ANALES DE ANTROPOLOGÍA

INSTITUTO DE INVESTIGACIONES ANTROPOLÓGICAS

Anales de Antropología 55-II (julio-diciembre, 2021): 13-22

www.revistas.unam.mx/index.php/antropologia

Artículo

Las tecnologías de Reconocimiento Automático de Voz y su incorporación a los métodos de transcripción de lenguas indígenas

Automatic Voice Recognition technologies and their incorporation into indigenous language transcription methods

Hilaria Cruz*

[1] *Universidad de Louisville, Bingham Humanities 303 2211, South Brook, University of Louisville, Louisville, CP 40292* USA.

Recibido el 15 de diciembre de 2021; aceptado el 3 de marzo de 2021.

Resumen

En agosto del 2018, un retiro en Quechee, Vermont, reunió a lingüistas, hablantes de idiomas en alto riesgo de desaparición, científicos de la computación especializados en procesamiento del lenguaje natural y activistas en el área, con el propósito de discutir las posibilidades de conjuntar esfuerzos para integrar las tecnologías de reconocimiento automático de voz (especialmente las redes neuronales artificiales), a los métodos de transcripción de estas lenguas.

En un ambiente ameno, donde el trabajo se mezcló con la diversión, los participantes tuvieron la oportunidad de conocerse, intercambiar ideas, conocimientos y experiencias, dialogando sobre los recursos que el trabajo de documentación lingüística y el reconocimiento de voz, podrían aportar para llevar a cabo dicha meta. Ambos campos compartieron sus últimos avances y las condiciones de sus respectivas áreas de investigación, incluyendo las condiciones de campo, la vitalidad de la lingüística de sus respectivas lenguas, al igual que temas relacionados con el "embotellamiento" de la transcripción del lenguaje.

Los lingüistas reportaron sobre su corpus y las horas que habían colectado. De igual manera, los participantes expresaron sus necesidades tecnológicas y sobre cómo debería ser un sistema de Reconocimiento Automático de Voz que pudiera ser usado por personas que no saben mucho sobre tecnología.

Abstract

In August 2018, a group of documentary linguists, speakers of endangered languages, endangered language activists, and natural language processing specialists came together at a retreat in Quechee, Vermont. The goal of the retreat was to help resolve the current bottleneck in natural languages transcription. Researchers at the retreat were particularly interested in ways to utilize new automatic speech recognition technologies, especially artificial neural networks, in the field. Prior work in natural languages transcription did not have access to such technology and these discussions were extremely fruitful.

In this welcoming environment, the participants from different fields and backgrounds had the opportunity to get to know one another, exchange ideas, knowledge, and experiences. The retreat hinged on discussions about the different resources each group used in the work of linguistic documentation and voice recognition. Both camps shared their latest advances and current conditions of their respective fields, including the linguistic vitality of their respective languages, the size of their *corpus*, their workflows, among other equally related themes.

Participants also expressed needs and wants related to implementation of such technologies in the field. In particular, speakers of endangered languages do not traditionally have access to advances natural language processing technology, while those with such technology are often unable to interact with the communities most in need. Bridging this gap was determined to be a key goal of all the groups in attendance.

Palabras clave: Lenguas con bajos recursos tecnológicos; lenguas en alto riesgo de desaparición; Tecnología RAV

Keywords: Low-resourced languages; Endangered Languages; ASR

* Correo electrónico: hilaria.cruz@louisville.edu

DOI: 10.22201/iia.24486221e.2021.77857

Antecedentes

El Reconocimiento Automático de Voz (RAV) (también llamado Reconocimiento Automático del Habla), es una tecnología floreciente con un potencial casi ilimitado. El reconocimiento de voz, que una vez fue considerado ciencia ficción, ahora es algo común: *Alexa*, *Siri* y *OK Google* viven en burós y mesitas de noche en nuestros hogares. Sin embargo, esta tecnología se ha limitado a lenguas dominantes como el inglés, el alemán y el español.

Desde los siglos XIX y XX, los Estados nacionales han erradicado poco a poco la diversidad lingüística y cultural, ocasionando el rápido declive global de las lenguas indígenas a nivel mundial. Se calcula que la gran mayoría de los 7.000 idiomas que actualmente se hablan en el mundo quedarán en desuso a finales de este siglo si la pérdida de lenguas continúa su ritmo actual, limitando de esta manera la característica de la humanidad "de apreciar todas las capacidades creativas de la mente humana" (Mithun 1998: 189).

El rápido desarrollo de las tecnologías RAV, especialmente de las redes neuronales artificiales, podrían ayudar a incrementar el *corpus*,[1] avanzar la investigación, la revitalización y la difusión de las lenguas minorizadas. Estas tecnología pudieran a acelerar la transcripción, la traducción y la anotación lingüística. La automatización del proceso tambien pueden darle mas consistencia a los textos y por consiguiente aliviar la ardua tarea del transcriptor, pudiéndolo convertir en un editor, más que solo un capturista de datos.

Las tecnologías RAV también podrían tener ramificaciones positivas en otras áreas de la revitalización lingüística al proporcionar una nueva visión de la naturaleza de las lenguas minorizadas ante los ojos de los hablantes y la sociedad en general. De igual manera podrían animar a los hablantes a capacitarse en la tecnología.

Para que dichas metas se puedan lograr, las partes interesadas (los lingüistas documentales,[2] computólogos, hablantes y activistas de lenguas) deberán superar múltiples obstáculos de índole tecnológica, metodológica, colaborativa, social y politica.

El trabajo a desarrollar por los computólogos y los lingüistas

Los computólogos interesados en el procesamiento del lenguaje natural (PLN), como los lingüistas documentales, tienen muchos intereses en común que pudieran ser áreas muy fértiles para la colaboración. Los dos usan como base de investigación el lenguaje natural, al igual que dependen de un *corpus* de lenguaje para llevar a cabo sus investigaciones. Por consiguiente, hay un creciente número de computólogos que desean integrar lenguas minorizadas a sus teorías y modelos. De igual forma, los lingüistas están buscando una manera de automatizar el análisis de las lenguas que estudian.

