

El empleo de una base de datos para el estudio morfológico del ckunsa: una herramienta digital para la revitalización de la lengua atacameña*

Eduardo Llanquiman Iturrieta

 <https://orcid.org/0000-0002-9175-0792>
Universidad de Chile, Chile
eduardo.llanquiman@ug.uchile.cl

Felipe Hasler Sandoval

 <https://orcid.org/0000-0003-2050-2481>
Universidad de Chile, Chile
fhasler@uchile.cl

Elizabeth Torrico-Ávila

 <https://orcid.org/0000-0002-9309-4055>
Universidad de Antofagasta, Chile
elizabeth.torrico@uantof.cl

RESUMEN

Este estudio se enfoca en la preservación y revitalización del ckunsa, lengua de los indígenas del pueblo atacameño-lickanantay¹. Partiendo de la intersección entre la tecnología y la revitalización lingüística, se propone la difusión de una base de datos lingüística de tipo morfológica. El proceso de construcción del corpus se dividió en tres fases: la compilación y análisis de una lista bibliográfica, la elaboración de un diccionario con palabras ckunsa extraídas de dicho análisis y el desglose morfológico de cada una de esas formas léxicas encontradas. Entre sus potenciales formas de análisis se encuentran: la frecuencia de grafemas y fonemas,

* Este trabajo forma parte del proyecto FONDECYT de Iniciación Nro. 11220225: *En búsqueda del ckunsa: el rescate de la lengua como parte de la identidad Atacameña*, financiado por la Agencia Nacional de Investigación y Desarrollo de Chile (ANID) y a cargo de la Dra. Elizabeth Torrico-Ávila.

¹ El nombre del pueblo originario es lickanantay o atacameño. Ambas palabras se usarán indistintamente a lo largo del texto.



la identificación de patrones silábicos, la evaluación de la coherencia de sus morfemas y la identificación de préstamos léxicos. Los posibles productos que resulten de estos recursos lingüísticos incluyen: aplicaciones digitales, materiales educativos y traducciones más eficaces. Este estudio aporta a la promoción de la diversidad lingüística y cultural, al tiempo que fortalece el vínculo de las comunidades indígenas con su lengua mediante el uso de tecnologías de la información y la comunicación. En particular, la digitalización de los estudios sobre el ckunsa busca avanzar hacia una mayor justicia lingüística para el pueblo atacameño.

Palabras clave: *Ckunsa, Tecnologías de la información, Base de datos morfofonológica, Preservación cultural, Revitalización lingüística*

Employing a Database for the Morpho-Phonological Study of Ckunza: A Digital Tool for the Revitalization of the Atacamean Language

ABSTRACT

This study focuses on the preservation and revitalization of Ckunza, the language of the Atacameño-Lickanantay Indigenous people. Based on the intersection between technology and language revitalization, it proposes the dissemination of a morphophonological linguistic database. The construction of the corpus was carried out in three phases: the compilation and analysis of a bibliographic list, the development of a lexicon containing Ckunza words extracted from that analysis and the morphophonological breakdown of each of those lexical forms. Potential lines of analysis include the frequency of graphemes and phonemes, the identification of syllabic patterns, the evaluation of morphemic consistency, and the detection of lexical borrowings. Possible outputs from these linguistic resources include digital applications, educational materials, and more effective translations. This study contributes to the promotion of linguistic and cultural diversity while strengthening the connection of Indigenous communities with their language using information and communication technologies. In particular, the digitalization of studies on Ckunza aims to advance linguistic justice for the Atacameño people.

Keywords: *Ckunza, Information technologies, Morpho-phonological database, Cultural preservation, Linguistic revitalization*

INTRODUCCIÓN

El ckunsa², antigua lengua de los indígenas del pueblo atacameño, quienes habitan actualmente en el norte de Chile, en el suroeste de Bolivia y en el noroeste de Argentina, ha tenido un desarrollo histórico y lingüístico complejo dentro de la región andina (Bustos, 1999). Clasificado como una lengua aislada, reconocida bajo el código ISO 639-3 kuz (SIL International, 2023), su estructura lingüística no presenta filiación genética con otras lenguas andinas, aunque ha mantenido contacto prolongado con el quechua, el aimara, y lenguas preincaicas ya no habladas naturalmente, como el kakan y el puquina (Adelaar & Muysken, 2004, pp. 350–376).

En efecto, desde su permanencia y expansión geográfica hasta su influencia lingüística (Martínez, 1998), el ckunsa ha sido moldeado por interacciones culturales que han dejado un rastro en su estructura lingüística y desarrollo social (Bustos, 1999). Marcado por la presencia de lenguas indígenas andinas y la lengua castellana (Cerrón-Palomino, 1994), permeándose de ellas en forma de contactos lingüísticos, el ckunsa ha experimentado como resultado un desplazamiento gradual a lo largo de los siglos, lo que llevó a la inexistencia de hablantes nativos (Terborg & García Landa, 2011). Sin embargo, el legado del ckunsa persiste en diversos registros documentales, en la toponimia, fitonimia y antropónimia del territorio, así como en el uso que le da su comunidad hoy en día, lo que abre una ventana hacia su vitalidad lingüística y cultural (Torrico-Ávila, 2020).

² El término <ckunsa> es empleado en este estudio en concordancia con los registros grafémicos contemporáneos y la propuesta del Consejo Lingüístico Ckunsa (2021), en adelante CLCK. A diferencia de las variantes <kunza> o <kunsa>, que han sido utilizadas en documentos históricos y estudios previos (por ejemplo, Mostny *et al.*, 1954; Álvarez, 1996), <ckunsa> refleja una propuesta de unificación ortográfica basada en criterios grafémicos y fonológicos que buscan mayor coherencia con la estructura fonémica de la lengua y su revitalización actual dentro de la comunidad atacameña.

El encuentro entre la sociedad incaica y los lickanantay no solo trajo consigo una expansión territorial, sino también un intercambio lingüístico que dejó marcas indelebles en el ckunsa (Cerrón-Palomino, 2003; Fernández, 2010; Torrico-Ávila, 2020). Este proceso de influencia mutua entre el ckunsa y las lenguas andinas ha dado lugar a una diversidad lingüística que merece ser preservada, pero también diferenciada correctamente. A su vez, la llegada del imperio español y la violencia lingüística impuesta sobre los pueblos indígenas, contribuyeron al desplazamiento definitivo del ckunsa (Torrico-Ávila, 2021). Los últimos hablantes nativos de los que se tiene registro fallecieron en la década de 1950, siendo la antropóloga Grete Mostny y su equipo (Mostny *et al.*, 1954) los últimos en dejar registros lingüísticos de aquellos hablantes.

En el contexto de los esfuerzos por preservar y revitalizar la lengua atacameña, diversos agentes han contribuido desde diferentes áreas del conocimiento y con una variedad no menor de acciones concretas. Sin embargo, aunque estas contribuciones son valiosas, no han logrado alcanzar un nivel de sistematización que permita obtener una visión integral y generalizada del estado actual del ckunsa. Por ello, este escrito presenta un proyecto cuyo objetivo es desarrollar una base de datos que reúna la mayor cantidad posible de partículas lingüísticas presentes en los registros disponibles, incorporando además sus principales características fonológicas y morfológicas.

