

# DISTRIBUCIÓN GEOGRÁFICA DE <Ü> EN EL MAPUDUNGUN DE *SOUNDS COMPARISONS*: UNA CONTRIBUCIÓN AL ESTUDIO DIALECTOLÓGICO MEDIANTE TÉCNICAS COMPUTACIONALES<sup>1</sup>

GEOGRAPHIC DISTRIBUTION OF ALLOPHONIC VARIATION  
OF <Ü> IN MAPUDUNGUN FROM SOUNDS COMPARISONS:  
A CONTRIBUTION TO THE DIALECTOLOGY USING  
COMPUTATIONAL TECHNIQUES

---

EDUARDO LLANQUIMAN ITURRIETA

Universidad de Chile

[eduardo.llanquiman@ug.uchile.cl](mailto:eduardo.llanquiman@ug.uchile.cl)

<https://orcid.org/0000-0002-9175-0792>

FELIPE HASLER SANDOVAL

Universidad de Chile

[fhasler@uchile.cl](mailto:fhasler@uchile.cl)

<https://orcid.org/0000-0003-2050-2481>

## RESUMEN

A partir de la base de datos *Sounds Comparisons*, en este trabajo se describe la distribución geográfica de la alofonía de <ü> en 15 palabras del mapudungun, mediante técnicas computacionales asociadas a la medición de Levenshtein, al *clustering aglomerativo* y los mapas de calor. Se propusieron cuatro contextos de aparición de la vocal: palabras bisílabas con <ü> en la primera de ellas, bisílabas con <ü> en la segunda, bisílabas con <ü> en ambas sílabas y palabras monosilábicas con <ü>; además, tres criterios de observación: cantidad de vocales en la palabra, cantidad de sílabas y posición del acento. Los resultados indican que la vocal no se relaciona geográficamente entre las localidades en estudio, aunque existen tendencias poco significativas entre las macrozonas de la Araucanía Norte, Los Ríos y Argentina, que sí presentan una similitud de pronunciación, especialmente, cuando la vocal es oxítona o monosilábica. De acuerdo con los criterios internos propuestos, se observó una alta frecuencia de [ə] en casi todos los contextos, seguido de [ʊ] que

<sup>1</sup>El presente estudio se enmarca en el Proyecto ANID/Fondecyt de Iniciación 11180078: Los Andes del sur: relaciones con los Andes del centro y del norte desde una perspectiva areal.

aparece siempre en la última sílaba. Estos resultados se asociaron a la distribución dialectal propuesta por Salas (1978) y Croese (1980) donde la distribución alofónica en *Sounds Comparisons* no se corresponde, aunque existan también mínimas tendencias relacionadas a las macrozonas mencionadas. En términos geográficos más generales, no se evidenciaron correlaciones significativas entre las pronunciaciones de las palabras con <ü> y las distancias entre las localidades donde estas se registran, lo que permite suponer que esta vocal no tiene un comportamiento isoglósico o tiene una distribución más compleja que la presentada en *Sounds Comparisons*.

*Palabras clave:* Mapudungun; distribución geográfica; distancia fonética; *clustering aglomerativo*; *Sounds Comparisons*; dialectos.

## ABSTRACT

Based on the Sounds Comparisons database, this study describes the geographical distribution of the allophony of <ü> in 15 words of Mapudungun, using computational techniques associated with Levenshtein measurement, agglomerative clustering, and heat maps. Four contexts of the vowel's occurrence were proposed: bisyllabic words with <ü> in the first syllable, bisyllabic words with <ü> in the last syllable, bisyllabic words with <ü> in both syllables, and monosyllabic words with <ü>. Additionally, three observation criteria were considered: the number of vowels in the word, the number of syllables, and the position of the accent. According to the proposed criteria, a high frequency of [ə] was observed in almost all contexts, followed by [ʊ], which always appears in the last syllable. The vowel is not geographically related among the localities under study, although there are minimal trends among the macrozones of Northern Araucanía, Los Ríos, and Argentina, which show some similarity in pronunciation, especially when the vowel is oxytone or monosyllabic. These results were compared to the dialectal distribution proposed by Salas (1978) and Croese (1980), where the allophonic distribution in Sounds Comparisons does not correspond, although there are also minimal trends related to the mentioned macrozones. In more general geographical terms, no significant correlations were found between the pronunciations of the words with <ü> and the distances between the localities where they are recorded, suggesting that this vowel does not have an isoglossic behavior or has a more complex distribution than presented in Sounds Comparisons.

*Keywords:* Mapudungun; phonetic and geographical distance; clustering agglomerative; *Sounds Comparisons*; dialects.

*Recibido:* 17/11/2023 *Aceptado:* 20/05/2024

## 1. INTRODUCCIÓN

*Sounds Comparisons* (SC) es un proyecto fonético creado por Paul Heggarty que posee más de 50.000 transcripciones fonémicas y fonéticas de más de 600 idiomas alrededor del mundo. Esta base de datos, de libre distribución, incluye transcripciones fonéticas del mapudungun de 37 localidades diferentes, 224 significados agrupados en 15 categorías semánticas y, en relación a estos significados, 7.372 archivos de audio y 7.611 notaciones fonéticas estrechas, geográficamente identificadas en base al Alfabeto Fonético Internacional (International Phonetic Association, 1999; en Adelante AFI). La digitalización de estos datos permite medir y comparar las pronunciaciones en diferentes sectores, relacionando las distancias fonéticas y las distancias geográficas (Heggarty *et al.*, 2019).

