

**La colaboración digital abierta para la transcripción de textos  
manuscritos de idiomas indígenas**

**J. Ryan Sullivant**  
University of Texas at Austin

*Memorias del VIII Congreso de Idiomas Indígenas de Latinoamérica,  
26-28 de octubre de 2017, Universidad de Texas en Austin*

*Disponible en la página de AILLA*

# La colaboración digital abierta para la transcripción de textos manuscritos de idiomas indígenas \*

**J. Ryan Sullivant**  
University of Texas at Austin  
sullivant@austin.utexas.edu

**Palabras Claves:** Mixteco, transcripción, crowdsourcing, manuscrito.

## 1. Introducción

Hoy en día, la investigación lingüística necesita textos digitales más y más. La utilidad y necesidad de textos digitales aparecen en varios campos de la disciplina: los textos mismos son el objeto de investigación para el procesamiento de lenguaje natural y otros estudios de la morfosintaxis de un idioma. Para otros investigadores, los textos digitales son imprescindibles para poder encontrar candidatos de análisis como los enunciados de ciertos variables sociolingüísticos o las porciones de una grabación donde aparecen ciertas secuencias consonánticas.

La práctica actual de la mayoría de investigadores lingüísticos es crear un texto digital en la primera etapa de transcripción de un texto grabado, ya sea en un documento de texto sencillo o mediante una aplicación como ELAN (Sloetjes y Wittenburg, 2008 i.a.). Esta práctica es bastante reciente en la historia de la investigación lingüística y todavía existe una gran cantidad de documentos que contienen transcripciones, traducciones y/o análisis de textos en idiomas indígenas que carecen de una versión con texto digital. En la Fig. 1, se puede ver unos ejemplos de textos manuscritos de los 1960 hasta 2000 que han sido digitalizados (en este caso escaneados) y preservados en un repositorio digital, el Archivo de los Idioma Indígenas de Latinoamérica (AILLA).<sup>1</sup> Se puede ver que constan de transcripciones bastante estrechas (1a) hasta transcripciones en una ortografía práctica (1b-e). Muchos documentos así tienen traducciones a otro idioma (en el contexto latinoamericano suele ser español o portugués, a veces acompañado con inglés) (1a-d) pero a veces sólo existe la transcripción misma (1e). Como estos textos contienen muchos datos de los idiomas (y seguro que unos lamentablemente ya no hablan) y replicar su transcripción y/o traducción desde cero sería sumamente costoso, vale la pena investigar cómo podemos crear textos digitales de ellos.

Los textos manuscritos tienen varios obstáculos que estorban la fácil preparación de textos digitales correspondientes. Con la excepción de textos tipografiados, el reconocimiento óptico de caracteres (conocido como OCR por sus signos en inglés) es casi imposible. De hecho, los documentos tipografiados que tenemos en idiomas indígenas latinoamericanos suelen ser difíciles para las soluciones más comunes de OCR. Por

---

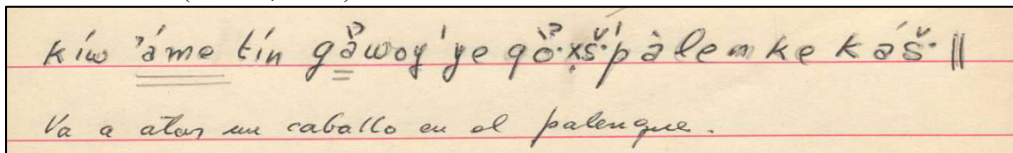
\* Agradezco a Tony Woodbury y Susan Kung, por su ayuda en varios aspectos de este reporte.

<sup>1</sup> <https://ailla.utexas.org/>

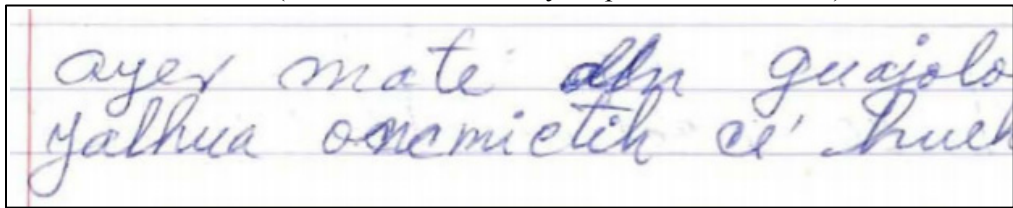
ejemplo, un texto representativo, como el de Fig. 2, sería difícil porque el OCR funciona mejor con idiomas y fuentes bien conocidos—muchas aplicaciones dependen de información acerca de las palabras y combinaciones de caracteres comunes para un idioma. Además, los documentos suelen ser plurilingües, como en Fig. 2 donde la primera columna es en triqui y en la segunda aparece su traducción al español. Ha habido avances en el OCR para tales documentos (e.g. Albert-Abrams, 2016), pero es importante notar que esos métodos exigen grandes cantidades de textos digitales para generar los modelos de idiomas que informan el OCR—un prerrequisito inalcanzable para la mayoría de idiomas indígenas latinoamericanos.