En definitiva, la creación de un espacio digital bibliográfico y morfonológico para el ckunsa representa un esfuerzo colaborativo y multidisciplinario orientado a sistematizar los registros dispersos de la lengua atacameña, esperando que este recurso sea utilizado para la enseñanza y el aprendizaje en contextos educativos, integrando el ckunsa como lengua con identidad propia dentro de la diversidad lingüística andina (Cooper, 1989). Al facilitar el acceso a su léxico y estructura, también intentamos contribuir al proceso de resignificación cultural de la comunidad atacameña, promoviendo el orgullo étnico y el reconocimiento del ckunsa como parte esencial de su identidad (Torrico-Ávila, 2017).

LA MOTIVACIÓN DETRÁS DEL PROYECTO

A lo largo de las décadas, se han realizado diversos esfuerzos académicos y comunitarios para documentar y revitalizar el ckunsa, incluyendo la creación de diccionarios (Álvarez, 1996; CLCK, 2021), intentos de unificación grafémica (CLCK, 2018; Llanquiman, 2023), reconstrucciones fonológicas (Schuhmacher, 1989; Adelaar & Muysken, 2004; Llanquiman, 2023), análisis morfosintácticos

(Peyró, 2005; Vidal, 2021; García, 2022; Torrico-Ávila, 2022), registros etnográficos (van Kessel, 1996; Bustos, 1999), digitalización y etiquetaje de piezas léxicas (Lehnert, 2021), entre otros trabajos en diferentes áreas del conocimiento.

Sin embargo, estos recursos permanecen dispersos y son de difícil acceso para los investigadores, educadores y miembros de la comunidad. Este fenómeno puede deberse a varios factores como la limitación de los canales de distribución tradicionales, dada la lejanía entre las localidades de la comuna; la escasez de financiamiento para proyectos de difusión masiva; o la falta de una plataforma unificada que centralice la información. Como resultado, la información existente, aunque es completa y detallada, resulta fragmentada y subutilizada, limitando su potencial impacto en la preservación cultural y el ‘despertar’ lingüístico.

A propósito de su vitalidad lingüística, los informes externos catalogan al ckunsa como una lengua extinta (Väisse *et al.*, 1896; Segovia, 2017), sin embargo, las comunidades atacameñas no concuerdan con esta clasificación. Desde una perspectiva interna, el ckunsa es una lengua ‘dormida’, cuyo proceso activo de recuperación, está impulsado por una política lingüística desde abajo (Hornberger, 1997). Este enfoque de revitalización busca reinsertar el ckunsa en la vida comunitaria a través de iniciativas educativas y tecnológicas, especialmente, en Atacama La Grande, fortaleciendo así su función identitaria en la contemporaneidad andina (Torrico-Ávila, 2017).

En un mundo cada vez más globalizado, las lenguas minorizadas enfrentan desafíos crecientes, como la presión de idiomas dominantes y la interrupción de la transmisión intergeneracional (Tollefson, 1991). En este contexto, la tecnología surge como un recurso clave para la preservación del ckunsa, dándole presencia en espacios digitales cotidianos (Arendt & Reershemius, 2024; Davies-Deacon, 2024).

La motivación de este proyecto radica en aprovechar la intersección entre la tecnología y la revitalización del ckunsa para ofrecer un marco teórico y práctico que responda a los desafíos lingüísticos actuales. Se busca adaptar la lengua a los entornos digitales contemporáneos, promoviendo su uso activo en nuevas plataformas de comunicación. En este sentido, el uso estratégico de la tecnología no solo contribuye a frenar un segundo proceso de desplazamiento lingüístico, sino que también empodera a las comunidades atacameñas en la gestión y proyección de su patrimonio lingüístico para el despertar de su lengua (Hodges y Prys, 2024; Terborg *et al.*, 2021; Terborg & García, 2011; Terborg & García, 2013; Torrico & Guerra, 2024).

Sobre este propósito, es importante mencionar el reciente trabajo de Lehnert (2021), donde se digitalizaron más de 800 palabras del ckunsa en la base de datos construida por Key y Comrie (2023) denominada *The Intercontinental Dictionary Series (IDS)*. Este corpus agrupa diferentes diccionarios de lenguas del mundo. Todo el léxico contenido en este marco de datos está etiquetado con códigos que clasifican categorías semánticas, tipos de estudios descriptivos y familias lingüísticas, con el objetivo de permitir comparaciones históricas y geográficas entre las lenguas. Los autores presentan el proyecto de la siguiente manera:

La Serie de Diccionarios Intercontinentales (*IDS*) es una base de datos en la que el material léxico de las lenguas del mundo está organizado de tal manera que se pueden realizar comparaciones. Los estudios históricos y comparativos, así como la investigación lingüística teórica, pueden basarse en esta documentación que está orientada hacia la comprensión y cooperación internacional. Este esfuerzo pionero tendrá un impacto global y su propósito también contribuye a preservar información sobre las lenguas poco conocidas y «no prestigiosas» del mundo, muchas de las cuales están en peligro de extinción (Key & Comrie, 2023).³

Mientras que la base de datos de Lehnert (2021) ofrece un diccionario de palabras etiquetadas del ckunsa y proporciona una base sólida para el léxico, nuestro proyecto se centra en su morfofonología, en donde se busca desarrollar un corpus que permita realizar análisis lingüísticos sobre la estructura interna de cada pieza léxica encontrada en los análisis bibliográficos.

Este corpus incluye más de 1.700 palabras del ckunsa y tiene como objetivo principal ser una herramienta de libre acceso para la comunidad local y otros agentes interesados en la lengua⁴. Aunque actualmente no forma parte de proyectos como *IDS*, *Glottolog* (Hammarström *et al.*, 2024) o *The World Atlas of Language Structures* (Dryer & Haspelmath, 2024), existe la posibilidad de que, en el futuro, este corpus pueda integrarse a repositorios más amplios. Aunque esta decisión dependerá de la comunidad atacameña, que será la encargada de determinar cómo llevar a cabo su inclusión en proyectos globales, respetando los derechos sobre su lengua, sus intereses como pueblo y los principios de autonomía y justicia cultural.

³ Traducción propia.

⁴ Para una visualización completa de los datos, véase el Anexo 1, donde se incluye el enlace al repositorio en *Open Science Framework (OSF)* con la base de datos y todos los archivos complementarios para su uso.

Incluir al ckunsa en proyectos internacionales de esta envergadura proporcionaría una nueva conceptualización respecto al significado de lo tradicional y lo moderno. La revitalización lingüística en este contexto digital integra a la lengua en un panorama mundial preocupado por las minorías lingüísticas (Arendt & Reershemius, 2024). Además, esto resalta la importancia de la investigación en el campo de la política y la planificación lingüística, así como en el lugar que ocupan las lenguas indígenas en la lucha contra la marginación social.

DISEÑO DEL CORPUS

Procedimiento para la construcción de la base de datos

La construcción del corpus forma parte del trabajo de investigación desarrollado para la maestría de Llanquiman (2023), quien propone un grafemario elaborado a partir de un análisis bibliográfico de las formas fonológicas del ckunsa. La construcción del corpus se llevó a cabo en tres fases: un análisis bibliográfico, una recopilación léxica y una categorización morfológica. A continuación, se detallan cada una de estas fases.