Los investigadores que crearon SC lo utilizan principalmente como una base de datos desde la cual intentan medir cuánto se han diferenciado los numerosos idiomas, dialectos y acentos de una familia lingüística. Estas medidas de divergencia se pueden utilizar en una variedad de técnicas cuantitativas para producir redes que representen esos patrones de divergencia a lo largo del tiempo. Estos datos lingüísticos pueden ayudar a comprender la historia de las poblaciones que hablaron las lenguas de las familias lingüísticas cubiertas por el proyecto (Heggarty *et al.*, 2019).

En esta base de datos hay 83 palabras del mapudungun que contienen la sexta vocal <ü><sup>2</sup> (inexistente en castellano) transcritas mediante el AFI. La base de datos ofrece muestras de habla de 37 localidades, aunque no para las 83 palabras. Solo 15 de ellas están asociadas a todas las localidades disponibles, es decir, 15 palabras poseen al menos una vocal <ü> y, a su vez, presentan un total de 37 variaciones (Sadovsky *et al.*, 2019).

Esta vocal es el segmento que ha presentado más diferencias entre las zonas donde se pronuncia. Su variación fonética es un fenómeno complejo que aún no se ha abordado adecuadamente en la dialectología. A pesar de los estudios previos de Croese (1980) y Salas (1976 y 1978), quienes intentaron delimitar y categorizar dialectos del mapudungun, no existe un consenso claro sobre las influencias geográficas y fonéticas que afectan a esta vocal. Incluso no existe consenso entre cuáles son y cómo están condicionados los alófonos de este fonema de manera interna (Mena, 2020).

<sup>2</sup>Dada la diversidad de interpretaciones fonémicas y fonéticas de este segmento en la literatura del mapudungun, hemos optado por utilizar la grafía <ü>, conforme al Alfabeto Mapuche Unificado (Sochil, 1986), para referirnos a nuestro objeto de estudio. Esta grafía se indica con paréntesis angulares para representar una unidad de la grafía. A lo largo de este trabajo utilizamos este mismo alfabeto para escribir todas las palabras en mapudungun. La notación fonémica se indicará entre barras diagonales (/ /), mientras que para la notación alofónica se utilizarán corchetes ([ ]), conforme a las convenciones del Alfabeto Fonético Internacional.

En este trabajo se propone describir la distribución geográfica de las realizaciones alofónicas de <ü> en SC, con el fin de determinar si este segmento se adecúa de alguna manera a las diferentes propuestas dialectológicas que se han elaborado para esta lengua. Para esto se propone medir las diferencias alofónicas de palabras que contengan <ü>, utilizando un algoritmo de distancia que permita identificar tendencias de pronunciación, relacionando distancias fonéticas y geográficas mediante la técnica del *clustering aglomerativo*, específicamente, el *clustering* jerárquico y los dendrogramas. En última instancia, se busca inferir si este segmento tiene un comportamiento isoglósico o si, por el contrario, su distribución geográfica es irregular en el espectro de SC.

Por último, y como se adelantaba más arriba, este estudio se centra exclusivamente en la sexta vocal debido a su alta variabilidad, lo que la convierte en un segmento clave para entender las diferencias dialectales y fonéticas del mapudungun. Al analizar detalladamente este segmento, aislándolo de otras potenciales variaciones segmentales, se espera proporcionar una herramienta para cuestionar y reevaluar las etiquetas dialectales tradicionales, utilizando datos cuantitativos que permitan visualizar las distancias fonéticas y geográficas. Este enfoque contribuye a una observación innovadora de las dinámicas internas del mapudungun, resaltando la necesidad de estudios del mismo tipo más específicos y detallados que puedan ofrecer nuevas perspectivas sobre la variación lingüística en esta lengua. A continuación se ofrece un resumen teórico de la vocal, así como una aclaración terminológica para los conceptos de distancias.

## 2. MARCO TEÓRICO

### 2.1. La sexta vocal del mapudungun <ü>

Dehnhardt *et al.* (2015) dividen la descripción de la sexta vocal en tres épocas: lacónicas, de transición y modernas. En las primeras, se mencionan trabajos de autores misioneros, donde destacan investigaciones como la de Valdivia (1606), Febrés (1764) y Havestadt (1777), quienes de forma impresionista describen esta vocal como “un sonido medio entre la /e/ y la /u/, pero distintivo de /u/”. Los autores (Dehnhardt *et al.*, 2015: 14), aun cuando no es contemporánea, clasifican también en este momento la investigación de Augusta (1903) que describe <ü> como “la *ü* alemana o la *u* francesa en *une* que se forma detrás de los dientes entre la lengua y el paladar, cuidando de mantener la lengua algo retirada y de no tocar con el extremo de ella en ninguna parte” (Augusta, 1903: 2). Ya desde esta época se puede apreciar que este fonema tenía realizaciones fonéticas diferentes.

