el fin de mantener la cultural lingüística en México.

Finalmente, es por ello que es de importancia realizar un seguimiento y trabajo constante para las lenguas indígenas en México, dada su decadencia y falta de información digital actual.

Referencias

Feng, K. (2019). Low-resource language identification from speech using transfer learning. 2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP), 1-6

Li, X., Dalmia, S., Li, J., Lee, M., Littell, P., Yao, J., y Metze, F. Universal phone recognition with a multilingual allophone system. ICASSP 2020-2020 IEEE Inter-

national Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.8249-8253, 2020.

López, J. M. M., y Herrero, J. G. Técnicas de análisis de datos. Aplicaciones prácticas utilizando Microsoft Excel y weka, 2006.

Medina, J. d. (1999). Explicación y Análisis: Taller de Comunicación I. México: UNAM.

Mukhamadiyev, A., Khujayarov, I., Djuraev, O., y Cho, J. Automatic speech recognition method based on deep learning approaches for Uzbek language. Sensors, 22(10), 3683, 2022.

Penner, K. (2019). Prosodic structure in Ixtayutla Mixtec: Evidence for the foot.

Rajendran, S. M. (2021). Language dialect-based speech emotion recognition through deep learning techniques. International Journal of Speech Technology, 24(3), 625-635.

Yu, D., y Deng, L. Automatic speech recognition (Vol. 1). Berlin: Springer, 2016.