En primer lugar, se procedió al análisis bibliográfico de los documentos oficiales del CLCK, es decir, el Grafemario Ckunza (2018) y el Diccionario Unificado Ckunsa (2021) y, separadamente, las investigaciones que sustentan sus elaboraciones. Estas fuentes, que abarcan desde finales del siglo XIX hasta el siglo XXI, constituyen la base teórica sobre la que se fundamenta el estudio del CLCK para sus publicaciones oficiales: Philippi (1860), Tschudy (1869), Moore (1878), San Román (1890), Vaïsse *et al.* (1896), Schüller (1908), Mostny *et al.* (1954), Munizaga (1958), Barthel (1959), Gunckel (1967), Álvarez (1996), Lehnert y Siarez (1997), Vilte (2004) y Torrico-Ávila (2017)⁵. Estas investigaciones fueron la principal base teórica para el inicio de la recopilación léxica.

En segundo lugar, se trasladaron todas las piezas léxicas presentadas de ambos documentos del CLCK a una hoja de *Excel*, donde se identificaron los aportes teóricos de cada estudio. Se formaba así un corpus que integraba tanto el grafemario (CLCK, 2018) como el diccionario (CLCK, 2021), enriquecida además con observaciones adicionales provenientes de estudios complementarios

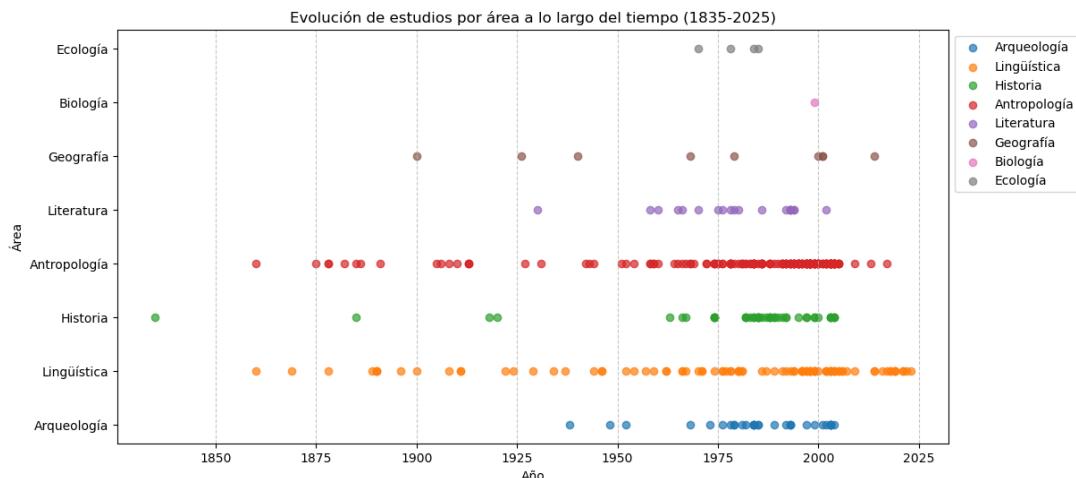
⁵ Existen otros estudios especializados en el área no considerados en estos documentos como los de Torero (2002) y Adelaar (2003) que sí se consideran en la elaboración de la base de datos.

como, por ejemplo, las categorías gramaticales, la identificación de préstamos lingüísticos, las formas silábicas y morfológicas, etc. (Llanquiman, 2023, p. 43).

Por último, se compiló una lista bibliográfica centrada en la cultura y lengua lickanantay, con el propósito de establecer una fuente desde la cual continuar extra yendo unidades léxicas. Esta tarea implicó el análisis de tres guías bibliográficas especializadas, entre ellas los compendios elaborados por Peyró (2006) y Fabre (2009) y las del sitio web *Glottolog* (Hammarström *et al.*, 2024). El resultado fue una recopilación de 383 citas bibliográficas, organizadas por áreas temáticas según la fuente de origen, donde potencialmente podrían existir palabras en ckunsa: lingüística, antropología, historia, literatura, geografía y arqueología. En la Figura 1 se puede observar un gráfico que relaciona la cantidad de referencias bibliográficas sobre lengua y cultura lickanantay encontradas, el área en la que se inscriben y su distribución en el tiempo.

Este compendio bibliográfico actualizado se convirtió en una herramienta práctica para agilizar la búsqueda de información relevante sobre trabajos relacionados a la lengua y su contexto cultural.

Figura 1. Gráfico de dispersión de las citas bibliográficas registradas en el corpus.



Descripción de la base de datos

Posterior a la construcción del corpus, se organizó la información de la siguiente manera:

En la hoja del análisis bibliográfico, por cada referencia encontrada, las columnas se nombraron de la siguiente manera: ‘referencia’, ‘origen’, ‘año’, ‘cantidad de páginas’, ‘área’, ‘universidad o ciudad’ y ‘revista o editorial’. En cuanto a la hoja con el lemario ckunsa, se etiquetó cada palabra encontrada en el análisis bibliográfico como sigue: ‘ckunsa’, ‘español’, ‘categoría gramatical’, ‘fuente de origen’, ‘transcripción fonológica’, ‘transcripción fonética españolizada’, ‘cantidad de sílabas’, ‘formas de sus sílabas’, ‘préstamo lingüístico’, ‘tipo de morfología’, ‘fenómeno morfológico’ y ‘observaciones’. Hoy en día, aún se trabaja en la búsqueda de palabras ckunsa en la extensión de la bibliografía, así como en su desglose lingüístico, proyectando un incremento de mil palabras, es decir, alrededor de 2.700.

Este proceso permitió la creación de un marco de datos que sirve como fundamento para el análisis y la interpretación de los datos lingüísticos del ckunsa a lo largo de la gran mayoría de sus registros. La combinación de técnicas de investigación documental y análisis lingüístico proporcionó una visión panorámica de la lengua, aportando datos para futuras investigaciones en el campo de la lingüística y la revitalización de lenguas indígenas, principalmente, en el contexto de la comunidad atacameña y la era digital.

En la Tabla 1, se presenta el marco de datos del análisis bibliográfico mediante un ejemplo de cinco filas, es decir, con la descripción de cinco referencias, las cuales se desglosan con información pertinente a su búsqueda. En la página siguiente, la Tabla 2 presenta el corpus con el lemario ckunsa en donde se exponen 8 filas seleccionadas al azar para exponer el desglose analítico de cada pieza léxica encontrada.

Actualmente, en la Universidad de Chile se han iniciado proyectos para ampliar este corpus: por un lado, un equipo de estudiantes de magíster de la Facultad de Filosofía y Humanidades incorpora nuevos datos y profundiza en los análisis lingüísticos; por otro, estudiantes de licenciatura de la misma facultad colaboran con Wikimedia Chile subiendo la información a *Wikcionario*; y también se explora su uso en un curso electivo de Ingeniería Lingüística en la Universidad de Antofagasta.