Las investigaciones de transición comienzan con los estudios de Rodolfo

Lenz, donde se dan descripciones articulatorias más técnicas de <ü> (cercanas a las de Augusta) y donde también se evidencian dos sonidos de un mismo fonema en distribución, al parecer, complementaria: “pronunciación como para *i*, pero el sonido a veces se debilita en una vocal sorda *ə*” (Lenz, 1910: 100). Luego de Lenz, en esta etapa es posible encontrar una serie de autores que han propuesto descripciones para la sexta vocal, junto a su variación alofónica: “[i] siempre es tónica, [ə] solo ocurre en sílabas átonas” (Suárez, 1959: 178); “vocal media [ə], casi áfona. [ü] mezcla la *u* con la *i*, la más gutural de todas” (Mösbach 1962: 26-27); “el sonido /i/ tiene un alófono alto posterior no redondeado [i] en posición tónica y un alófono medio central no redondeado [ə] en posición átona” (Echeverría, 1964: 132); “la vocal *ü* posee dos alófonos [i] y [ə]. El alófono [i] siempre se le encontrará en posición inicial de palabra, posición que jamás tomará el alófono [ə]” (Salas, 1978). De este último autor es preciso destacar los contextos de aparición de los alófonos mencionados, los cuales siguen las pautas de los autores anteriores:

- (1) Se pronuncia como una *u*, pero con los labios puestos en la posición de la *i*.
- (2) También como una especie de *e* muy breve y poco audible, articulada con el dorso de la lengua (no el ápice) ligeramente arqueado hacia arriba. En posición inicial absoluta, debe preferirse (1); en posición final absoluta, sirven (1) y (2). En posición interna entre dos consonantes, ocurre solo (2) en habla rápida o casual, en posición interna entre dos consonantes es particularmente breve, lo que produce al oído la impresión de que no hay vocal entre dos consonantes. En habla rápida y deliberada, la vocal recupera su nivel normal de audibilidad. (Salas, 1992: 73-74)

Por último, es importante destacar en este periodo el estudio fonológico que se realizó para el huilliche, dialecto controversial dentro de los estudios de transición, donde también se describe un sexto segmento: “el inventario del huilliche agrega un fonema vocálico /i/. Este se articula igual que /u/ en castellano, cambiando solamente la posición de los labios a no-redondeada”. (Álvarez Santullano, 1986: 45)

Según el trabajo de Dehnhardt *et al.* (2015), al final de la época de transición, comienzan los estudios modernos y científicos sobre el mapudungun, en donde se entregan definiciones aún más específicas de la variación de este segmento:

Vocal alta, posterior no redondeada [i], siempre tónica en posición inicial y/o final de palabra. Se realiza también como vocal media central no redondeada [ə] con cualquier contexto, o bien puede realizarse como una vocal más breve en sílaba libre ante consonante [ə] que nunca aparece en inicio de sílaba. En posición final de palabra ambos alófonos pueden alternar (Lagos, 2001: 241, siguiendo a Salas, 1978 y 1992)

La variación de <ü> en diferentes contextos es abordada por varios estudios de esta época, como se destaca en las siguientes citas: “A comienzo de palabra la sexta vocal se realiza como [i], al final como [ə]. En otras posiciones alternan” (Zúñiga, 2006: 54); “La pronunciación de *ii* tiende a ser media [ə] en posición átona y alta [i] en posición tónica” (Smeets, 2008: 25); “La sexta vocal /i/ posee dos alófonos, correspondiendo [i] para la realización vocoide alta posterior no redondeada silábica y [ə] para la realización vocoide media central no redondeada silábica. Al final de palabra están en distribución libre” (Salamanca y Quintrileo, 2009: 19); “[ü] alta, posterior, no redondeada, ocurre en posición inicial y final absolutas, mientras que [ə]: media, central, no redondeada, ocurre en todos los contextos, excepto en posición inicial”<sup>3</sup>. (Salamanca y Urrea, 2021)

De este momento es importante el estudio de Sadowsky *et al.* (2013) que invierte la distribución de los alófonos: “se invierte la interpretación tradicional de la tónica [ə] en lugar de [i] y la átona [i] en lugar de [ə] alófonos de la sexta vocal del mapudungun. En posición átona siempre se evidencian alófonos más cerrados”. (93).

A pesar de que existen descripciones más sofisticadas de la vocal en cuestión, como la de Álvarez *et al.* (2020) que incluso estudia sus formantes acústicos<sup>4</sup>, Mena (2020: 95) menciona que aún no existe consenso para una descripción articulatoria definitiva según los contextos de aparición de sus principales alófonos: [ə] y [u]<sup>5</sup>. Para identificar estos contextos de variación, la autora propone tres criterios de observación según los contextos fonéticos que propiciaron variación alofónica en un trabajo previo de la autora:

Our proposal consists of assessing whether acoustic evidence obtained from /ə/ justifies the hypothesis of two allophones ([u] and [ə]), and whether the variability of the data can be explained as an effect of the linguistic variables *stress*, *position of the syllable within the carrier word* and *phonetic context*. Mena *et al.* (2019: 1426)

Para identificar los contextos de variación alofónica de <ü>, se han considerado las variables lingüísticas de acento, posición de la sílaba dentro de la palabra y el contexto fonético, como propone Mena *et al.* (2019). Este estudio

<sup>3</sup> Sobre todas estas definiciones es importante mencionar el trabajo de Croese *et al.* (1978), quienes proponen un sistema de transcripción unificado para la fonemización del mapudungun, la cual siguen varios de los autores mencionados en este trabajo.