Figura 1. Transcripciones manuscritas de idiomas indígenas en un repositorio digital.

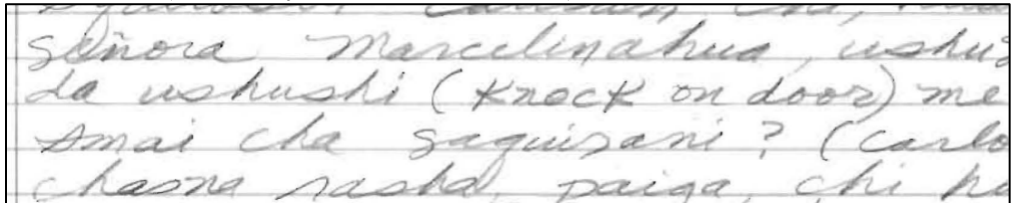
a. Tehuelche (Suárez, 1968)



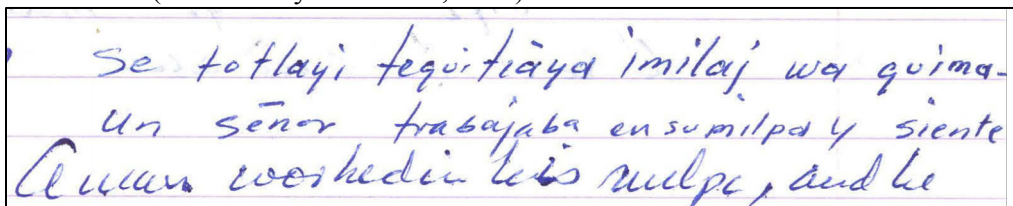
b. Náhuatl del centro (Hill, Hill, Casa Vélez y Zepeda Serrano, 1975)



c. Quechua (Nuckolls y Anachu Cadena, 1988)



d. Náhuatl (Sandstrom y Sandstrom, 1990)



e. Sharanahua (Déléage, Martínez Tello y Mélenhez Mélenhez, 2002)

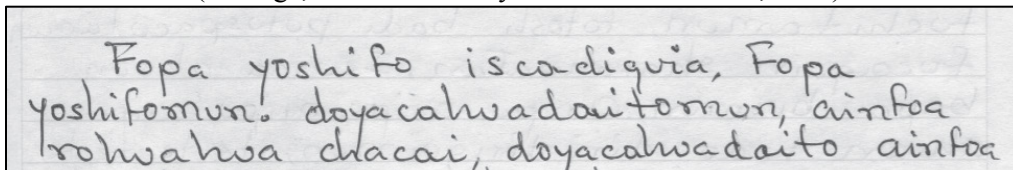


Figura 2. Un texto tipografiado de un idioma indígena (Belmar 1987)

	13	
Nu nghui tsiiahu?	Qué hace?	
Kata ghue ne naha.	Dile que estoy acá.	
Kata kunén rua kamiha.	Dile que le quiero hablar.	
Kata kunén dughéten trianghue.	Dile que se asome á la puerta.	
Andá kaá ha?	A dónde fué?	
Amá kanaha?	Cuándo vendrá?	
Kana rañianni?	Vendrá pronto?	
Kanghui kanaghui ma- ni?	Le podré aguardar?	
Anda naríhi maha?	A dónde lo hallaré?	
Kana tinén.	Vendrá á la tarde.	
Kana nihini.	Vendrá á la noche.	
Kana hayoni.	Vendrá mañana.	
Kana guanigán.	Vendrá temprano.	
	2.	
Nunghui tsiganaha?	Por qué no vienes?	
Nunghui tsi gánso?	Por qué no vas?	
Nunghui go ya tsunso?	Por qué no trabajas?	
Nunghuigo ka kia tsun- so?	Por qué no vas á traba- jar?	
Nunghui tsi gunama?	Cómo se llama eso?	
Nu gunaso?	Cómo te llamas?	
Anda nghuiso?	De dónde eres?	
Ama kuchumare?	Cuándo llegaste?	
Rua kia sunreni?	Quieres trabajar?	
Ruare tighiñiureni?	Quieres aprender oficio?	
Ruare kuna kichire ria- ni.	Quieres servirme?	
Kúrughue dughuere da- da ioho, dada yaghu- hi.	Te pagaré cada año, ca- da mes.	