Tabla 1. *Guía bibliográfica principal.*

Referencia	Guía	Año	Nº pág.	Área	Universidad/Ciudad	Revista/Editorial
Adelaar Willem. 2003. La clasificación genética y tipológica del idioma atacameño. Ponencia del simposio ALL-15, 51° CIA, Santiago de Chile.	F&G&P	2003	12	LING	Santiago	Congreso Internacional de Americanistas
Álvarez Cristina y Grebe María Ester. 1974. La trifonía atacameña y sus perspectivas interculturales. RMCh, 126–127.	P	1974	1	ANTR	U. de Chile	Revista Musical Chilena
Bennett Wendell. 1946. The Atacameño. En Hand South American Indian 1, 599–618.	F&P	1946	17	ANTR	U. de Jaén	Smithsonian Institution Bureau of American Ethnology
Gundermann H. 2014. Guía para Educadores Tradicionales Cultura Licanantai y Lengua Kunsa: Programa de Educación Intercultural Bilingüe. Gobierno de Chile. Ministerio de Educación.	NA	2014	103	LING	Santiago	Programa de Educación Intercultural Bilingüe
Larraín Horacio y Ross Eric. 1988. Historia y evolución de los comités de artesanos en pueblos situados en el hinterland de la II Región, Antofagasta. HD 2, 89–104.	P	1988	15	HIST	U. de Antofagasta	Hombre y Desierto

Fuente: Llanquiman (2023, p. 42). Elaborada a partir de Peyró (2005), Fabre (2009), *Glottolog* (2024) e investigaciones propias. Se presenta un ejemplo de 5 referencias de 378 restantes.

Tabla 2. Base de datos principal. Ejemplo de 8 filas de 1397 restantes.

ckunsa	español	categoría grammatical	fuente	t. fonológica	t. fonética españóлизada	cantidad sílaba	forma sílaba	prestamos	raíz o derivado	tipo de derivado	observaciones
ckabar	ropa	sustantivo oficial	Grafemario	'ka.bar	'ka.βai	2	cv.cvc	no	raíz	no aplica	«sin distinción con ckabar»
ckabatur	esconder	verbo oficial	Diccionario oficial	ka.ba.'tur	ka.βa.ʈuɾ	3	cv.cv.cvc	no	derivado	infinitivo	«sin observaciones»
ckabi	apellido	sustantivo oficial	Grafemario	'ka.bi	'ka.βi	2	cv.cv	no	raíz	no aplica	«apellido licanantay de 1613»
ckabur	montaña	sustantivo oficial	Diccionario oficial	ka.'bur	ka.βuu	2	cv.cvc	no	raíz	no aplica	«también cerro elevado»
ckabutur	elevar	verbo oficial	Diccionario oficial	ka.bu.'tur	ka.βu.ʈuu	3	cv.cv.cvc	no	derivado	infinitivo	«sin observaciones»
ckacka	frente	sustantivo oficial	Diccionario oficial	'ka.ka	'ka.ga	2	cv.cv	no	raíz	no aplica	«se refiere a la parte superior de la cara»
yuscka	resbaloso	adjetivo	Lehnert 1978	'jus.ka	'juh.ka	2	cvc.cv	quechua: llucha	raíz	no aplica	«sin observaciones»
yampu	suave	adjetivo	Lehnert 1978	'jam.pu	'jam.pu	2	cvc.cv	quechua: llampu	raíz	no aplica	«sin observaciones»

Fuente: Llanquimán (2023, p. 43).

ANÁLISIS DE DATOS

Análisis preliminares de interpretación

El primer análisis que se llevó a cabo en el corpus se centró en la categorización del texto, facilitando la observación de tendencias lingüísticas en diferentes niveles. En el estudio de Llanquiman (2023), se empleó un enfoque inductivo a los datos con el propósito de unificar las variaciones grafémicas, así como identificar patrones y procesos fonológicos dentro del marco de datos principal. Siguiendo el modelo propuesto por el CLCK (2021), se buscó establecer la consistencia y sistematización de las representaciones fonológicas en la escritura. Este método, similar al empleado por Lehnert (1987), y actualizado recientemente por Viegas Barros (2024) sobre el trabajo con datos precientíficos, se caracterizó por una observación grafémica detallada de cada léxico registrado antes de abordar su sistema fonológico.

Es importante aclarar que los fonemas (y sus interpretaciones fonéticas) presentados en este estudio se extraen de investigaciones previas que registraron las características de la lengua cuando aún se tenía como materna. Sin embargo, debido a la ausencia de hablantes nativos, estas interpretaciones deben considerarse como especulaciones, ya que no es posible comprobarlas de manera definitiva. En este sentido, las interpretaciones gráficas han sido fundamentales para entender el sistema fonológico de la lengua, pues representan una de las aproximaciones más cercanas al haberse realizado en contacto directo con los hablantes originarios (Lehnert, 1987; Viegas-Barros, 2024).

Para remediar en cierta medida estas especulaciones, es preciso destacar el trabajo de Mostny *et al.* (1954), quienes utilizaron la convención fonética del Instituto de Etnografía de la Universidad de París (Cohen & Meillet, 1952) para describir los sonidos que escucharon durante su visita a la localidad atacameña de Peine. A diferencia de otros estudios que utilizaron grafemarios latinos basados en el español, francés o alemán, la convención fonética utilizada por estos investigadores ha sido crucial para una interpretación más precisa de los fonemas del ckunsa. La combinación de los registros gráficos y los símbolos fonéticos utilizados por Mostny y su equipo han permitido obtener una aproximación sólida sobre cómo podría haber sido la lengua en términos fonológicos, constituyéndose como la base principal para su reconstrucción actual y el tratamiento de los datos extraídos de este corpus.

En cuanto a su morfología, los datos resultan más alentadores. A pesar de que los registros disponibles han sido transcritos utilizando diversos sistemas de escritura, su análisis comparativo ha permitido identificar patrones consistentes. Estudios recientes, como el de Vidal (2021) y García (2022) sobre la gramática del ckunsa y la nomenclatura de las partes del cuerpo, han demostrado que es posible reconstruir aspectos gramaticales y léxicos de la lengua a partir de la relación de estos datos. La combinación de enfoques interdisciplinarios, que incluyen la lingüística histórica, la documentación etnográfica y la colaboración con descendientes de la comunidad, ha sido clave para esclarecer dichas estructuras.

Análisis cuantitativos en el corpus

Los análisis presentados a continuación se realizaron principalmente mediante herramientas de búsqueda y conteo de texto simple, complementadas con funciones que permiten calcular datos cuantitativos básicos. Entre las posibles variables a analizar se encuentra, por ejemplo, la frecuencia de aparición de fonemas, la cual se ilustra en la Figura 2, que muestra el porcentaje de ocurrencia de cada fonema en las palabras ckunsa que conforman el corpus.

Para llevar a cabo análisis más complejos, como el cálculo de la frecuencia de formas silábicas, los datos fueron convertidos a formato *CSV* (valores separados por comas), lo que permitió procesarlos en hojas de cálculo. Dado que la segmentación silábica había sido registrada mediante puntos (p. ej., CV.CV), fue posible separar automáticamente las sílabas en columnas individuales. Esta estructuración facilitó una manipulación más eficiente del corpus y permitió extraer información específica sobre la posición y composición de las sílabas dentro de cada palabra.

A partir de este procedimiento, fue posible examinar la estructura silábica del corpus, identificando los tipos de márgenes y sus frecuencias tanto en las cudas como en los ataques silábicos, como se muestra en la Figura 3. Asimismo, se analizó la longitud silábica de las palabras, observándose una mayor frecuencia de formas bisílabas, seguidas por las trisílabas, como se aprecia en la Figura 4.