<sup>4</sup> Alvar *et al.* (1997) realizó un análisis acústico de la sexta vocal 23 años antes, reafirmando lo dicho por los autores mencionados hasta el momento de transición. Posteriormente, Soto-Barba *et al.* (2016) realizan también una caracterización acústica de la vocal para el mapudungun hablado en el Alto Bío-Bío, donde se siguen las mismas conclusiones.

<sup>5</sup> Sobre estos alófonos es importante mencionar el estudio de Catrileo (2014) quien explica que la variación entre ambos alófonos está condicionada por el estilo de habla (5).

pretende contribuir al esclarecimiento de estas variaciones alofónicas a través del uso de técnicas computacionales avanzadas y análisis cuantitativos, enfocándose en las distancias geográficas y fonéticas. Al observar las agrupaciones zonales de las alofonías de <ü> en el corpus de *SC*, se espera proporcionar una descripción más precisa de su comportamiento fonético y su relación con las propuestas dialectológicas del mapudungun, abordando así un vacío en la literatura existente sobre esta vocal.

## **2.2. Distancias geográficas y distancias fonéticas**

En este estudio se aborda el concepto de distancia en dos dimensiones diferentes: la distancia geográfica y la distancia fonética. La distancia, entendida como la separación o diferencia entre entidades, juega un papel crucial en la variación de las lenguas (Levenshtein, 1966). La distancia geográfica se refiere a la separación física entre las localidades donde se habla un mismo idioma, mientras que la distancia fonética hace referencia a las diferencias en la pronunciación de ese idioma (Moreno Fernández, 2018). En este trabajo, estas dos dimensiones se relacionan para determinar la variación geográfica del sonido <ü> en hablantes del mapudungun.

### **Distancia Geográfica**

La distancia geográfica en este estudio se refiere a la separación física entre las localidades donde se recolectaron los datos del mapudungun. Esta variable se calculó en función de la distancia en kilómetros entre los diferentes lugares, considerando que esta distancia puede influir en la diversidad lingüística según el proyecto de *SC*. Estudios previos han demostrado que las variedades dialectales tienden a ser más similares cuanto más cerca están las comunidades, debido a interacciones sociales y migratorias (Moreno Fernández, 2018). Sin embargo, también pueden ocurrir diferencias significativas incluso en localidades cercanas, dependiendo de factores históricos y culturales.

## **Distancia Fonética**

La distancia fonética se refiere a la diferencia en la pronunciación de los sonidos del mapudungun. En este estudio, la distancia fonética es calculada mediante el algoritmo de Levenshtein, que mide la distancia entre dos secuencias de sonidos. Esto permite cuantificar qué tan diferentes son las pronunciaciones de las palabras en diferentes lugares. Este enfoque es crucial para entender cómo los hablantes varían en la producción de sonidos, incluso en comunidades que están geográficamente cercanas. La distancia fonética es fundamental para explorar la variación dialectal y para entender mejor cómo los factores geográficos influyen en la distribución de las formas lingüísticas.

En la siguiente sección se detalla la metodología utilizada para calcular tanto la distancia geográfica como la distancia fonética entre las localidades estudiadas. Se explican en detalle los métodos empleados, incluyendo el algoritmo de Levenshtein para calcular la distancia fonética, así como los criterios y datos geográficos utilizados en *SC* para calcular la distancia entre localidades.

## **3. METODOLOGÍA**

### **3.1. Descripción de los datos**

Como se mencionó anteriormente, de todas las palabras con la vocal en estudio (83), se seleccionaron solo aquellas que tienen el número total de localidades asociadas. Estas localidades (37) se dividen en siete macrozonas al sur de Chile y Argentina, especificadas en la Tabla I. De las 15 palabras que cumplen con los tres requisitos mínimos para un cálculo homogéneo de distancias entre localidades (organizadas según el número de sílabas y la posición de <ü> en la Tabla II) se seleccionó una muestra de dos palabras por grupo para realizar la comparación y los gráficos.



**Tabla I.** 37 localidades divididas entre las macrozonas de Chile y Argentina en SC.

Localidades según macrozonas en <i>Sounds Comparisons</i>						
<i>Biobío</i>	<i>Araucanía Norte</i>	<i>Araucanía Centro</i>	<i>Araucanía Sur</i>	<i>Los Ríos</i>	<i>Los Lagos</i>	<i>Argentina</i>
Santa Bárbara	Angol	Chol Chol	Nueva Toltén	Mariquina	San Pablo	Chalileo
Cañete	Ercilla	Dollinco	Villarrica	Lanco	J. de la Costa	Picunches
Tirúa	Galvarino	Vilcún	Curarrehue	Panguipulli		Zapala
Alto Biobío	Victoria	Puerto Saavedra		Valdivia		Aluminé
	Lonquimay	Truf Truf				J. de los Andes
	Lumaco	Icalma				Huiliches
		Freire				Lago Rosario
		Cunco				Jacobacci
						Cushamen
						Futaleufú

Tabla II. Palabras por Localidad según Sílabas y Posición de <ü>.