El OCR con textos manuscritos se enfrenta con muchos retos. Por uno, a diferencia de textos tipografiados, los caracteres manuscritos (como los de Fig. 1a) pueden tener formas diferenciadas (Fig. 3), y letra corrida (Fig 4) que son problemáticos para la mayoría de aplicaciones OCR. Los textos manuscritos en idiomas indígenas tienen aún más dificultades debido al frecuente uso de símbolos que son poco comunes en los idiomas más estudiados en el mundo, y un autor puede combinar muchos signos y diacríticos en un alfabeto práctico o para indicar algún aspecto fonético de un enunciado (Fig. 5).

Figura 3. Diferencias en forma de letra.

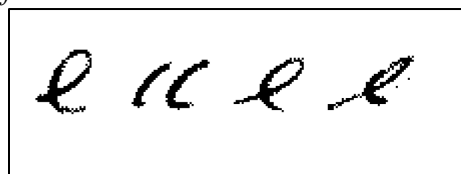


Figura 4. *Letras conectadas*

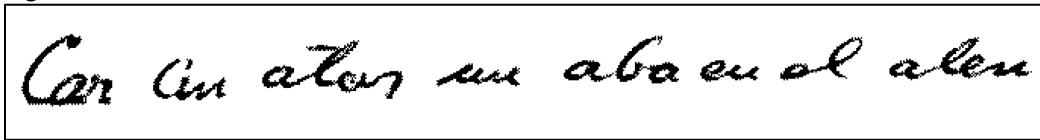
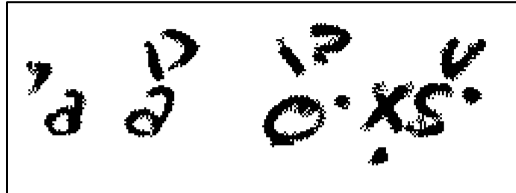


Figura 5. *Símbolos y combinaciones de símbolos poco comunes.*



Si el reconocimiento de caracteres manuscritos sería tan difícil, ¿cómo se puede producir textos digitales? La manera más utilizada hasta la fecha es la transcripción: un transcriptor crea un texto digital signo por signo por un teclado con referencia al texto manuscrito. Ha habido unos proyectos que trabajan en idiomas latinoamericanos que sí aprovecharon de este método (Lillehaugen et al., 2016; Matthew, 2017). Pero aquéllos trataban de documentos coloniales escritas en idiomas indígenas bastante estudiados (el zapoteco colonial y dos idiomas náhuatl), y los integrantes de los proyectos o ya eran expertos en los idiomas y la paleografía de la época o estaban en proceso de adquirir el conocimiento de éstos. Por muchas colecciones de documentos manuscritos, no hay un cuerpo de especialistas tan numerosas y motivadas para sostener un proyecto de transcripción experta. La pregunta que surge de esta situación (y el enfoque de este reporte) es ¿son necesarios los especialistas para la transcripción de textos manuscritos en idiomas indígenas? O más general, podemos preguntar ¿puede uno transcribir un texto manuscrito en un idioma que no comprende?

## 2. El cuestionario dialectal del mixteco

Decidí investigar la pregunta por tratar de transcribir unos textos del cuestionario dialectal del mixteco.<sup>2</sup> Entre 1978 y 1980, Kathryn Josserand y sus colaboradores se dedicaron a la recopilación de respuestas a un cuestionario dialectal de los idiomas mixtecos para estudiar la diversificación y desarrollo histórico de los idiomas mixtecos (Josserand 1983). El cuestionario consta de palabras y frases castellanas diseñadas para elicitación de ciertas estructuras gramaticales mixtecas. Grabaron respuestas al cuestionario en casete en 137 lugares distintos por todo el oeste del estado de Oaxaca en México y algunos lugares en los estados de Puebla y Guerrero. Los casetes fueron digitalizados y están disponibles en una colección pública en AILLA (Josserand s/f). También se hicieron