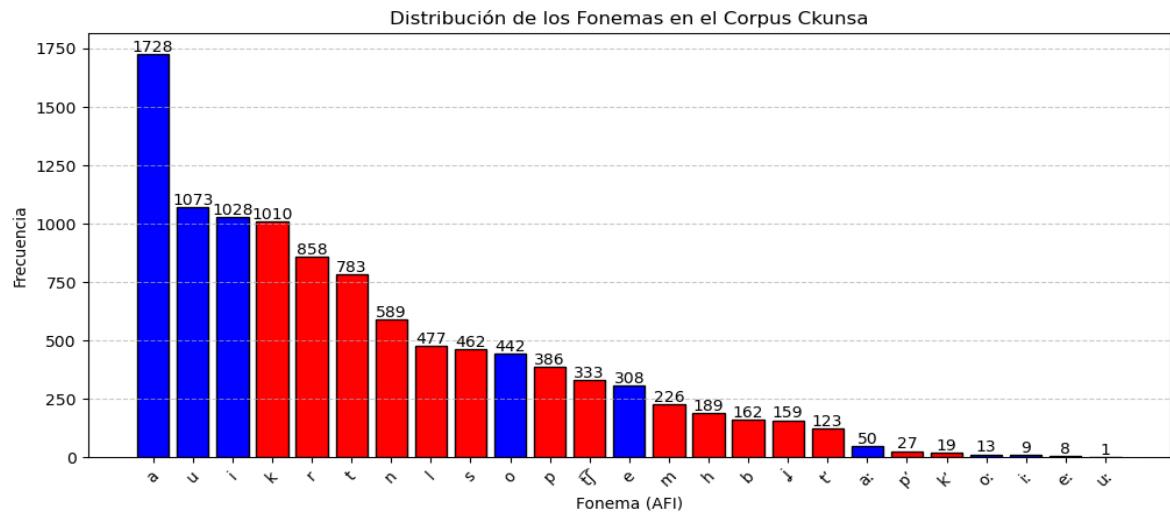
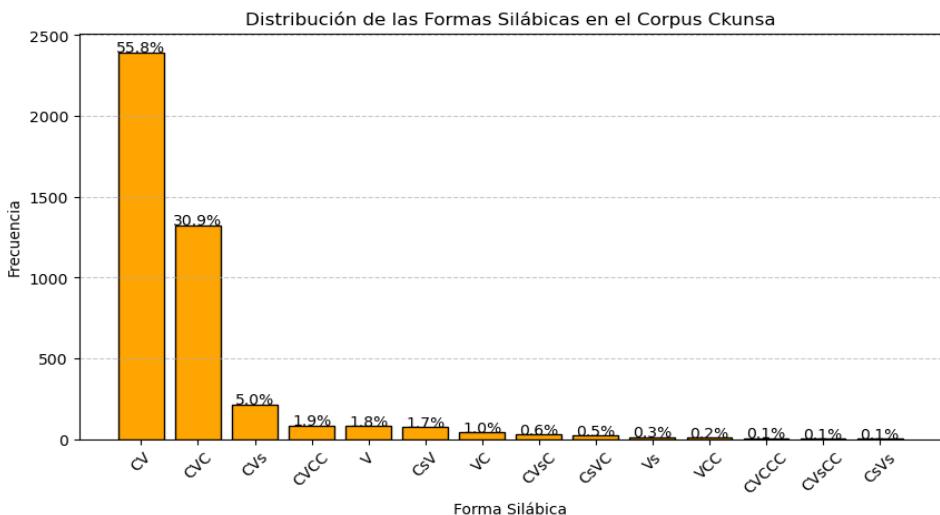
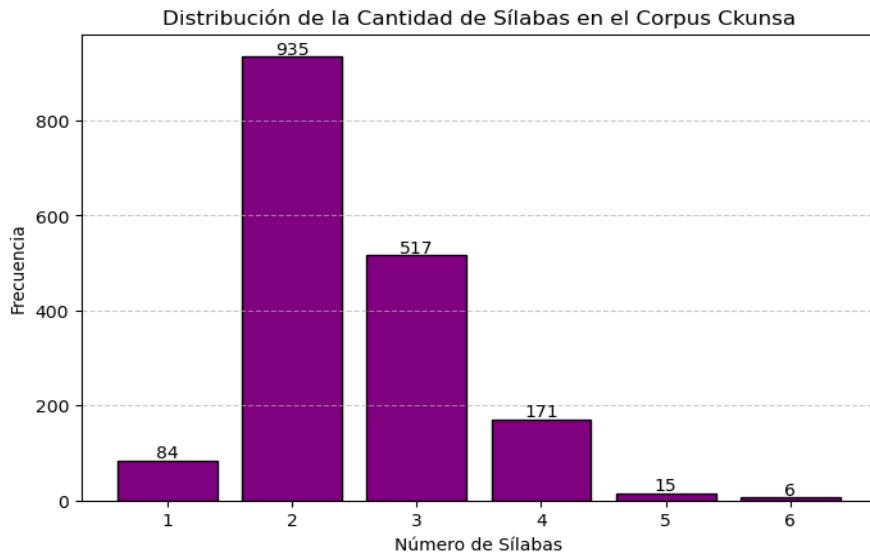
Figura 2. Frecuencias fonémicas en palabras ckunsa.**Figura 3.** Frecuencias de las formas silábicas en palabras ckunsa.

Figura 4. Frecuencias silábicas de las palabras ckansa.

Otro ejemplo de análisis en esta línea se realizó a partir de la columna correspondiente a la transcripción fonológica de las palabras, presentada en notación del Alfabeto Fonético Internacional (AFI), como en *ckabar* /'ka.bar/. Mediante un breve código de programación en *Python* (véase Anexo I), se descompusieron las palabras en sílabas y se distribuyeron en columnas individuales. Como la palabra más extensa del corpus cuenta con seis sílabas, se organizaron seis columnas, cada una correspondiente a una posición silábica (véase Tabla 3).

De este modo, cada fila contenía la segmentación completa de una palabra. Esta estructuración facilitó el tratamiento de los datos fonémicos, permitiendo identificar, por ejemplo, cuántas sílabas presentaban márgenes consonánticos y de qué tipo.

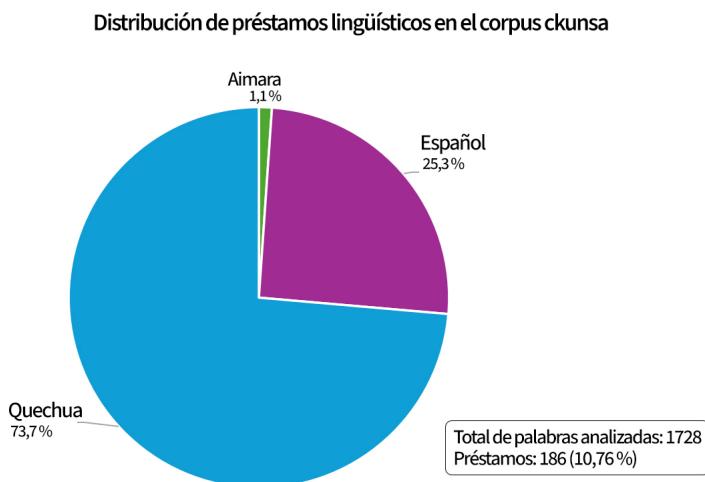
Tabla 3. Separación silábica de las formas fonológicas registradas en el corpus.

una	dos	tres	cuatro	cinco	seis
'ka	<i>βa.ɹ</i>				
'an	ka	<i>tʃi</i>			
a	'ri	tsa	ro		
ka	ma	le	'iʃe	ke	
,ku	ri	ka	san	'tu	ria

Por otro lado, en este marco de datos también fue posible realizar análisis más específicos, como la clasificación de palabras según su morfología, la búsqueda de pares mínimos entre fonemas o la identificación de la cantidad de préstamos lingüísticos. Por ejemplo, de este último análisis se puede observar en la Figura 5, que los préstamos lingüísticos más frecuentes del ckunsa provienen de la lengua quechua. Estos análisis contribuyen a la comprensión de la estructura de la lengua y aportan datos al campo de la investigación lingüística que hoy se empeña en establecer la estandarización de sus sonidos y de su escritura.