Los lingüistas documentales trabajan cercanamente con hablantes para registrar diversos usos de la lengua para diferentes fines. Éstos incluyen elicitaciones gramaticales, grabaciones de usos formales y de la vida diaria, incluyendo cuentos, historias, rezos, oratorias, intercambios formales entre dos o más hablantes y conversaciones de la vida cotidiana. Unos lo hacen para registrar la lengua para la posterioridad, otros para llevar a cabo análisis y desarrollar teorías gramaticales, y otros lo hacen para llevar a cabo estudios sociolingüísticos, tipológicos y de discurso. Los computólogos que hacen investigación en el PNL, por el contrario, crean algoritmos y modelos para automatizar el proceso de análisis de una lengua o de grandes datos. También usan grandes *corpus* de datos para enseñarle una lengua natural a una computadora.

Algunas de las barreras que obstaculizan la colaboración entre estos investigadores incluyen que las partes interesadas muchas veces desconocen el flujo de trabajo de cada área de estudio. Los lingüistas documentales sienten que los computólogos siempre andan persiguiendo o tratando de conquistar nuevas fronteras y rara vez se detienen a conocer una lengua con mayor profundidad.

Tanto los lingüistas como los computólogos asumen que los investigadores de otras áreas conocen "naturalmente" las metodologías, terminologías y *softwares* de sus respectivas áreas de estudio, pero éste no es el caso. Por ejemplo, *software* como *ELAN*, *Transcriber*, o *Fieldworks* que los lingüistas usan comúnmente para transcribir, anotar y analizar audios o textos, por lo general no son conocidos por lo computólogos.

De igual manera los computólogos asumen que los lingüistas documentales saben programar en varias lenguas o que éstos están familiarizados con las plataformas de colaboración, tales como github, slack y otros lugares donde los computólogos depositan, comparten y adquieren sus códigos y sus *corpus*. Éstas son áreas que los lingüistas documentales por lo general desconocen.

Tanto los lingüistas documentales como los computólogos tienen conocimientos y metodologías importantes qué compartir y contribuir para avanzar en la integración del RAV a las lenguas en alto riesgo de desaparición. Los computólogos con su habilidad de crear, navegar y modelar la tecnología del lenguaje (p. ej. creación de *software*, plataformas de colaboración) al igual que su habilidad de trabajar con *corpus* grandes. Los lingüistas documentales por el otro lado, tienen mucha experiencia trabajando con comunidades de hablantes, conocen de teorías gramaticales y algunos se pasan mucho tiempo o muchas veces toda su carrera trabajando una sola familia lingüística.

Para que la colaboración sea exitosa entre estos investigadores y sus áreas de estudios, las dos partes deben ser

[1] "Un conjunto grande y estructurado de textos legibles por máquina que se han producido en un entorno comunicativo natural. Se pueden derivar de diferentes formas, como texto que originalmente era electrónico, transcripciones del lenguaje hablado y reconocimiento óptico de caracteres, etc." (NLP-Linguistic Resources-Tutorialspoint 2021)

[2] "Un professional que se ocupa de la elaboración y el mantenimiento de registros de los idiomas del mundo y sus patrones de uso" (Woodbury 2003: 35).

flexibles y tienen que estar dispuestas a aprender de la otra. Los lingüistas tendrán que aprender de las herramientas y métodos de los computólogos (p. ej., aprender a programar) y a su vez los computólogos tendrán que detenerse a aprender más sobre la agencia local y las necesidades de las comunidades (Bird 2020). De igual manera, pudieran detenerse y familiarizarse con lo que se ha escrito en las lenguas que están trabajando y mucho mejor si pudieran hacer trabajo de campo.

Corpus

Todos los modelos RAV que hasta ahora existen, como las cadenas Márkov, la alineación forzada, los modelos ocultos de *Márkov*, *Kaldi* y *ESPnet*, requieren de enormes cantidades de *corpus* de entrenamiento para poner en marcha sus sistemas (Watanabe *et al.* 2021; Michaud *et al.* 2018).

Actualmente existen un sin fin de obstáculos para obtener *corpus* en lenguas minorizadas para entrenar a los sistemas RAV. Esto se debe a una convergencia de factores de índole político, social y tecnológico que han resultado de siglos de racismo institucional hacia dichas lenguas y sus hablantes.

Los estados nacionales han establecido e implementado políticas publicas dirigidas a desaparecer la diversidad lingüística y cultural de sus naciones. Así como lo mencionábamos antes, dichas políticas han hecho obligatorio el uso de las lenguas dominantes como lenguas de instrucción en las escuelas públicas, forzando a los hablantes de lenguas minorizadas a aprender, hablar y escribir dichos idiomas.

Cada día que pasa, las lenguas indígenas van perdiendo más campo de uso y, como consecuencia, se está volviendo más difícil encontrar contextos donde todavía se usen estas lenguas, como se puede observar en la casi nula presencia de las lenguas minorizadas en los medios masivos de comunicación –ya sea en los medios impresos, audiovisuales o digitales– tales como libros, periódicos, revistas, películas, programas de radio, televisión, y redes sociales.

Desde la Colonia hasta el siglo XXI, la investigación lingüística así como la capacitación de los hablantes en el uso de tecnología para crear, organizar y traducir *corpus* para entrenar modelos RAV, lucen por su abandono. Todo esto es parte del daño colateral de estas políticas públicas que han provocado el bajo prestigio de estas lenguas, las cuales son vistas como lenguas que no tienen valor o futuro, y que los padres y madres se rehúsen a transmitirlas a las nuevas generaciones por miedo a que les haga daño.

Obstaculos para transcribir lenguaje natural

La traducción, transcripción, y anotación manual del habla natural en cualquier lengua es un proceso lento y tedioso. Aunque los que realizan estas tareas sean hablantes nativos de la lengua, el/la transcriptor(a) deberá escuchar muy atentamente el audio y escribir cada sonido o palabra que escuche, tratando de no omitir algo. El proceso no presenta mayores complicaciones cuando el audio es relativamente limpio, cuando hay una sola persona hablando claramente y cuando el audio no tiene ruido en el *background*. Si la lengua en cuestión tiene una ortografía establecida, y el transcriptor sabe leer y escribir esta lengua, se facilitará mucho la tarea.

El proceso se complica rápidamente si la calidad del audio no es buena, si hay mucho ruido en el contexto del habla o si las personas en la grabación no hablan claramente. De igual manera, las conversaciones entre varias personas o conversaciones solapadas complican el proceso.

En estos casos, los transcriptores tendrán que pausar, regresar y repetir el audio muchas veces, haciendo la tarea aún más lenta y tediosa. Estas complicaciones son muy comunes en los contextos donde existen las lenguas en alto riesgo de desaparición.