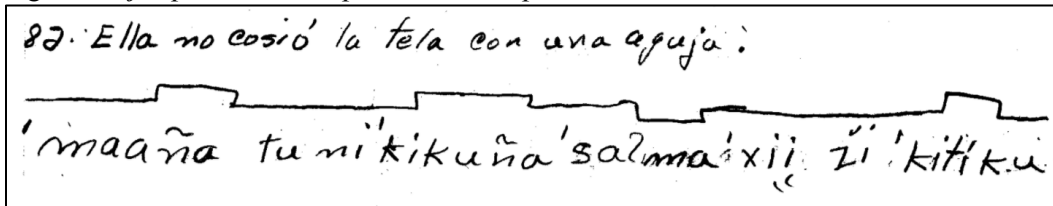
---

<sup>2</sup> A pesar del uso de “dialectal” aquí y en la obra de Josserand (p.ej. *Mixtec Dialect History*), el mixteco no es una sola lengua, sino una familia de diversos idiomas, y la diversidad interna de varios de estos hace difícil saber con certeza cuántos idiomas mixtecos hay. Por el número de “zonas dialectales” que identifica Josserand, tenemos 12 como un número mínimo (1983). El conteo más alto que conozco es 52 (Simons y Fennig, 2017).

transcripciones manuscritas de las respuestas, pero solamente una pequeña porción de los datos se publicó (Josserand 1980, 1983).

Como pueden ver en la Fig. 6, las transcripciones fueron hechas con una letra bastante clara, pero con los símbolos fonéticos de su tradición lingüística que son infrecuentes en muchos alfabetos (e ignoro uno en que aparece toda la gama). Por ejemplo, el apóstrofe indica énfasis lexical, <ʔ> el cierre glotal, y el gancho polaco debajo de una vocal <ɥ> señala su calidad nasal. También, muchos transcripores agregaron una línea arriba del enunciado mixteco para señalar su impresión de las subidas y bajadas de tonía en estos idiomas tonales. La transcripción en la Fig. 6 se puede reproducir y analizar como en (1a). Una transliteración de esta oración según las normas del Ve'e Tu'un Savi (Guadalupe Joaquina, 2017) aparece en (1b).

Figura 6. Ejemplo de transcripción de una respuesta al cuestionario dialectal de Josserand.



- (1) a. 'maña tu ni'kikuña 'saʔma 'xij ži 'kitiku  
 SPEC.3FEM NEG CPL.sew.3FEM ropa PREP aguja  
 'Ella no cosió la tela con una aguja.'<sup>3</sup>

- b. *Maa-ña tu nikikú-ña sa'mà jiiin yiki tiku.*

Las transcripciones de Josserand entonces son buenos candidatos para la transcripción no experta. La letra es muy clara y moderna, los documentos son bastante estructurados y el conjunto de símbolos es relativamente común. Pero, como un corpus, las transcripciones todavía tienen retos para la transcripción no experta. Más que una mano transcribió las respuestas, y es posible que un transcriptor pueda tener dificultades en interpretar unas manos más que otras. También, la representación del tono presenta dificultades..

En lenguas tonales, como los mixtecos, el tono es relativo, y son las diferencias en tonía que son importantes. Los escalones claramente indican bajadas y subidas, pero ¿cómo podríamos interpretar si una bajada es del nivel alto a medio, alto a bajo, o medio a bajo. El problema se complica aún más si haya cuatro niveles tonales en un idioma mixteco como se ha reportado por el mixteco de Yoloxóchitl entre otros (DiCanio y Bennett 2017). Además, hay cuestiones de cómo alinear los tonos con las vocales, y diferenciar entre tonos de contorno y secuencias de tonos nivelados.<sup>4</sup> Para no detenerme en estos problemas, limité

<sup>3</sup> Abreviaturas usadas: CPL = aspecto completivo; FEM = femenino; NEG = negativo; PREP = preposición de varias funciones; SPEC = demostrativo específico.

<sup>4</sup> La mora es el portador del tono en casi todas las descripciones fonológicas de idiomas mixtecos. Ya que Josserand escribe vocales largas con dos vocales (p.ej. <xij> /hí:/) cada vocal escrita equivale a una mora.

la tarea para este proyecto a la transcripción de un solo cuestionario y dejé al lado la representación tonal.

### 3. La prueba

Para determinar si la transcripción no experta de un idioma indígena latinoamericano fuera posible, extendí una oportunidad de transcribir uno de los cuestionarios de Josserand a una clase de estudiantes universitarios. Estos estudiantes asistían a una clase de interés general dirigida a alumnos del primer año universitario. Es decir que a pesar de haber inscrito en una clase que trataba de temas lingüísticos, los alumnos no habían recibido instrucciones en la transcripción lingüística. Ninguno sabía un idioma mixteco. Por fines del curso, cada alumno tuvo que escoger uno de cuatro idiomas para luego redactar unos ensayos sobre él, y entre las opciones fue el mixteco de Chalcatongo.