Como se puede evidenciar, este formato de datos es compatible con análisis en entornos de programación como *R* (R Core Team, 2024) y *Python* (Python Software Foundation, 2024), lenguajes conocidos por su capacidad para el análisis estadístico y la manipulación de datos. Estos lenguajes automatizan los cálculos mencionados, optimizando la investigación y brindando herramientas adicionales para el análisis morfológico. De hecho, existen librerías específicas en *R*, como *TMI* (*Text Mining Infrastructure*), y en *Python*, como *NLTK* (*Natural Language Toolkit*), diseñadas precisamente para el procesamiento de lenguaje natural y análisis textual; este marco de datos, aunque simple, abarca gran cantidad de datos textuales (*strings*), lo que sin duda es un aporte a este tipo de construcciones analíticas más avanzadas.

Figura 5. *Frecuencias de préstamos lingüísticos en el ckunsa.*



Por último, como se mencionó anteriormente en relación con el trabajo de Lehner (2021) en el *Intercontinental Dictionary Series (IDS)*, futuras investigaciones podrían avanzar en la etiquetación de estos datos mediante códigos específicos que integren categorías semánticas, contextos de uso y lugares de origen. Este enfoque se alinea con iniciativas como la base *Sound Comparison* (Heggarty *et al.*, 2019) o la plataforma colaborativa de libre acceso *Wikidata*. La estructura del corpus resulta además compatible con proyectos internacionales como *Glottolog* (Hammarström *et al.*, 2024) y *WALS* (Dryer & Haspelmath, 2024), lo que le otorga un alto potencial para su incorporación en sistemas de inteligencia artificial y aprendizaje automático. En estos entornos, los datos lingüísticos etiquetados son clave para el desarrollo de modelos de procesamiento del lenguaje natural capaces de producir análisis más sistemáticos y reconstrucciones más precisas.

DISCUSIÓN: UNA HERRAMIENTA DIGITAL PARA LA ESCUELA Y LA INVESTIGACIÓN

La creciente influencia de las tecnologías digitales en la preservación y revitalización de las lenguas minoritarias y minorizadas se pueden estudiar en el libro *Heritage Languages in the Digital Age: The Case of Autochthonous Minority Languages in Western Europe*, editado por Birte Arendt y Gertrud Reershemius (2024). Esta obra analiza el impacto de las herramientas digitales y las redes sociales en la comunicación y el uso de lenguas minorizadas en Europa Occidental, destacando casos de éxito en la revitalización lingüística. A través de diversos estudios de casos, las autoras demuestran cómo las tecnologías digitales pueden ser fundamentales para documentar, enseñar y promover lenguas que de otro modo estarían completamente en desuso. Es de especial interés el capítulo *Teachnig a Regional Language ni Online Course* que pone de relieve la educación en línea como medio principal para la adquisición de lenguas autóctonas minorizadas.

Nuestro proyecto para la lengua ckunsa comparte esta misma visión y metodología, ya que también busca utilizar tecnologías de la información para la preservación y revitalización. Al igual que en los ejemplos presentados en dicho capítulo, este corpus busca ser de libre acceso para permitir un análisis flexible de los datos lingüísticos del ckunsa, ideal para ser tratados, por ejemplo, en cursos en línea (véase el Anexo I). Esto promovería la participación de la comunidad en la preservación de su patrimonio lingüístico y reforzaría el potencial de las TIC para fomentar la enseñanza, el aprendizaje y la visibilidad de las lenguas en peligro de extinción.

Los productos generados por este estudio serán recursos lingüísticos tangibles que se transformarán en herramientas para la preservación cultural y el empoderamiento comunitario. Además de las aplicaciones en cursos en línea, con estos datos es posible elaborar materiales educativos adaptados a las necesidades y preferencias de la comunidad atacameña, siguiendo el legado pedagógico dejados por autores como Lehnert y Bustos (2005, 2007 y 2013) en las ediciones de los libros de estudio *Licana: Textos de Lengua y Cultura Atacameña*, utilizados en las escuelas básicas de la comuna.

En este contexto escolar podría ser más rápido y práctico la traducción de los libros de texto o materiales de enseñanza interactivos, así como una herramienta personal de capacitación para educadores tradicionales y el personal general de los establecimientos de la zona. Asimismo, puede ser utilizado en otros contextos educativos de la comunidad, como para la traducción de material audiovisual de difusión. Por último, estos datos también están orientados para el uso de investigadores, educadores y activistas de lenguas indígenas en todo Chile, contribuyendo así a un diálogo global sobre la importancia de la tecnología en la unificación lingüística.

El aporte final de este estudio se extiende a la promoción de la diversidad lingüística y cultural en todo el continente americano, en línea con las actividades que realiza el grupo *Rising Voices* (Centro de Cultura Digital, 2022) y *Activismo Digital de Lenguas Indígenas* (Global Voices, 2024). Al demostrar el potencial transformador de las tecnologías de la información y la comunicación en el proceso de revitalización lingüística, este estudio puede servir como un modelo para otras comunidades indígenas que enfrentan desafíos de unificación en Chile, como los de la lengua kakan del pueblo diaguita o el de las lenguas huarpes. Además, al integrar principios de participación comunitaria, sostenibilidad y equidad cultural, este proyecto también contribuye a la construcción de un futuro más inclusivo y justo para el pueblo lickanantay y su lengua.

REFLEXIONES FINALES

Aunque aún está por verse la aplicación plena de todas las potencialidades de este corpus, especialmente considerando el vertiginoso avance de la tecnología, en términos generales este estudio constituye un testimonio conmovedor de la capacidad humana para adaptarse y prosperar en contextos adversos, asumiendo la lengua como un elemento central de la identidad comunitaria (Goller & Harteis, 2017, p. 2).

A través de la preservación y revitalización del ckunsa, se honra la memoria de aquellos que lo hablaron y se allana el camino para las generaciones futuras, asegurando que su legado perdure en el consciente mismo de su comunidad, así como en la cultura inmaterial de la humanidad. En un mundo cada vez más interconectado, es fundamental recordar y celebrar la riqueza de la diversidad humana honrando su herencia y, a la vez, adaptándose a las tecnologías de la modernidad. Este estudio nos recuerda que cada cultura tiene un valor intrínseco que merece ser reconocido y protegido, asegurando su legado para las futuras generaciones.