En los pueblos indígenas, es difícil encontrar espacios que sean idóneos para grabar audios libres de ruido. La gente toca música a un volumen muy alto, los aparatos de sonido gritan mensajes a cualquier hora del día, aunado al ruido de animales y motores. De igual manera, las grabaciones en las lenguas minorizadas suelen colectarse con ancianos –quienes, en la mayoría de los casos, son los últimos hablantes de la lengua– que muchas veces no tienen dientes, lo cual dificúlta la enunciación clara de los sonidos de la lengua.

Creación de *corpus* en lenguas en alto riesgo de desaparición

La transcripción de las lenguas minorizadas son un proceso aún más largo y complicado porque tiene que considerar todos los factores arriba mencionados, entre otros. La manera en la que tradicionalmente se han creado *corpus* de las lenguas indígenas es a través del trabajo de campo etnográfico y la documentación lingüística de los últimos 20 años. Un investigador, por lo general externo a la comunidad con una formación occidental, se traslada a los lugares donde tradicionalmente se habla la lengua y tiene que ganarse la confianza de la comunidad de hablantes para poder obtener permiso para grabar. Muchas veces estos lugares son difíciles de acceder: lugares plagados de conflictos, brotes de malaria y dengue.

Una vez obtenida la grabación, el investigador o investigadora deberá trabajar con un hablante. Juntos deberán escuchar y discutir los sonidos y palabras en el audio: el lingüista toma notas y formula preguntas, y el consultado repite las palabras una y otra vez. A esta tediosa actividad se le ha denominado el embotellamiento de la transcripción. Como hablante nativa de la lengua Chatina, me toma un promedio de 30 minutos transcribir un minuto de audio. Para un investigador no nativo, este proceso es mucho más lento.

El hecho de que una persona sepa hablar su lengua, no lo hace experto en el análisis y transcripción de la misma. Los hablantes de lenguas minorizadas que actualmente están creando *corpus* para estudios RAV en sus lenguas han aprendido a leer y escribir siendo ya adultos. Cuando una persona se alfabetiza como adulto, tiene más barreras que contender que si lo hubiera hecho de niño. Las dificultades de aprender a leer y escribir la lengua materna como adultos incluyen barreras técnicas y cognitivas. Un adulto muchas veces no tiene el tiempo, la experiencia, la tradición, o la práctica para escribir y reconocer los sonidos de su lengua y, por consiguiente, le tomará más tiempo aprender y familiarizarse con el sistema fonético de la misma.

De igual manera, la gran mayoría de los hablantes de lenguas indígenas en México, por ejemplo, apenas empiezan a escribir su lengua. Esta autora fue alfabetizada en español e inglés mucho antes de que pudiera leer y escribir en Chatino, ya que el Chatino de San Juan Quiahije no tenía un alfabeto de trabajo hasta que la autora tuvo cuarenta años. Como decíamos antes, los hablantes de lenguas indígenas de México han sido alfabetizados en español y solo han aprendido la gramática del español, una lengua con un sistema fonémico bastante diferente a las lenguas indígenas de México. El equipo de Watanaby y Jonathan Amith (Watanaby *et al.* 2021) ha notado que un transcriptor principiante de mixteco de Yoloxóchilt (YM), se le dificulta reconocer los tonos y paradas glotales de su lengua.

Creación de corpus en lenguas dominantes

Hay variantes de lenguas dominantes, como el español mexicano, que hasta hace poco carecían de *corpus* preparados especialmente para usarse para entrenar los modelos RAV. Sin embargo, este vacío lo pudieron remediar rápidamente, dado que en México todos los medios impresos y audiovisuales están pensados y hechos para una audiencia que solo habla en español. Por ejemplo, el dr. Carlos Hernández Mena, un informático de la UNAM que participó en el retiro del que trata este artículo, relató que colabora con estudiantes que realizan su servicio social obligatorio transcribiendo y editando los materiales (tales como telenovelas) que Hernández descarga de YouTube (Hernández Mena & Herrera 2017). Este proyecto le ha permitido a su equipo adquirir un enorme *corpus* de español mexicano en poco tiempo. Los investigadores de lenguas en peligro de extinción no tienen este mismo privilegio. Debido a estos obstáculos, el diálogo interdisciplinario entre lingüistas y científicos informáticos es una necesidad.

Barreras para la colaboración

Los científicos informáticos que desarrollan modelos RAV a menudo desconocen las dificultades y obstáculos para obtener *corpus* en las lenguas minorizadas. Por otro lado, los lingüistas, con frecuencia, no son conscientes de cómo preparar datos para modelos RAV, o muchas veces se rehúsan a compartir su *corpus* para estos objetivos.

De igual manera, tanto las industrias de alta tecnología, como los lingüistas computacionales, no desarrollan modelos RAV para lenguas minorizadas por la falta de *corpus* y porque estas labores no se consideran rentables. Mientras que los lingüistas computacionales a menudo citan la presión de obtener la permanencia universitaria como una razón para renunciar a la investigación sobre lenguas minoritarias. De cualquier forma, los conocimientos de RAV y otras áreas de PNL siguen siendo un conocimiento privilegiado en manos de unos cuantos individuos de compañías e instituciones elitistas del occidente. Lo poco que se ha escrito sobre las herramientas para "lenguajes de bajos recursos" en la literatura de la informática se centra en las necesidades de los lingüistas occidentales, con muy poca mención de los hablantes nativos y su papel en el avance de estas herramientas.

Tal pareciera que a muchos de los tecnólogos que actualmente están trabajando en esta área les preocupa más mantener su ego por ser los primeros autores de los papeles que publican, en vez de sentarse a hablar con las comunidades. Al parecer no hay mucho interés por saber si los artículos que publican o los modelos que desarrollan podrán ser leídos, entendidos y usados por los hablantes de estas lenguas. Dichas actitudes siguen perpetuando el colonialismo de la ciencia europea.

Las conversaciones colaborativas no han tenido lugar, en parte también, porque los hablantes nativos carecen de la influencia, la financiación y las conexiones para convocar a los investigadores con una visión común. La mayoría de las conversaciones sobre RAV tienen lugar en entornos formales, como conferencias, talleres y foros, en los cuales no son bien atendidos los hablantes nativos.

La necesidad de automatizar la transcripción de Chatino se hizo más urgente para esta autora cuando comenzó a transcribir grabaciones de audio que la doctora Lynn Hou, profesora asistente de lingüística en la Universidad de California en Santa Bárbara, hizo en San Juan Quiahije trabajando con familias de niños sordos. La doctora Hou también es sorda y, por lo tanto, depende de la transcripción e interpretación de cualquier lenguaje hablado que encuentre, ya sea inglés, español o Chatino.