Hice un formulario en el servicio Google Forms (Fig 7.) para recolectar las transcripciones de los alumnos del cuestionario de Chalcatongo. También les pedí información acerca de cuál fue el idioma escogieron para lectura adicional (para determinar si haya un efecto de un poco de familiarización con las formas del idioma) y cuánto tiempo tardaron en hacer la actividad.

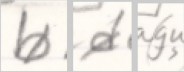

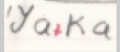
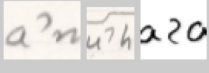

Figura 7. *El formulario para recolectar transcripciones.*

The image shows a screenshot of a Google Form titled "Chalcatongo Mixtec Transcription". The form is titled "Part 2" and contains instructions for transcribing Mixtec text. The instructions state: "You will see images of the original pages from the second portion of the Mixtec survey at this address: <https://drive.google.com/open?id=0B6LpL-WFZ7UscEVoMIQtdWRLTUE>". It also provides a link to a Google Docs document for transcription conventions: <https://docs.google.com/document/d/1xYWo-yJQq4SrRg-7p4sTh9RITjYr6vl6lY6KAKvLDGM/edit?usp=sharing>. The form includes a question: "39. Fue su yerno de ella dentro de su casa. [Her son-in-law went in her house.]" and a text input field labeled "Your answer".

Las instrucciones claramente indicaron que los alumnos únicamente tenían que transcribir las respuestas mixtecas a cada ítem del cuestionario. Mediante los enlaces en el

formulario, los alumnos podían ver una imagen del cuestionario en otra pestaña o ventana de su browser, y podían ver también una página de instrucciones acerca cómo entrar los símbolos que aparecen en las transcripciones pero no son fáciles de entrar con el teclado estadounidense. En la Figura 8 se puede ver unas de las instrucciones. Nótese que por tomarse lugar en una clase universitaria estadounidense, todas las instrucciones y las frases del cuestionario aparecen en o fueron traducidos al inglés.

Figura 8. *Detalle de las instrucciones para los transcriptores.*

Special symbols.			
The transcriptions were done with a version of the Americanist notation system (APA). The transcribers used a lot of symbols that are difficult to write on a keyboard. Use this key to type the symbols you find.			
APA	Example	Type:	
b d g		b d g	A line across a letter indicates a fricative. Simply write the plain letter.
ç		ts	Type <ts> for the <c> with a stroke through it.
<sup>x</sup> y		XY	Write any superscript symbols as normal e.g. <sup>n</sup> d → nd
X.Y		X_Y	A small plus indicates a word boundary, type an underscore instead.
ʔ		ʔ	Type 7 for glottal stops
ˈ		ˈ	Type a backtick <`> (above tab) for the apostrophe used to mark the beginning of a stressed syllable.

Las entradas de los alumnos se guardaron en una hoja de cálculo del servicio Google Sheets (Fig. 9). Tener los datos en esta forma facilitó la comparación del desempeño de los alumnos con una transcripción de oro que hice yo.

Figura 9. *Detalle de la hoja de cálculos de las transcripciones.*

	1. sol [sun]	2. luna [moon]	3. estrella [star]
1s	no`ika`dii	`yoo	`lusero
1f	`noikadii	`yoo	`lusero
1s	`bikasii	`yoo	`lusero
1f	oika'o~ii	yoo	lusero