Antes de concluir, es importante señalar que, bajo nuestra perspectiva, la integración de tecnologías emergentes, como la inteligencia artificial, en los procesos de revitalización lingüística no debe percibirse como una amenaza, sino como una herramienta con alto potencial. Al aplicar técnicas avanzadas de procesamiento de lenguaje natural y aprendizaje automático, es posible llenar los vacíos dejados por la ausencia de hablantes nativos, permitiendo reconstruir aspectos estructurales de la lengua de manera más precisa y efectiva. Así, la «IA» puede ser un aliado en la preservación de lenguas en peligro, proporcionando nuevas oportunidades para que las lenguas como el ckunsa sigan siendo un vínculo vivo y vibrante de identidad.

Con estas últimas palabras, se espera que la futura investigación aproveche la amplia, diversa y flexible utilidad de esta base de datos para el rescate lingüístico y cultural del ckunsa, y que, a su vez, la comunidad lickanantay se empodere de estos elementos fundamentales de su identidad. En última instancia, este corpus no solo busca preservar una lengua, sino también fortalecer una comunidad, promover la justicia lingüística y fomentar el empoderamiento de los pueblos indígenas en la era digital. Los avances en la sistematización de la información sientan las bases para la creación de una herramienta educativa accesible, orientada a garantizar la transmisión del ckunsa y su herencia cultural a las generaciones futuras.

REFERENCIAS

- Adelaar, W. (14-18 de julio de 2003). *La clasificación genética y tipológica del idioma atacameño* [Ponencia del simposio ALL-15]. 51º Congreso Internacional de Americanistas, Santiago de Chile.
- Adelaar, W., & Muysken, P. (2004). *The languages of the Andes*. Cambridge University Press.
- Álvarez, A. (1996). *Diccionario ckunsa-español y español-ckunsa*. Imprenta Impresos Universitaria, S. A.

- Arendt, B., & Reershemius, G. (2024). *Heritage Languages in the Digital Age: The Case of Autochthonous Minority Languages in Western Europe*. Multilingual Matters. <https://doi.org/10.21832/arendt4228>
- Barthel, T. (1959). Ein Frühlingsfest der Atacameños. *Zeitschrift für Ethnologie*, 84, 25-45.
- Barthel, T. (1986). El agua y el festival de primavera entre los atacameños. *Allpanchis*, 28, 147-184.
- Bustos, A. (1999). *Etnografía Atacameña*. Instituto de Investigaciones Antropológicas, Universidad de Antofagasta.
- Centro de Cultura Digital. (2022). *Fortalecer tu lengua en internet a través del activismo digital. Foro Internacional Lenguas Indígenas e Innovación*. UNESCO, Rising Voices. <https://centroculturadigital.mx/actividad/fortalecer-tu-lengua-en-internet-a-traves-del-activismo-digital>
- Cerrón-Palomino, R. (1994). *Quechumara: estructuras paralelas de las lenguas quechua y aimara*. Cuadernos de Investigación 42. Centro de Investigación y Promoción del Campesinado.
- Cerrón-Palomino, R. (2003). *Lingüística quechua* (2da ed.). Centro de Estudios Regionales Andinos Bartolomé de las Casas.
- Cohen, M., & Meillet, A. (1952). *Les Langues du Munde*. La société de Linguistique de Paris XVI.
- Consejo Lingüístico Ckunsa. (2018). *Grafemario Unificado Ckunsa*. Corporación Nacional de Desarrollo Indígena (CONADI).
- Consejo Lingüístico Ckunsa. (2021). *Diccionario Unificado Ckunsa*. Corporación Nacional de Desarrollo Indígena (CONADI).
- Cooper, R. (1989). *Language Planning and Social Change*. Cambridge University Press.
- Davies-Deacon, M. (2024). Breton in the Online Context: A new Speaker Community? En B. Arendt & G. Reershemius (Eds.), *Heritage Languages in the Digital Age: The Case of Autochthonous Minority Languages in Western Europe* (pp. 77-102). Multilingual Matters. <https://doi.org/10.21832/arendt4228>
- Dryer, M., & Haspelmath, M. (Eds.). (2024). *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology. <https://wals.info/>
- Fabre, A. (2009). Diccionario etnolingüístico y guía bibliográfica de los pueblos indígenas sudamericanos: Kunza. *Academia Edu*. <https://www.academia.edu/3611584/>
- Fernández, V. (2010). Lenguas en el norte grande de Chile: antecedentes históricos y situación actual. *Tinkuy*, 12. 121-142. <https://dialnet.unirioja.es/servlet/articulo?codigo=3304050>

- García, C. (2022). *Estrategias de formación de palabras en lengua ckunsa* [Tesis de licenciatura, Universidad de Chile].
- Global Voices. (2024). *Activismo Digital para Lenguas Indígenas*. Decenio Internacional de Lenguas Indígenas. <https://rising.globalvoices.org/lenguas/>
- Goller, M., & Harteis, C. (2017). Human Agency at Work: Towards a Clarification and Operationalisation of the Concept. En M. Goller & S. Paloniemi (Eds.), *Agency at Work, Professional and Practice-based Learning 20* (pp. 85-103). Springer. https://doi.org/10.1007/978-3-319-60943-0_5
- Gunckel, H. (1967). Fitonimia atacameña, especialmente cunza. *Revista de la Universidad del Norte*, 30, 1-38.
- Hammarström, H., Forkel, R., Haspelmath, M., & Bank, S. (Eds.). (2024). *Glottolog 4.8*. Max Planck Institute for Evolutionary Anthropology. <https://glottolog.org/>
- Heggarty, P., Shimelman, A., Abete, G., Anderson, C., Sadowsky, S., Paschen, L., Maguire, W., Jocz, L., Aninao, M. J., Wägerle, L., Dërmaku-Appelganz, D., Pheula, A., Lawyer, L. C., Michalsky, J., Suelly, A., Walworth, M., Koile, E., Runge, J., & Bibiko, H. J. (2019). *Sound Comparisons: Exploring Diversity in Phonetics across Language Families*. <https://soundcomparisons.com>
- Hodges, R., & Prys, C. (2024). Language, Education and Community in a Digital Age: A Welsh Digital Resources Case Study. En B. Arendt & G. Reershemius (Eds.), (2024), *Heritage Languages in the Digital Age: the case of autochthonous minority languages in western Europe* (pp. 103-127). Multilingual Matters. <https://doi.org/10.21832/arendt4228>.
- Hornberger, N. (1997). Indigenous Literacies in the Americas. En N. Hornberger *et al*, *Indigenous Literacies in the Americas* (pp. 3-16). Mouton de Gruyter.
- Key, M., & Comrie, B. (2023). *The Intercontinental Dictionary Series*. Max Planck Institute for Evolutionary Anthropology. <https://ids.clld.org>
- Llanquiman Iturrieta, E. (2023). *Planificación lingüística formal lickanantay: perfil fonológico-grafémico de la lengua ckunsa* [Tesis de maestría, Universidad de Chile].
- Lehnert, R. (1978). Préstamos del quechua y el castellano a la lengua Kunza. *Revista de Lingüística Teórica y Aplicada*, 16, 135-140.
- Lehnert, R. (2021). Kunza Dictionary. M. R. Key & B. Comrie (Eds.), *The Intercontinental Dictionary Series*. Max Planck Institute for Evolutionary Anthropology. <http://ids.clld.org/contributions/308>
- Lehnert, R., Reyes, W., & Siares, J. (1997). *Rescate y recopilación de la lengua kunza*. San Pedro de Atacama, Chile: Asociación Indígena Atacameña Zahli Likan Capur, Instituto de Investigaciones Antropológicas, Universidad de Antofagasta y CONADI.