La autora se esforzó en proporcionar a la doctora Hou anotaciones cuidadosas de los materiales de Chatino que Hou recopiló en San Juan Quiahije, para que pudiera tener datos fiables para analizar. La tarea fue extremadamente laboriosa, lenta y agotadora. La autora se encontró escribiendo las mismas palabras una y otra vez (*no*^{A} *qan*^{E} *lyuq*^{H} 'niña', *chaq*^{F} 'cosa, palabra', *na*^{F} *jin*^{C} 'mmm') y ella anhelaba poder automatizar el proceso. Empezó a preguntar a los lingüistas qué se necesitaría para automatizar la transcripción de Chatino. La mayoría le dijo que la RAV no era posible para las lenguas minoritarias porque carecían de grandes *corpus* para entrenar los modelos RAV.

Necesidad de automatizar el Chatino

Esta respuesta la llevó a colaborar con Damir y Malgorzata Cavar, de la Universidad de Indiana, en la creación del primer *corpus* de textos de Chatino para la formación de RAV (Cavar *et al.* 2016). La autora leyó en voz alta numerosos textos previamente transcritos en el chatino, en el sótano de la organización de *Linguist List*. Este *corpus* tiene una licencia *Creative Commons*, lo que significa que cualquiera persona lo puede descargar y está disponible en Recursos Abiertos Globales e Información para el Análisis del Lenguaje y la Lingüística (*GORILLA*).[3] Cuando la autora se convirtió en becaria de la fundación *NEUKOM* en *Dartmouth College*, propuso mejorar y expandir este *corpus*, mientras buscaba maneras de desarrollar RAV para Chatino y otras lenguas indígenas.

Este y otros *corpus* fueron incorporados al proyecto Corpora Lingüística Unificada de diversas fuentes de datos, manejados por el dr. Raphael Finkel, del Departamento de Ciencias de la Computación de la Universidad de Kentucky, y el dr. Daniel Kaufman de la Alianza de Idiomas en Peligro.[4] De igual manera, en el 2020 la autora, en colaboración con Aryaman Arora, estudiante de pregrado de George Washington, subió un *corpus* de 180 inflexiones verbales chatinas en *Wiktionary*.[5]

En otra parte del mundo, el lingüista Alexis Michaud, que trabaja, con Yongning Na, una lengua hablada en el suroeste de China, tuvo objetivos similares: automatizar el proceso de transcripción de las lenguas Na (Michaud *et al.* 2018). Como fonetista, Michaud diseñó sus grabaciones para que eventualmente pudieran ser utilizadas para el entrenamiento de RAV. Al igual que la autora, Michaud comenzó a indagar sobre posibles colaboraciones con científicos informáticos. Esta búsqueda lo llevó a entablar una exitosa colaboración con Oliver Adams, un científico informático con sede en la Universidad de Melbourne.

En el camino, Adams desarrolló un *kit* de herramientas RAV de código abierto llamado *Persephone* (Adams *et al.* 2018), que se basa en redes neuronales artificiales. Adams y su equipo comenzaron a producir resultados prometedores en la transcripción de lenguas Na. La herramienta estaba produciendo una tasa de error de 20%, y Michaud comenzó a desplegar *Persephone* en su proceso de trabajo lingüístico (Adams *et al.* 2018). También encontraron que *Persephone* podía alcanzar una precisión razonable para un solo orador con tan solo treinta minutos de datos, un signo auspicioso para las lenguas en alto riesgo de desaparición (Adams *et al.* 2018).

Posterior a ello, Michaud y Adams deseaban probar el modelo en un lenguaje tonal comparable, y encontraron su camino en el *corpus* de chatino en GORILLA que habíamos desarrollado con los Cavars. Me invitaron a evaluar los resultados de *Persephone* en chatino (Adams *et al* 2018). Para mi gran sorpresa, el sistema funcionó bien para el Chatino. El sistema obtuvo un error de 30% con tan solo 3.5 horas de grabación.

En este momento, *Persephone* solo es accesible para los científicos informáticos, lo que requiere una revisión de la interfaz para llegar a un público más amplio. Al ver los resultados de la comparación cruzada de Yongning Na y Chatino, esta autora se motivó a seguir buscando colaboraciones con especialistas en PNL para seguir avanzando y mejorando RAV para Chatino y otras lenguas indígenas de las Américas. Así comenzó la idea del retiro.

El retiro

La Fundación William H. Neukom en *Dartmouth College* financia grupos de trabajo interdisciplinarios para fomentar el diálogo recíproco entre participantes para avanzar en la investigación científica. A menudo organizados a través de "retiros", los investigadores con una misión común se reúnen en un lugar agradable e íntimo (generalmente el *Upper Valley* en New Hampshire y Vermont), para discutir soluciones a un problema o una pregunta que han estado meditando.

Daniel Rockmore, decano de ciencias de *Dartmouth College* y director del Instituto Neukom, animó a la autora a organizar un retiro para discutir el desarrollo de RAV para lenguas en peligro. Con el apoyo de Rockmore, ella procedió a invitar a científicos de la computación, lingüistas, hablantes nativos y activistas del idioma a reunirse en Quechee, Vermont, donde discutiríamos maneras de avanzar en el RAV para idiomas menos estudiados. El evento tuvo lugar del 12 al 14 de julio de 2018.

Las veinte personas que participaron en el retiro contaban con diversas experiencias, profesiones y provenían de diferentes instituciones. Participaron personas del Instituto John Hopkins, de la Universidad Carnegie Mellon, la Universidad de Yale, La Universidad del Norte de Texas, la Universidad de Texas en Austin, la Universidad Autónoma de México y el Centro de Investigaciones y Estudios Superiores en Antropología Social. La reunión marcó un equilibrio entre ingenieros, lingüistas, hablantes nativos y activistas de lenguas de alto riesgo de desaparición.

Los participantes hablaban o estudiaban idiomas de seis familias lingüísticas y de cuatro continentes: de Australia: Nyulnyulan (Bardi), Pama-Nyungan (Djambarrpuyngu, Djapu); de Africa: Masso (Burkina Faso); de la India: Tibeto-Birmania (Manipuri y otras 17); y de América: Otomangue (Chatino y Otomí), Maya (Tzetsal, Tzotzil y Mocho). Muchas de estas lenguas están gravemente en peligro. Las lenguas australianas, por ejemplo, tienen un promedio de cuatro a cinco hablantes, mientras que una de las dos variedades de lengua mocho tiene dos hablantes.