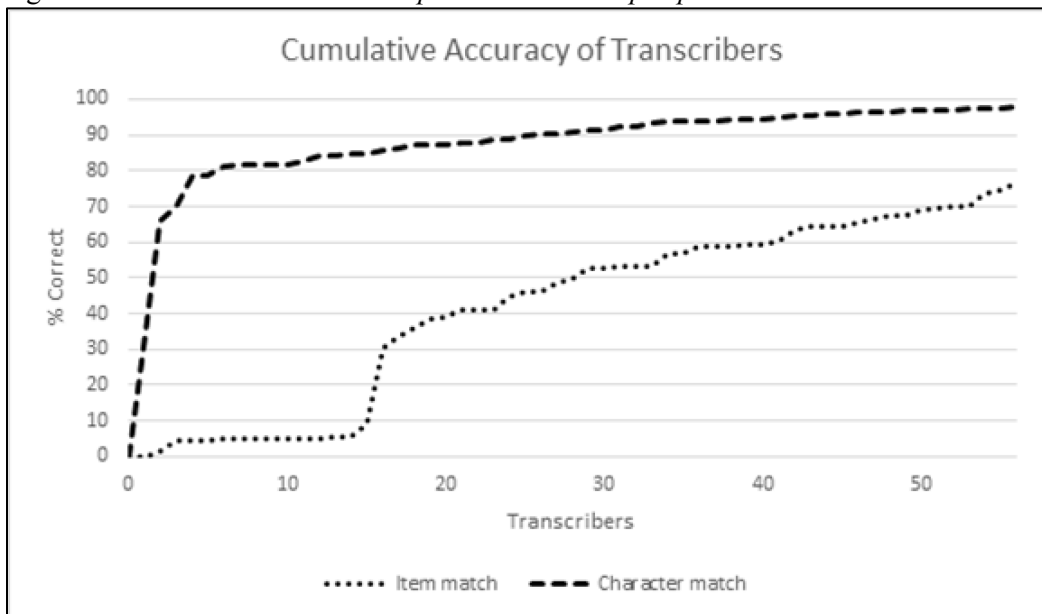
#### 4. Resultados

En total, 56 alumnos entregaron transcripciones, y las comparé contra mi transcripción de la misma encuesta. La distancia Levenshtein entre cada transcripción y la de oro fue calculado, y un porcentaje de precisión fue calculado como en (2).

$$(2) \text{ precisión}(\text{alumno}) = 100 \left( 1 - \left( \frac{\text{distancia Levenshtein}(\text{alumno,oro})}{\text{longitud}(\text{oro})} \right) \right)$$

La precisión fue calculada dos veces. Una vez tomando la precisión carácter por carácter y otra vez tomando la precisión ítem por ítem. Esta segunda medida de precisión considera una respuesta equivalente a la de oro únicamente si toda la palabra o frase del cuestionario fue igual a la transcripción de oro. Ya que la mayoría de las respuestas son oraciones enteras, ésta es una medida de precisión muy conservador. La Fig. 10 muestra las precisiones de los transcriptores ordenados del menos al más preciso.

Figura 10. *Precisión de los transcriptores ordenados por precisión.*



Nótese que la precisión por caracteres empieza alrededor del 65%, sube rápidamente al 80% y de ahí se aumenta paulatinamente hasta llegar casi al 100% de concordancia con la transcripción de oro. Como se esperaría, la medida de precisión por ítems es bien más baja. No supera el 80%, y unos 14 transcriptores tienen una bajísima precisión por ítem. Después de revisar las transcripciones de éstos, noté que ignoraron las instrucciones de sustituciones totalmente. O entraron caracteres sencillas (i.e. <s> por <š>) o entraron los caracteres complejas en lugar de seguir las reglas de sustituciones (i.e. <ñ> por <n~>). Si dejamos al lado los transcriptores que no siguieron esas instrucciones, la precisión por ítem promedia es alrededor del 56%.

La precisión promedia (de caracteres) de los que escogieron el mixteco de Chalcatongo para lectura adicional en la clase (91.85%, N=13) sí era mayor que la de los que no (87.89%, N=43), pero la diferencia no es significativa ( $p = 0.54805$ ).

## 5. Discusión

Los hallazgos de esta prueba indican que la transcripción no experta de textos manuscritos sí es posible, y ofrece algunas sugerencias acerca de cómo preparar una plataforma para tal transcripción.

Es notable que una gran porción de alumnos no hicieron caso al sistema de substitución de múltiples caracteres por glifos complejos. Entonces, recomiendo que cualquier proyecto así busque otras maneras de entrar símbolos menos comunes. Una solución sencilla puede ser meter un cuadro de texto con los símbolos encontrados en el corpus en la página donde entran las transcripciones, para darles una posibilidad de copiar y pegar los necesitan a los transcriptores. Otra solución un poco más elegante y más compleja sería poner botones en la página de transcripción donde uno tiene que nomás dar clic en un botón para entrar un símbolo (en vez de hacer una serie de acciones para copiar y pegar).

También es importante notar que estos alumnos universitario echaron mucha fuerza en acabar las transcripciones con sólo el motivo extrínseco del regalo de puntos adicionales. Unos alumnos me escribieron después para decirme que se divirtieron en y querían saber si fuera más actividades parecidas que podían hacer por ocio. Como eso fue un estudio piloto, no tenía listo cuestionarios adicionales para ofrecerles. Así sugiero que cualquier que empieza un proyecto de transcripción así tenga disponible unos documentos adicionales listos para ofrecer los interesados para puedan continuar transcribiendo por su gusto.

## 6. Conclusión

Con la confirmación que la transcripción no experta si es posible (y con el apoyo y la motivación adecuada es producto es bastante fiel), decidí continuar el proyecto de transcripción con el software FromThePage (Fig. 11).<sup>5</sup> Incorporando las lecciones del estudio piloto, decidí minimizar la cantidad de substituciones de caracteres y en su lugar poner los símbolos menos comunes en el texto de ayuda, con instrucciones de copiar y pegar los símbolos que aparecen para eliminar la posibilidad de recibir transcripciones de estilos divergentes.