Palabras en estudio			
<i>Bisílaba A</i>	<i>Bisílaba B</i>	<i>Bisílaba C</i>	<i>Monosílaba</i>
<i>küla</i> ‘tres’	<i>kelü</i> ‘rojo’	<i>ngürü</i> ‘zorro’	<i>nün</i> ‘boca’
<i>müt’a</i> ‘cuerno’	<i>kurü</i> ‘negro’	<i>rüpü</i> ‘camino’	<i>üy</i> ‘nombre’

Esta tabla muestra palabras asociadas a todas las localidades, categorizadas según el número de sílabas y la posición de <ü>. ‘Bisílaba A’ corresponde a palabras de dos sílabas con <ü> en la primera sílaba; ‘Bisílaba B’ con <ü> en la última sílaba; ‘Bisílaba C’ con <ü> en ambas sílabas; y las palabras monosílabas tienen un solo núcleo <ü>. Para los cálculos y gráficos, se seleccionaron solo las dos primeras palabras de cada grupo.

3.2. Procesamiento de los datos

La construcción de los datos consistió en la unión de 15 archivos *csv* descargados de SC, uno por cada palabra. Cada archivo *csv* tiene la siguiente información en columnas: la localidad de la pronunciación, la longitud y latitud de su ubicación, la traducción en castellano, su notación fonética estrecha, la palabra en mapudungun<sup>6</sup> y la categoría semántica. A este marco de datos se le adicionó la relación dialectal propuesta por Croese (1980) y Salas (1976). En su conjunto, los datos se organizaron en 532 filas y 10 columnas. Se puede ver una descripción general en la Tabla III con un ejemplo de cuatro filas.

Tabla III. Base de datos principal: primeras cuatro filas de 528 restantes.

Organización de los datos									
<i>índice</i>	<i>localidad</i>	<i>latitud</i>	<i>longitud</i>	<i>español</i>	<i>AFI</i>	<i>mapuche</i>	<i>categoría</i>	<i>Croese</i>	<i>Salas</i>
0	S. Bárbara	-37.6	-71.8018	rojo	'k <sup>h</sup> ë.lə	kelü	colores	II	pehuenche
1	Cañete	-37.9	-73.3928	rojo	'c <sup>h</sup> ë.lə	kelü	colores	I	lafkenche
2	Tirúa	-38.3	-73.4906	rojo	'c <sup>h</sup> ë.lë	kelü	colores	I	lafkenche
3	A. Biobío	-38.4	-71.3634	rojo	k <sup>h</sup> ë.'lə	kelü	colores	II	pehuenche

Para manipular estos datos se utilizaron las siguientes bibliotecas de *Python*: *Pandas* (The Pandas, 2020) para calcular distancias geográficas; *Jellyfish* (Turk y Stephens, 2015) para medir distancias fonéticas y *Scipy* (Virtanen *et al.*, 2020) para los *clustering*, mapas de calor y dendrogramas.

<sup>6</sup>Como se mencionó anteriormente, la escritura del mapudungun en este trabajo se realiza a partir del Alfabeto Mapuche Unificado (Sochil, 1986).

### 3.3. Distancia de Levenshtein

La distancia de Levenshtein es una medida que se utiliza para cuantificar las diferencias entre dos cadenas de texto, en este caso, diferentes pronunciaciones de una misma palabra. Esta métrica cuenta el número mínimo de operaciones necesarias para transformar una cadena de texto en otra. Las operaciones permitidas son agregar, eliminar o reemplazar un elemento. Este método se utilizó para medir las diferencias alofónicas entre distintas pronunciaciones de una misma palabra (Levenshtein *et al.*, 1966). En el Anexo 5 se presenta la fórmula matemática específica.

Para ilustrar cómo funciona este cálculo, se consideraron dos pronunciaciones diferentes de la palabra *wün* 'boca': [wən] y [gwən]. La distancia de Levenshtein entre estas dos pronunciaciones es de 1 punto, ya que solo se necesita una operación (agregar o eliminar un elemento: [g]) para que ambas cadenas sean iguales. En este estudio, el cálculo se centró específicamente en la vocal <ü>, por lo que las diferencias en consonantes no se consideraron en la distancia principal. Por ejemplo, [wən] y [gwən] tienen una distancia de cero, dado que ambas pronunciaciones representan la misma realización de la vocal [ə]; aun cuando se guarde la distancia entre la palabra completa.

Para llevar a cabo este análisis, se utilizó un código de programación que implementa la función de distancia de Levenshtein. Este código, adjunto en el Anexo 6, recibe dos cadenas de texto a comparar y devuelve el número de operaciones necesarias para transformarlas una en otra. Posteriormente, se realizó esta misma comparación métrica en términos numéricos de latitud y longitud, asociando cada pronunciación con su ubicación geográfica. Se generaron nuevos datos que relacionan la distancia en kilómetros entre las localidades comparadas y la diferencia de pronunciación descrita anteriormente. En el Anexo 1 se presentan las distancias geográficas exactas.