Los participantes se alojaron en un albergue de esquí llamado Nido del Búho. El ambiente fue íntimo, fami-

[3] https://gorilla.linguistlist.org/code/ctp/

[4] http://www.kratylos.org/

[5] https://en.wiktionary.org/wiki/Category:San_Juan_Quiahije_Chatino_verbs

liar y relajado. No hubo presentaciones de *PowerPoint*, la reunión se llevó a cabo en la sala de la casa. Durante los descansos, los participantes se dividieron orgánicamente en grupos mixtos y continuaron las conversaciones del día.

El primer día lo pasamos compartiendo intereses de investigación y estableciendo una agenda para el fin de semana. Tanto los lingüistas como los computólogos estaban ansiosos por conocer los métodos y la rutina de trabajo del otro; mientras que la mayoría de los lingüistas de campo se someten a un proceso iterativo de recopilación de datos y documentación, los científicos informáticos generalmente comienzan su investigación leyendo artículos académicos y luego replicando los métodos del articulo leído.

Antes de la cena, en el primer día, hicimos una excursión remando en canoas en el río Connecticut.

El segundo día estuvo marcado con descripciones más detalladas de la investigación de todos. Las discusiones se centraron en los recursos que teníamos a nuestra disposición para el entrenamiento de los modelos RAV, tales como los idiomas en los que habíamos colectado datos, los participantes en los audios o videos y las horas grabadas. Las preguntas que se plantearon en el retiro también iban enfocadas en la clarificación del *corpus* que hasta ahora había en las lenguas que estaban trabajando los participantes, incluyendo el formato de los audios (análogos o digitales) la calidad de los mismos y los tipos de equipos en los que éstos fueron colectados (p. ej. *Wax Cilinder*, casetes). De igual manera se habló de los tipos de archivos de los documentos (*Word*, PDF o *EAF*, *Excel*, *elan, toolbox, Flex*).

También se indagó sobre la información demográfica de los participantes en dichas grabaciones, la edad y el sexo, ya que idealmente el *corpus* de entrenamiento de los modelos RAV deben incluir un número diversos de hablantes. La fotografía en páginas siguientes, muestra las notas que tomamos el día de trabajo. Las conversaciones también se dieron en torno a si los *corpus* contaban con anotación, transcripción y traducción. Ya que como lo hemos mencionado, la gran mayoría de los materiales que hasta ahora existen en los archivos de lenguas carecen de esos elementos.

La escases de datos, se suma el hecho de que muchos de los materiales y documentos de legacía que existen en algunas lenguas indígenas, como resultado de estudios previos hechos en estas lenguas, se encuentran en formatos muy diversos y que muchas veces no se pueden interpretar (leer o escuchar) con tecnología y formatos contemporáneos. Muchas veces estos formatos son extremadamente delicados y muy caros de migrar a formatos contemporáneos, como el caso de las grabaciones hechas en *Wax Cilinders* al principio del siglo XIX.

Los primeros etnólogos y lingüistas que documentaron las lenguas indígenas en las Américas en el siglo XIX tales como Franz Boas y John Peabody Harrington, colectaron textos a mano, porque en aquel tiempo las grabadoras no eran comunes. De igual manera, muchos de los materiales están escritos a mano, los cuales son muy difíciles de automatizar y de migrar a formatos digitales contemporáneos. Para poder migrar documentos escritos a mano a formatos digitales, es necesario hacerlo a mano. Un ejemplo de esto es el esfuerzo que Ryan Sullivant, del Archivo de Lenguas Indígenas de Latinoamérica, está llevando a cabo para transcribir encuestas colectadas a mano por Kathryn Josserand en el mixteco de Oaxaca a través de voluntarios de *crowdsourcing*.

Otra complejidad de los textos escritos a mano es que muchas veces no se pueden leer porque las letras de los creadores son ilegibles. Éste es el caso de los materiales que colectó John Peabody Harrington en lenguas indígenas de California, tales como Hupa, Karuk, and Mutsun, entre muchas otras (Warner *et al.* 2006). Mientras que Harrington fue un gran fonólogo, su caligrafía era pésima y los materiales que dejó en estas lenguas son muy difíciles de descifrar por hablantes que están trabajando por revivir sus lenguas (Warner *et al.* 2006: 260).

Muchos de los textos de legacía están escritos en diversas ortografías. Muchas veces estas ortografías fueron creadas por las mismas personas que escribieron dichos textos. Muchos de ellos no proveyeron una clave para poder leer dicha ortografía, un ejemplo son los textos escritos por Hilario Canseco en el Chatino de San Juan Quiahije, presentados en el texto the Carmen Cordero de Duran (Cruz 2014).

Algunos miembros de comunidades indígenas de Norteamérica, quienes están empezado a integrar el RAV a sus lenguas, han tenido que hacer mucho trabajo para sistematizar y regular los documentos de legacía antes de poder integrarlos a los *corpus* para entrenar los modelos RAV. Un caso concreto es de Kwk'wala, una lengua indígena hablada en British Columbia (Roland *et al.* 2020).

Estas conversaciones fueron sumamente importantes para poder planear y reflexionar sobre los materiales que los modelos RAV pudieran usar para avanzar en la transcripción de las lenguas minorizadas.

El segundo día se concluyó con una visita al Instituto de Ciencias Naturales de Vermont, un centro de educación de rapaces y santuario de aves.

Esfuerzos actuales en el ámbito de las tecnologías RAV

Como lo hemos mencionado, los modelos RAV son una tecnología ya bastante madura en las lenguas mayoritarias, cada día hay un mayor interés en esta área de investigación. Parte del entusiasmo por esta tecnología se debe al impulso que le han dado los productos comerciales, tales como la búsqueda por voz de Google, Alexa de Amazon, Siri de Apple. Aunado a esto está la gran actividad de códigos abiertos que están surgiendo para esta área de investigación, tales como *Kaldi, HTK* , *Sphinx, Julius, RASR* y últimamente ESPnet (Watanabe *et al.* 2018).

También está creciendo el interés de parte de los computólogos por integrar las tecnologías RAV y otras tecnologías de PNL a las lenguas indígenas de las Américas. Por ejemplo, hay un equipo entusiasta trabajando con *machine learning* y la creación de *corpus* en lenguas Indígenas de las Américas (Mager 2018).