Todavía quedan unos retos que enfrentar a todo lo largo del proyecto de transcripción. Uno es cómo enseñar a los transcriptores hacer las interpretaciones correctas del texto manuscrito. Por ahora, sigo poniendo ejemplos de símbolos comunes tomados de los documentos mismos (como en la Fig. 8 arriba) para ilustrar los símbolos raros o idiosincráticos.

Otra tarea sería el reclutamiento de nuevos transcriptores. De hecho, el reclutamiento y motivación de voluntarios es una gran parte del trabajo involucrado en un proyecto de colaboración abierta digital (Ferriter, 2016; Parilla y Ferriter, 2016). La oferta de puntos adicionales en un examen sirvió para motivar un conjunto de alumnos una vez, pero no sería suficiente para reclutar voluntarios que seguirán transcribiendo documentos ni para

---

<sup>5</sup> <https://fromthepage.com/>

reclutar voluntarios fuera de los muros universitarios. Así sería importante incorporar la tarea de transcripción dentro de un módulo de aprendizaje o de promoción cultural para buscar voluntarios que serían animados a continuar con la actividad de transcripción y buscar la participación de personas que son miembros de las comunidades donde recopilaron los documentos y las personas que tienen los idiomas indígenas de los documentos como lengua materna de lengua de herencia.

Figura 11. Nuevo sistema para recolectar transcripciones.

The screenshot shows the 'FROM THE PAGE' web interface. At the top, there is a navigation bar with 'Dashboard', 'Collections', and 'FAQ' links, and a user profile 'Signed In As Ryan Sullivant'. The main content area is titled 'The Kathryn Josserand Mixtec Language Surveys' and includes a breadcrumb trail '→ Story-Conejo-sn-San\_Mateo\_Tunuchi'. Below this, there is a '1' indicator and a 'Page 1 of 1' button. A navigation menu contains 'Overview', 'Transcribe', 'Versions', and 'Settings', with 'Transcribe' selected. The main content is split into two columns: 'Facsimile' and 'Transcription'. The 'Facsimile' column shows a scanned image of a handwritten document with several lines of text. The 'Transcription' column shows the corresponding typed text with phonetic annotations, such as 'kachiri, chi zhachiay chi zhachiay, zho ri tu ra looy' and 'a bichi `nuka `kachi zhoo chi baa dañura'.

Como el software disponible está en inglés actualmente, la internacionalización o traducción del interfaz e instrucciones también sería necesario para divulgar la actividad en las comunidades latinoamericanas (e hispanoparlantes dentro de Estado Unidos) donde

estos idiomas se hablan. Por ahora, la meta es la presentación de instrucciones bilingües donde sea posible.

Finalmente, considero que otro reto es la atribución de las contribuciones de los transcritores. Es importante que las personas sean reconocidas por el labor que hacen cuando están dispuestos a este reconocimiento. Hay que elaborar un sistema de metadatos administrativos para monitorear quiénes han contribuido a una transcripción y quiénes quieren que su nombre (o seudónimo) aparezca al lado de su contribución.

Todo este trabajo está por hacer para crear textos digitales que podemos utilizar para estudios lingüísticos (como análisis de la variación morfosintáctica tras toda la familia mixteca) y que servirán a facilitar la reutilización de las grabaciones en audio de los cuestionarios mismos que ya están preservados en Josserand (s/f). Con el éxito de este proyecto, renovaremos la utilidad del tesoro de dialectología de Josserand.