El punto de referencia en la base de datos inicialmente marca -40.0 de latitud y -75.5 de longitud, una ubicación en el Océano Pacífico sur a 193 kilómetros de la costa de Valdivia. Este punto se modificó para ubicarse en la ciudad de Temuco, siguiendo a Croese (1980), quien realizó su estudio dialectológico comenzando en esta ciudad, ya que es el núcleo urbano más importante: "simplemente empezamos cerca de Temuco, que es el centro geográfico, sociopolítico, demográfico e institucional del territorio mapuche. Desde allí trabajamos en diferentes direcciones" (Croese, 1980: 11).

Para relacionar los resultados fonéticos y geográficos se desarrolló un código que permitió realizar esta operación para cada una de las localidades reunidas por parejas. De esta forma, se asignó un valor de distancia para cada uno de los cruces entre todas las localidades, resultando en 1,369 (37x37) valores para cada una de

las 15 palabras analizadas. Esta relación se graficó mediante un *clustering aglomerativo*, analizando las diferencias en función de varios factores como el número de sílabas (una o dos), la posición de la vocal en la sílaba (inicial o final), el número de vocales <ü> en la palabra (una o dos) y la posición del acento (primera o segunda sílaba). Es importante señalar que estos gráficos consideraron, para un cálculo homogéneo, una muestra de dos palabras por grupo, es decir, ocho y no 15, escogiendo las palabras con sílabas abiertas para intentar reducir la influencia de los sonidos consonánticos y así enfocarse en la variación alofónica de la vocal en estudio.

### 3.4. Clustering aglomerativo

El *clustering* es un procesamiento de datos que agrupa las características similares de la información en conjuntos (*cluster*). Estos grupos de datos se generan a partir de un algoritmo que discrimina similitudes y diferencias, en este caso, las distancias de pronunciación en diferentes sectores. Estos grupos comienzan con un punto propio para cada valor calculado y luego combinan continuamente los dos grupos más similares hasta que cumple con algún criterio para detener el cálculo, generalmente, cuando el número de conglomerados es 1. En este trabajo, el ‘enlace máximo’ se considera el criterio de agrupación que fusiona dos grupos con la menor distancia máxima entre sus puntos (Müller y Guido, 2016: 174).

Esta forma de agrupación genera una jerarquía en el ordenamiento de los datos, ya que el algoritmo asigna un grupo para cada valor de distancia y luego ese grupo a otro similar inmediatamente arriba, formando ‘grupos de grupos’ que permiten mostrar todas las divisiones posibles entre los datos. Además, esto permite observar en qué áreas una misma palabra tiene pronunciaciones similares o diferentes. Una forma muy frecuente de representar esta agrupación es con mapas de calor y dendrogramas.

Los dendrogramas muestran puntos de datos en la parte inferior y lateral del mapa de calor que, en este caso, corresponden a las 37 localidades específicas. A los lados del gráfico aparecen ramas que indican los grupos formados en un primer ciclo de asociación. Luego, las siguientes ramas se agrupan encima de las primeras para agrupar los clústeres más similares hasta formar uno solo, dibujando una jerarquía en los grupos. A su vez, todos los grupos de la jerarquía se determinan con distintos grados de calor en el centro del mapa, donde los colores más intensos muestran una mayor distancia entre las pronunciaciones, mientras que los más tenues una menor distancia, es decir, pronunciaciones similares.

### 3.5. Relación dialectal

Por último, se elaboraron dos categorías de agrupación dialectal a partir de dos investigaciones que han propuesto similitudes y diferencias entre las pronunciaciones del mapudungun. Croese (1980) divide los dialectos entre los hablados en el norte (zona I y II), en el centro (zonas III, IV, V y VI) y en el sur (zonas VII y VIII). Salas (1978), en cambio, clasifica los dialectos como *picunche* (norte), *mapuche* (centro), *pehuenche* (centro-noreste), *lafkenche* (noroeste), *puelche* (sureste)<sup>7</sup>, *huilliche* (centro-sureste) y *moluche* (suroeste)<sup>8</sup>.

Estas categorías establecen límites geográficos de las propuestas dialectológicas existentes del mapudungun, y se adjuntaron a los datos principales con el fin de comparar los resultados de la agrupación resultante del *clustering* y observar si las tendencias de variación coinciden (o no) con las categorías propuestas por los autores. Una vez relacionados los datos con la información dialectal, se utilizó el algoritmo de Levenshtein para medir la distancia fonética entre las distintas realizaciones alofónicas de las palabras. Luego, se aplicaron técnicas de *clustering* jerárquico para agrupar las localidades según sus similitudes fonéticas con los dialectos mencionados.

En las Figuras 1 y 2 se presentan mapas geográficos con la ubicación de estos dialectos según las divisiones de Croese y Salas, respectivamente. Estos mapas proporcionan una visualización clara de las áreas geográficas asociadas a cada dialecto, facilitando la comparación con los resultados del *clustering* y la distribución geográfica en *SC*.

La Figura 3 muestra el mapa de *SC*, que incluye las localidades en estudio. Este mapa es importante para visualizar cómo se distribuyen geográficamente las muestras fonéticas utilizadas en el análisis.