Además, hay varios equipos de diferentes universidades e instituciones del mundo contribuyendo a estos esfuerzos por integrar el RAV a las lenguas minorizadas. Dichos trabajos se están enfocando tanto en la creación de sistemas RAV como también en *API web* (interfaz de programación de aplicaciones), para que los sistemas *Kaldi* y *ESPnet* puedan ser utilizados por un público más amplio.

Allosaurus

Una de estas iniciativas es el Reconocimiento Fónico Universal con un Sistema Alofónico Universal (*Allosaurus*) (Li *et al.* 2020). Ésta es una iniciativa de miembros de la Universidad de Carnegie Mellon y tiene como objetivo aprovechar los modelos multilingües para la transcripción fonética del lenguaje. El sistema tiene como objetivo modelar conjuntamente los *fonos* independientes de lengua al igual que los fonemas dependientes de la misma. Para poder hacer uso de este sistema, los usuarios(as) primero tendrán que insertar el sistema fonético de la lengua que desean transcribir al sistema. Una vez hecho esto, el sistema estará listo para arrojar una transcripción muy preliminar de la lengua. Desarrolladores esperan aliviar el problema de embotellamiento que ahora es una traba en las lenguas minorizadas. Los desarrolladores argumentan que Allosauros está pre-entrenado para reconocer más de 2000 idiomas hasta ahora.

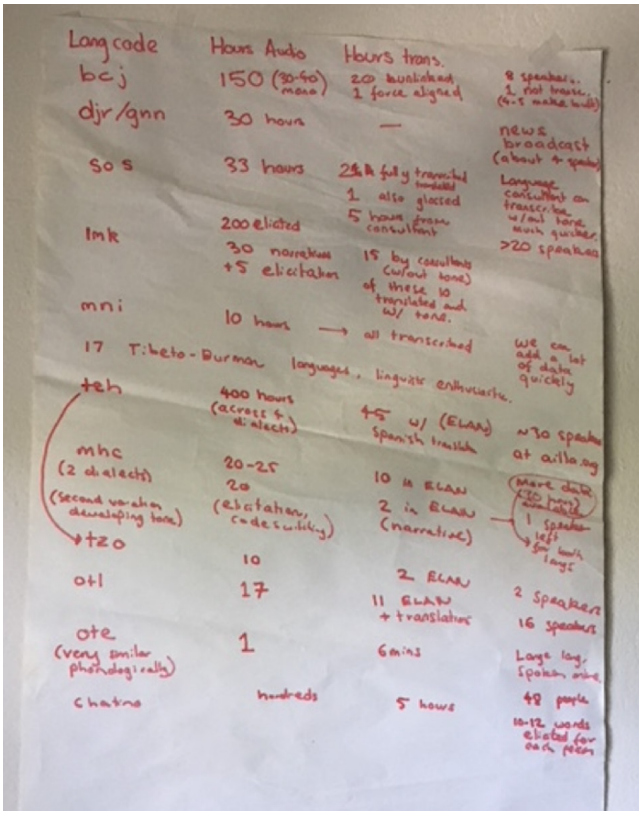

Figura 1. *Notas que se tomaron en el retiro. Fotografía de Hilaria Cruz.*

Transcriptor principiante (mixteco de Yoloxóchitl y *end-to-end* ASR

Otro equipo colaborativo conformado por los computólogos Shinji Watanabe, Jiatong Shi, y otros de Carnegie Mellon, al igual que Jonathan Amith del departamento de antropología en Gettysburg College, y Rey Castillo un lingüista y hablante de la lengua del mixteco de Yoloxóchitl, han comenzado a experimentar la efectividad del modelo RAV de extremo-a-extremo (*end-to-end* ASR) en el YM, una lengua Otomangue hablada en el estado de Guerrero, México. Este mismo equipo también está empezando a experimentar con el náhuatl de Puebla (Watanabe *et al.* 2021).

Ante la escases de transcriptores competentes que desarollen un *corpus* robusto para entrenar los sistemas RAV en lenguas minorizadas, estos investigadores reportan un experimento empírico que llevaron a cabo comparando la transcripción de un nativo hablante, sin experiencia en escribir en su lengua, con la de un sistema RAV *end-to-end* ASR en MY. Su objetivo era calcular el tiempo y el costo de entrenar estos dos recursos: el humano (principiante) y el sistema RAV. Para los dos contextos usaron una misma *corpora*, libre de ruido, de 8.36 horas.

Para la parte principiante, contrataron a Esteban Guadalupe Sierra, un hablante de MY, quien no tenía experiencia previa en escribir su lengua. Los investigadores observaron el tiempo que les tomo alfabetizar al hablante así como también los errores que éste incurría en su transcripción. Después de un año transcribiendo, a medio tiempo, se dieron cuenta que Guadalupe Sierra todavía tenía muchos problemas distinguiendo algunos de los segmentos (paradas glotales), tonos, y la morfología de su lengua. De igual manera se le dificultaba separar con el símbolo [=] los enclíticos, los prefijos con un guion y poner entre paréntesis los tonos elididos en la forma subyacente. Con el transcriptor principiante obtuvieron una tasa de error (CER) de 6%.

Con el mismo *corpus* entrenaron el sistema RAV *end-to-end*. Con este obtuvieron un CER de 8.2%. Aunque el transcriptor principiante (humano) obtuvo mejores resultados que el sistema de RAV, el sistema RAV pudo distinguir los tonos, las paradas glotales y de igual manera pudo separar con el símbolo [=] los enclíticos, separar los prefijos con un guion y poner entre paréntesis los tonos elididos en la forma subyacente de una manera mucho más consistente que el principiante. Los investigadores también se dieron cuenta de que mientras más datos se le inyectaba al sistema RAV, éste arrojaba mejores resultados. Esto se resume en el cuadro 1.

Figura 2. *Colaboradores en el retiro, de izquierda a derecha: Nestor Hernández Green, Carlos Hernández Mena, Joseph Waring, Zach Yu-Hsiang Lin, Sarah Gupta, Hilaria Cruz, Dolly Goldenberg, Oliver Adams, Isabelle Strong, Laura McPearson, Shalui Abeles, Emiliana Cruz, Jaime Pérez Gonzáles, Claire Bowern, Shruti Rijhwani, Shobhana Chelliah, Manuel Mager, Sol Aréchiga Mantilla. Fotografía de Michael Abramov.*

Cuadro 1. Trancripcion del humano principiante con la del RAV end-to-end en YM.