- Lehnert, R. & Bustos, A. (2005) *Licana: Texto de Lengua y Cultura Atacameña NB2.* Universidad de Antofagasta.
- Lehnert, R. & Bustos, A. (2007) *Licana: Texto de Lengua y Cultura Atacameña NB3.* Fundación Minera Escondida.
- Lehnert, R. & Bustos, A. (2013) *Licana: Texto de Lengua y Cultura Atacameña NB4.* Universidad de Antofagasta.
- Martínez, J. (1998). *Pueblos del chañar y el algarrobo. Los atacamas en el siglo XVII.* Ediciones de la Dirección de Bibliotecas, Archivos y Museos.
- Moore, T. (1878). Vocabulaire de la langue atacameña. *Actas del Congreso Internacional de Americanistas*, 2, 44-56. Luxemburgo.
- Mostny, G., Jeldes, F., Gonzalez, R., & Oberhauser, F. (1954). *Peine, Un Pueblo Atacameño.* Instituto de Geografía, Universidad de Chile.
- Munizaga, C. (1958). *Notas etnobotánicas del pueblo atacameño de Socaire. Publicaciones del Centro de Estudios Antropológicos*, 5. Universidad de Chile.
- Peyró García, M. (2005). Estructuras Gramaticales en el Glosario de la Lengua Atacameña (1896). *Línguas Indígenas Americanas*, 5, 25-42. <https://periodicos.sbu.unicamp.br/ojs/index.php/liames/article/download/1437/1427/0>
- Peyró García, M. (2004). Relación de estudios y documentos sobre la antigua lengua de Atacama. *Colección Aalten*, (540), 1-12.
- Python Software Foundation (2024). Python Documentation. *Python*. <https://docs.python.org/>.
- Phillipi, R. (1860). *Viage al desierto de Atacama hecho de orden del Gobierno de Chile en el verano 1853-1854.* Halle, Anton.
- R Core Team (2024). R Documentation. *R*. <https://www.R-project.org/documentation/>
- San Román, F. (1890). *La lengua cunza de los naturales de Atacama.* Imprenta Gutenberg.
- Schüller, R. (1908). *Estudio de la lengua de los indios Lican Antai (Atacameños) - Calchaqui.* Imprenta Cervantes.
- Schuhmacher, W. (1989). Reconstrucción interna del kunza. *Chungará: Revista de Antropología Chilena*, 22, 113-115.
- Segovia, W. (2017). Creación de material para la enseñanza de la lengua Kunza. En E. Torrico-Ávila (Ed.), *Taller de sensibilización para el rescate de la lengua Kunza* (2.ª ed., pp. 35-44). Corporación Nacional de Desarrollo Indígena.
- SIL International. (2023). ISO 639-3 Registration Authority: kuz. *Kunza*. <https://iso639-3.sil.org/code/kuz>

- Tollefson, J. (1991). *Planning Language, Planning Inequality: Language Policy in the Community*. Longman Inc.
- Terborg, R., & García Landa, L. (2011). *Muerte y vitalidad de las lenguas indígenas y las presiones sobre sus hablantes*. Centro de Enseñanza de Lenguas Extranjeras.
- Terborg, R., & García Landa, L. (2013). The Ecology of Pressures: Towards a Tool to Analyze the Complex Process of Language Shift and Maintenance. En A. Massip-Bonet & A. Bastardas-Boada (Eds.), *Complexity Perspectives on Language* (pp. 219-239). Springer.
- Terborg, R., Velázquez, V., & Trujillo, I. (2021). *Presiones que obligan a los hablantes de lenguas originarias, indígenas y minorizadas a abandonar sus lenguas*. Universidad Nacional Autónoma de México. <https://altexto.mx/presiones-que-obligan-a-los-hablantes-de-lenguas-originarias-indigenas-y-minorizadas-a-abandonar-sus-lenguas-2yrc5.html>
- Torrico-Ávila, E. (2017). *Taller de sensibilización para el rescate de la lengua kunza* (2da ed.). Corporación Nacional de Desarrollo Indígena (CONADI).
- Torrico-Ávila, E. [PAR Explora]. (9 de agosto de 2020). *Ciencia abierta: Elizabeth Torrico-Ávila. Revitalización de la Lengua Ckunza en Atacama* [Video]. Youtube. <https://www.youtube.com/watch?v=-rU8oGgRd2s&t=1581s>
- Torrico-Ávila, E. [Centro de Políticas Públicas UC]. (18 de noviembre de 2021). *El estado de vitalidad y desplazamiento de las lenguas originarias de Chile. Conversatorio virtual Plurilingüismo y revitalización de lenguas: experiencias desde los territorios* [Video]. Youtube. <https://www.youtube.com/watch?v=vpRvptq8qVY&t=16s>
- Torrico-Ávila, E. (2022). Los elementos de la gramática de la lengua ckunza de San Pedro de Atacama. *Signo y Pensamiento*, 41. <https://doi.org/10.11144/Javeriana.syp41.eglc>
- Torrico Ávila, E., & Guerra Mejía, R. (2024). Medición de la reversión del desplazamiento del ckunza de San Pedro de Atacama: una primera aproximación. *América Latina Hoy*, 94, e31159. <https://doi.org/10.14201/alh.31159>
- Torero, A. (2002). *Idiomas de los Andes*. Editorial Horizonte.
- Tschudi, J. (1869). La lengua kunza (Trad. y notas por L. Sáez Godoy [1971]). *Signos*, 5(1), 15-20.
- Vaïsse, E., Hoyos, F., & Echeverría y Reyes, A. (1896). *Glosario de la Lengua Atacameña*. Imprenta Cervantes.
- Van Kessel, J. (1996). Poblaciones indígenas de Atacama. En H. Lehuedé (Ed.), *Etnografía: sociedades indígenas contemporáneas y su ideología* (pp. 48-110). Andrés Bello.

- Vidal, A. (2021). *Gramática general de la lengua kunza* [Tesis de licenciatura, Universidad de Chile]. <https://repositorio.uchile.cl/handle/2250/186442>
- Viegas Barro, P. (2024). *Una metodología para la recuperación de la fonética y la fonología de lenguas registradas de manera pre-científica*. Instituto de Lingüística, Universidad de Buenos Aires, Argentina.
- Vilte, J. (2004). *Kunza: diccionario kunza-español, español-kunza*. Programa Codelco.

ANEXO

Como material suplementario, se ha dispuesto en el repositorio de *Open Science Framework (OSF)* el conjunto de recursos utilizados en este estudio, incluyendo los códigos en *Python* empleados para la elaboración de los gráficos, así como los datos que conforman el corpus presentado. Estos insumos se encuentran organizados en archivos *CSV* individuales que comprenden: (i) el corpus léxico principal, (ii) la guía bibliográfica, (iii) un resumen del cuadro fonológico, (iv) un grafemario que sirvió de base para la construcción de dicho cuadro, y (v) el código de *Python*. Todo el material está disponible para su libre consulta y reutilización en el siguiente enlace: https://osf.io/u764r/?view_only=4ce5b31696284a78a0d0cc6f73444755.