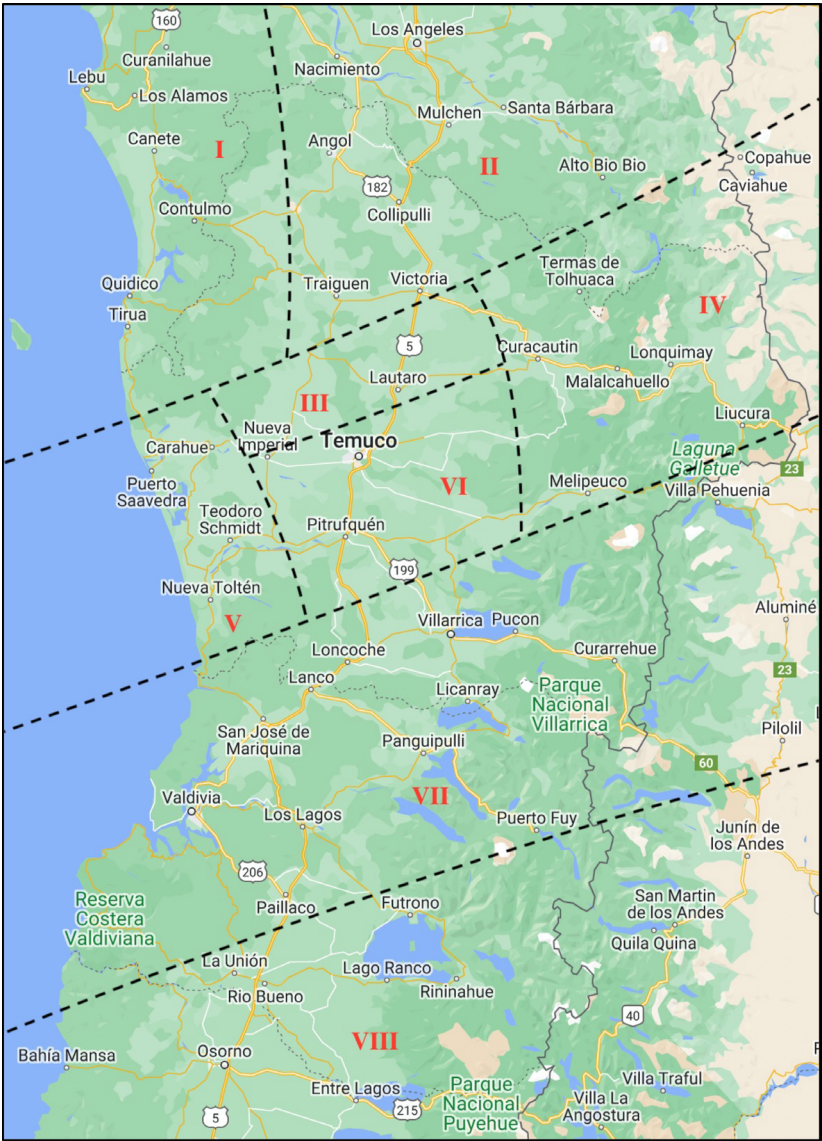
En la Tabla IV se presenta la división propuesta por Croese relacionada con estas localidades, basada en los datos de Heggarty *et al.* (2019). Esta tabla intenta establecer las delimitaciones geográficas de las localidades de *SC* y los grupos dialectales definidos por Croese, permitiendo una comparación directa con los resultados del *clustering*.

Finalmente, en la Tabla V se presenta la relación propuesta por Salas (1978), detallando cómo se distribuyen las localidades en las diferentes categorías

<sup>7</sup> Viegas Barros (2005) no hace una diferenciación clara de estas etiquetas cuando presenta la lengua mapuche en las raíces lingüísticas de la Patagonia, tomando como sinónimos mapuche, moluche y pehuenche. Sin embargo, hace una interesante diferenciación dialectal entre el habla pehuenche y puelche (78).

<sup>8</sup> Erize (1960) presenta un diccionario mapuche-español, donde presenta variedades como 'Araucano Pehuenche', 'Pampa Picunche' y 'Racülche Huilliche'. Es imprescindible un consenso entre todas estas nominaciones para los diferentes espectros del mapudungun y su cultura.

dialectales de *SC* según su propuesta. Esta tabla ofrece una visión complementaria a la de Croese, proporcionando una base sólida para evaluar las concordancias y discrepancias con los resultados del análisis fonético.