	CER con transcriptor humano principiante	CER con RAV (E2E-conformer)
Con audio limpio (clean-dev set)	6%	8.2%

La reflexión de los resultados del experimento llevó al equipo a proponer que una manera de aminorar la escases de *corpus* en la lenguas en alto riesgo de desaparición es que un transcriptor humano trabaje conjuntamente con el sistema RAV para transcribir audios. Si se sabe desde el principio que el sistema RAV se utilizará para corregir transcripciones de un transcriptor principiante en áreas donde éste encuentra dificultades, bien se pudiera entrenar al sistema RAV en aquellas áreas que desafían el aprendizaje de los principiantes, tales como los tonos o las paradas glotales, como lo vimos en el MY .

Elpis

También hay iniciativas para crear interfaces que faciliten el uso de las tecnologías RAV por un grupo más amplio de usuarios con mínimo conocimiento o experiencia en computación. Ésta es otra posibilidad para facilitar la obtención de *corpus* para entrenar los sistemas RAV en lenguas minorizadas.

El equipo Elpis está conformado por diversos investigadores de diferentes instituciones educativas de diversos países, tales como la organización de Lenguas y Civilizaciones con Tradición Oral de Francia, las universidades Australianas de Queensland y Brisabane, y la Universidad de Estudios Orientales y de África en Londres (Adam *et al.* 2021), las cuales están creando una interfaz gráfica que permitirá al usuario construir su propio modelo de reconocimiento y, de esta manera, puedan transcribir automáticamente su audio (Foley *et al.* 2018). Elpis fue originalmente diseñado para usarse con el *toolkit* de reconocimiento automatic Kaldi, pero en su última generación están haciendo una interfaz para la integración *ESPnet*.

Experiencia multidisciplinaria

Las experiencias multidisciplinarias en el retiro fueron extensas y muy diversas. Ésta es una de las primeras veces que los computólogos interesados en integrar los sistemas RAV en las lenguas en alto riesgo de desaparición se encontraron (en un ambiente relajado) con lingüistas documentales para conocerse, dialogar, intercambiar ideas y experiencias, y ver posibilidades de futuras colaboraciones. Ésta es tambien la primera vez que un evento de esta naturaleza es convocado por una hablante nativa de una lengua en alto riesgo de desaparición. Históricamente a los nativos hablantes se les ha negado el derecho a alfabetizarse y aprender de la gramática de sus lenguas. Como resultado, las lenguas minorizada están subrepresentadas y no tienen recursos en la academia, en la industria o instituciones del Estado.

Esta autora, por ejemplo, pudo darse cuenta de que el RAV es una pequeña parte dentro de muchas áreas de investigación dentro del área del PNL. Por ejemplo, en

el retiro participaron especialistas en *Optical Character Recognition* (Shruti Rijhwani) y *machine translation* (Manuel Mager). Estos investigadores invitaron a los participantes a que diversificaran sus intereses en otras áreas de PNL. Este conocimiento le permite a la autora canalizar las preguntas hacia donde desea llevar su investigación.

El retiro también permitió a los participantes dialogar abiertamente sobre cómo sería un sistema de RAV que pudiera ser de utilidad para aliviar el problema del embotellamiento en las lenguas minorizadas. Los lingüistas documentales expresaron que los modelos deberán ser accesibles a los usuarios que no son muy adeptos a la tecnología; idealmente el sistema debería permitir al usuario subir su audio al sistema RAV en su ordenador, este sistema deberá ser capaz de transcribir el contenido de manera independiente. También dijeron que los sistemas RAV deben, idealmente, poder usarse en diferentes plataformas y ser de acceso libre, sin dejar de lado que los participantes expresaron su deseo de que fueran herramientas que pudieran usarse en contextos donde no hay acceso a internet.

Consideraciones finales

El evento fue un éxito rotundo. El encuentro fue productivo y agradable. Se forjaron amistades y se hicieron planes de futuras colaboraciones y retiros. Los participantes se fueron muy ansiosos por continuar la conversación sobre cómo integrar el RAV para los idiomas en alto riesgo de desaparición.

Reafirmamos la urgencia de integrar las herramientas RAV para acelerar la transcripción de lenguas en alto riesgo de desaparición, ya que, como lo notaron varios participantes del retiro (p. ej. Jaime Pérez González con el Mocho Maya, una variante de la que solo quedan dos hablantes), hay lenguas que ya casi no tienen hablantes. Algunas de las conclusiones a las que arribamos en el retiro fueron que deberíamos buscar la existencia de herramientas de voz que ayuden a expeditar la transcripción de las lenguas en peligro.

Cuando el retiro se llevó a cabo, la integración de los sistemas de RAV a las lenguas minorizadas apenas comenzaba a tomar auge. Pero los trabajos que están surgiendo en esta área, tales como el equipo que trabaja con YM y el Nahuatl de Puebla (Watanabi *et al.* 2021), así como los esfuerzos del equipo Elpis para hacer más accesible el RAV a un grupo más grande de usuarios (Adams *et al.* 2021), están avanzando, clarificando e informando la metodología y las teorías de RAV en lenguas minorizadas. A través de dichos esfuerzos se están realizando experimentos para superar los obstáculos de trabajar con lenguas que tienen escasos *corpus*, hablantes, transcriptores y preparación técnica.

Como se dijo previamente, actualmente los modelos RAV todavía requieren de mucha experiencia computacional para ser utilizados por un publico más amplio. El éxito de estos modelos depende mucho del financiamiento, seguimiento y mantenimiento de los mismos. Idealmente tiene que haber toda una infrastructura detrás de cada sistema. Sin embargo, no existe ningún obstáculo teórico para interfaces que permitan acelerar el proceso de anotación y transcripción en lenguas minorizadas. Esta labor se trata, en gran medida, de que un ingeniero de *software* profesional desarrolle dichas herramientas. Esto nos llevó a concluir que era importante conseguir financiamiento para crear una interfaz para el *Persephone* y, de esta manera, facilitar su uso para las personas que no son computólogos. Las tecnologias RAV, al igual que otras tecnologías, cambian rápidamente. De acuerdo con Adams *et al.* (2021), *Persephone* no ha podido conseguir un API web (interfaz de programación de aplicaciones), así que *Persephone* ya no tiene mucha viabilidad porque no tiene técnicos que le sigan dando mantenimiento (Adam *et al.* 2021).