## 6. Referencias bibliográficas

- Alpert-Abrams, Hannah. (2016). Machine Reading the Primeros Libros. *Digital humanities quarterly*. 10(4).
- Archivo General de Centro América, Guatemala (AGCA), A3.16 Leg 2897 Exp 42986, “Carta de tlatoque,” *Nahuatl/Nawat in Central America*, accessed October 21, 2017, <http://nahuatl-nawat.org/items/show/27>.
- Belmar, Francisco. (1897). *Ensayo sobre la lengua trike*. Imprenta de Lorenzo-San Germán: Oaxaca.
- Déleage, Pierre (investigador), Mario Martínez Tello (actor) y Oswaldo Mélenendez Mélenendez (transcritor, traductor). (2002). MCD003R041I101.pdf. Mito: El antepasado de la libélula (L’ancêtre de la libellule). *Colección de la Lengua Sharanahua de Pierre Déleage*. Archivo de los Idiomas Indígenas de Latinoamérica, [ailla.utexas.org](http://ailla.utexas.org). Acceso: público. PID ailla:209641. Accedido el 21 de octubre de 2017.
- DiCanio, Christian T. y Ryan Bennett. (2017). Prosody in Mesoamerican Languages. In: *The Oxford handbook of language prosody*. Gussenhoven, Carlos y Aojun Chen (eds). Oxford University Press: Oxford.
- Ferriter, Meghan. (2016). Inviting engagement, supporting success: How to manage a transcription center. *Collections: A Journal for Museum and Archives Professionals*. 12(2), pp 97-116.
- Guadalupe Joaquina, Amadeo. (2017). Normalización del sistema de escritura de Tu’un Savi: Un proceso desde el enfoque lingüístico. Ms.
- Hill, Jane (investigadora) Kenneth Hill (investigador), Rosendo Casa Vélez (hablante) y Alberto Zepeda Serrano (transcritor, traductor). NHN004R080I301.pdf. Entrevista con Rosendo Casa. *Colección Nahuatl de Jane Hill y Kenneth Hill*. Archivo de los Idiomas Indígenas de Latinoamérica, [ailla.utexas.org](http://ailla.utexas.org). Acceso: público. PID ailla:161084. Accedido el 21 de octubre de 2017.
- Josserand, J. Kathryn. (1980). Morphological and syntactic variation in Mixtec: Transitive sentences. Paper presented to the 19th Conference on American Indian Languages, Symposium on Syntax, 79th Annual Meeting of the American Anthropological Association, Washington, D.C.

- Josserand, J. Kathryn. (1983). *Mixtec Dialect History*. Ph.D. dissertation. Tulane University.
- Josserand, J. Kathryn. (sin fecha). *The MesoAmerican Languages Collection of Kathryn Josserand*. Archive of the Indigenous Languages of Latin America.
- Lillehaugen, Brook Danielle, George Aaron Broadwell, Michel R. Oudijk, Laurie Allen, May Plumb, y Mike Zarafonetis. (2016). Ticha: a digital text explorer for Colonial Zapotec, first edition. Online: <http://ticha.haverford.edu>
- Matthew, Laura. (2017). *Nahuatl/Nawat in Central America*, accessed October 21, 2017, Nuckolls, Janis B. (investigadora) y Eloise Anachu Cadena (hablante). (1988). QVZ009R085I001.pdf. Esteban y Marcelina Nuñez. *Colección Quechua de Janis B. Nuckolls*. Archivo de los Idiomas Indígenas de Latinoamérica, [ailla.utexas.org](http://ailla.utexas.org). Acceso: público. PID ailla:200547. Accedido el 21 de octubre de 2017.
- Parilla, Lesley y Meghan Ferriter. (2016). Social media and crowdsourced transcription of historical materials at the Smithsonian Institution: Methods for strengthening community engagement and its tie to transcription output. *The American Archivist*. 79(2), pp. 438-460.
- Sandstrom, Alan R. (investigador) y Pamela Efrein Sandstrom (investigadora). (1990). NAWA001R015I101.pdf. Un hombre y un zopilote. *La Colección de Nahuatl de Alan y Pamela Sandstrom*. Archivo de los Idiomas Indígenas de Latinoamérica, [ailla.utexas.org](http://ailla.utexas.org). Acceso: público. PID ailla:203079. Accedido el 21 de octubre de 2017.
- Shimelman, Aviva (compiladora, investigadora). (2012). Recopilación de transcripciones Yauyos. (QUX001R100I001.txt). *Colección Quechua de Yauyos de Aviva Shimelman*. Archivo de los Idiomas Indígenas de Latinoamérica, [ailla.utexas.org](http://ailla.utexas.org). Acceso: público. PID ailla:135135. Accedido el 21 de octubre de 2017.
- Simons, Gary F. y Charles D. Fennig (eds.). 2017. *Ethnologue: Languages of the World*, Twentieth edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>. Accedido el 29 de enero de 2018.
- Sloetjes, H., & Wittenburg, P. (2008). Annotation by category – ELAN and ISO DCR. In: *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.
- Suárez, Jorge (investigador). (1968). TEH002R036I001.pdf. Cuaderno 10. *Colección de Lenguas Argentinas de Jorge Suárez*. Archivo de los Idiomas Indígenas de Latinoamérica, [ailla.utexas.org](http://ailla.utexas.org). Acceso: público. PID ailla:143392. Accedido el 21 de octubre de 2017.

Archive of the Indigenous Languages of Latin America  
 University of Texas at Austin  
 2300 Red River Street  
 SRH 1.108 (S5410)  
 Austin, TX 78712  
 USA

[sullivant@austin.utexas.edu](mailto:sullivant@austin.utexas.edu)