Algunas de las conversaciones que han surgido a raíz de los experimentos que se han desarrollado en esta área incluyen la calidad de *corpus* que se necesita para echar a andar los modelos de RAV, así como la manera de representar los préstamos de las lenguas homogenizantes, como el español, en estas lenguas. Las discusiones también se centran sobre la mejor manera de sistematizar y preparar los documentos de legacía para que estos puedan ser utilizados para entrenar los sistemas de RAV más eficazmente.

Por ejemplo, el equipo de trabajo de YM (Watanabe *et al.* 2021) reporta que *ESPnet end-to-end* obtiene muy buenos resultados con 50 horas de *corpus*. Mientras que el equipo de Elpis confirma que, cuando hay una sola persona hablando con audios que no tienen mucho ruido, se obtienen mejores resultados en el RAV. Éstos son muy buenos indicios, ya que cuando esta autora, en 2015 le preguntaba a los computólogos cuántas horas de *corpus* se necesitaban para entrenar a los modelos RAV, nadie podía contestar con certeza, porque estos experimentos apenas comenzaban a hacerse.

Los hablantes nativos deben ser incluidos en conversaciones, diseños, e implementación de los sistemas RAV, ya que aportan perspectivas orientadas a la comunidad y la responsabilidad hacia la misma. Éstos son temas que muchas veces son pasados por alto o ignorados por lingüistas externos y científicos informáticos.

Algunas accciones concretas que las instituciones educativas pudieran tomar para mejorar este problema es incluir a nativos hablantes en los laboratorios de ingenierías lingüísticas y PNL que abundan en los centros computacionales en la actualidad. Los departamentos de lingüística tambien podrían contratar a especialistas de PNL que no neceriamente tienen que ser lingüistas. Por último, las instituciones educativas deben dar crédito a los investigadores que trabajan para preparar *corpus*.

Agradecimientos

Agradezco a José Luis Hernández Jiménez, al dr. Fidel Hernández Mendoza, a Andrés Pérez Pérez y al dr. Javier Flores Gómez por la revisión y edición del artículo. De igual manera deseo agradecer infinitamente al Instituto Neukom y el Departamento de Lingüística en Dartmouth College, a los dictaminadores anónimos, a Michael Abramov y a tod@s l@s que participaron en el retiro.

Referencias

Adams, O., T. Cohn, G., Neubig T., Hilaria Cruz, S. Bird y A. Michaud. (2018). Evaluating phonemic transcription of low-resource tonal languages for language documentation. Proceedings of Language Resources and Evaluation Conference (LREC) 2018. Disponible en https://halshs.archives-ouvertes.fr/halshs-01709648 [Consulta: mayo de 2021].

Bird, S. (2020). Decolonising Speech and Language Technology. Proceedings of the 28th International Conference on Computational Linguistics.

Ćavar, M. E., Cavar, D. y Cruz, H. (2016). Endangered Language Documentation: Bootstrapping a Chatino Speech Corpus, Forced Aligner, RAV. *LREC*, 4004-4011.

Cruz, H. (2014). Linguistic Poetics and Rhetoric of Eastern Chatino of San Juan Quiahije. Unpubl. Tesis. Austin: University of Texas at Austin.

Hernández Mena, C. D. y Herrera, A. (2017). Corpus de Investigación en Español de México del Posgrado de Ingeniería Eléctrica y Servicio Social Light. *Linguistic Data Consortium*. Disponible en https://catalog.ldc.upenn.edu/LDC2017S23 [Consulta: mayo de 2021]. Kuhn, R., Davis, F. Alain Désilets, Eric Joanis, Anna Kazantseva, Rebecca Knowles, Patrick Littell, Delaney Lothian, Aidan Pine, Caroline Running Wolf, Eddie Santos, Darlene Stewart, Gilles Boulianne, Vishwa Gupta, Owennatékha Brian Maracle, Akwiratékha' Martin, Christopher Cox, Marie-Odile Junker, Olivia Sammons, Delasie Torkornoo, Nathan Thanyehténhas Brinklow, Sara Child, Benoît Farley, David Huggins-Daines, Daisy Rosenblum, and Heather Souter. "The Indigenous Languages Technology project at NRC Canada: An empowerment-oriented approach to developing language software". In Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020), Disponible en https://www.aclweb.org/anthology/2020.coling-main.516/ [Consulta: 13 de diciembre de 2020].

Li, Xinjian, Dalmia, Siddharth, Li, Juncheng, Lee, Matthew, Littell, Patrick, Yao, Jiali, Anastasopoulos, Antonios, Mortensen, David R., Neubig, Graham, Black, Alan W, & Metze, Florian. Universal Phone Recognition with a Multilingual Allophone System. *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing*. Disponible en: https://par.nsf.gov/biblio/10175774.

Mager, M., Gutierrez-Vasques, X., Sierra, G., Meza, I. (2018). Challenges of language technologies for the indigenous languages of the Americas. *arXiv* 1806.04291

Michaud, A., Adams, O., Cohn, T. A., Neubig, G. y Guillaume, S. (2018). Integrating automatic transcription into the language documentation workflow: experiments with Na data and the Persephone toolkit. *Language Documentation and Conservation* 12. 393-429.

Mithun, M. (1998). The significance of diversity in language endangerment and preservation. L. Grenoble y L. Whaley (Eds.), *Endangered languages: Current Issues and Future Prospects* (pp. 163-191). Cambridge: Cambridge University Press.

NLP - Linguistic Resources - Tutorialspoint. (2021). Tutorials Point. Disponible en: https://www.tutorialspoint.com/natural_language_processing/natural_language_processing_linguistic_resources.htm [Consulta: 20 de mayo de 2021].

Warner, N., Butler, L. y Luna-Costillas, Q. (2006). Making a Dictionary for Community Use in Language Revitalization: The Case of Mutsun. *International Journal of Lexicography*, 19 (3), 257-285.

Watanabe, S., Hori, T., Shigeki K., Hayashi, T., Nishitoba, J., Unno, Y., Enrique, N., Soplin, Y., Heymann, J., Wiesner, M., Chen, N., Renduchintala, A. y Ochiai, T. (2018). *End-to-End Speech Processing Toolkit. arXiv* 1804.00015.

Woodbury, T. 2003. Defining Documentary Linguistics. In Austin, P. K. 2003 (Ed.). Language documentation and description. Londres: Hans Rausing Endangered Languages Project.