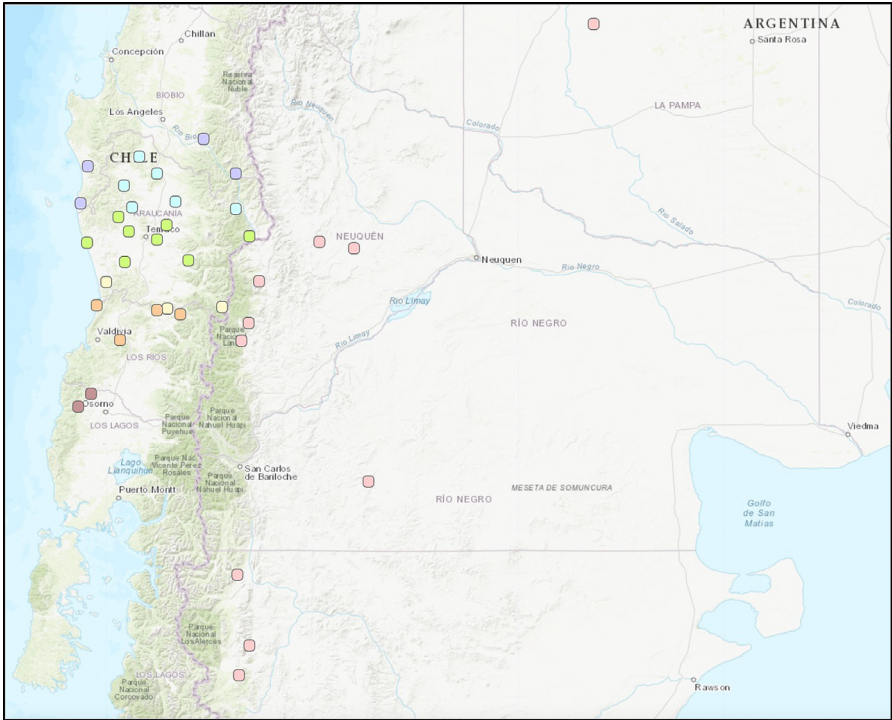


**Figura 1.** Reelaboración del mapa geográfico del sur de Chile según las zonas dialectales del mapudungun propuestas por Croese (1980), mediante la herramienta de Google Maps (2024).



**Figura 2.** Reelaboración del mapa geográfico del sur de Chile según las zonas dialectales del mapudungun propuestas por Salas (1978), mediante la herramienta de Google Maps (2024).





**Figura 3.** Mapa geográfico presentado en *Sounds Comparisons* con los puntos y colores específicos de las localidades en estudio.

Sobre esta figura, es importante mencionar que los colores que identifican macrozonas en SC se replican en las Tablas IV y V para facilitar la comparación con las propuestas dialectológicas de Croese y Salas. Sin embargo, existen divergencias entre las localidades estudiadas por Croese, Salas y SC, ya que no hay unanimidad en la selección de localidades. Algunas de ellas aparecen en SC pero no en las propuestas de Croese ni Salas. Para estas localidades, se ha utilizado ‘n.r.’ (no registrado) para indicar que no están presentes en las clasificaciones de estos autores, pero sí en SC, que cuenta con 37 localidades en total.



**Tabla IV.** Agrupación dialectal del mapudungun propuesta por Croese (1980) en la que se clasifican las localidades geográficas de *Sounds Comparisons* identificadas en colores; n/r quiere decir ‘no registrado’ en las localidades estudiadas por Croese (1980) y que solamente se encuentra en SC.

Agrupación dialectal de Croese									
Norte			Centro				Sur		Argentina
I	II	III	VI	IV	V	VII	VIII		
Cañere	Santa Bárbara (n.r.)	Galvarino	Dollinco (n.r.)	Lonquimay	Puerto Saavedra	Villarica	San Juan de la Costa		Picunches (n.r.)
Tirúa	Alto Biobío (n.r.)	Chol - Chol	Truf-Truf (n.r.)	Vilecún	Nueva Toltén	Curarrehue (n.r.)	San Pablo (n.r.)		Zapala (n.r.)
	Angol		Freire (n.r.)	Icalma		Mariquina (n.r.)	Huilliche (n.r.)		Aluminé (n.r.)
	Lumaco			Cunco		Lanco (n.r.)	J. de los Andes (n.r.)		Lago Rosario (n.r.)
	Ercilla (n.r.)					Panguipulli			Jacobacci (n.r.)
	Victoria (n.r.)					Valdivia			Cushamen (n.r.)
	Chalileo (n.r.)								Futaleufú (n.r.)

Localidades en SC

- Chile: Biobío
- Chile: Araucanía Norte
- Chile: Araucanía Centro
- Chile: Araucanía Sur
- Chile: Los Ríos
- Chile: Los Lagos
- Argentina

**Tabla V.** Agrupación dialectal del mapudungun propuesta por Salas (1978) en la que se clasifican las localidades geográficas de *Sounds Comparisons* identificadas en colores; n/r quiere decir ‘no registrado’ en las localidades estudiadas por Sala (1978) y que solamente se encuentra en SC.

Agrupación dialectal de Salas						
<i>Picunche</i>	<i>Mapuche</i>	<i>Pehuenche</i>	<i>Lafkenche</i>	<i>Puelche</i>	<i>Huilliche</i>	<i>Moluche</i>
Chalileo (n/r)	Dollinco (n/r)	Lonquimay	Puerto Saavedra	Villarica	San Juan de la Costa	Zapala (n/r)
Picunches (n/r)	Truf-Truf (n/r)	Vilcún	Nueva Tolén	Curarrehue (n/r)	San Pablo (n/r)	Aluminé (n/r)
	Freire (n/r)	Icalma	Cañete	Mariquina (n/r)	Huilliche (n/r)	Jacobacci (n/r)
	Chol - Chol	Cunco	Tirúa	Lanco (n/r)	Junín de los Andes (n/r)	
	Galvarino	Santa Bárbara (n/r)		Panguipulli	Lago Rosario (n/r)	
	Angol	Alto Biobío (n/r)		Valdivia	Cushamen (n/r)	
	Lumaco				Futaleufú (n/r)	
	Ercilla (n/r)					
	Victoria (n/r)					

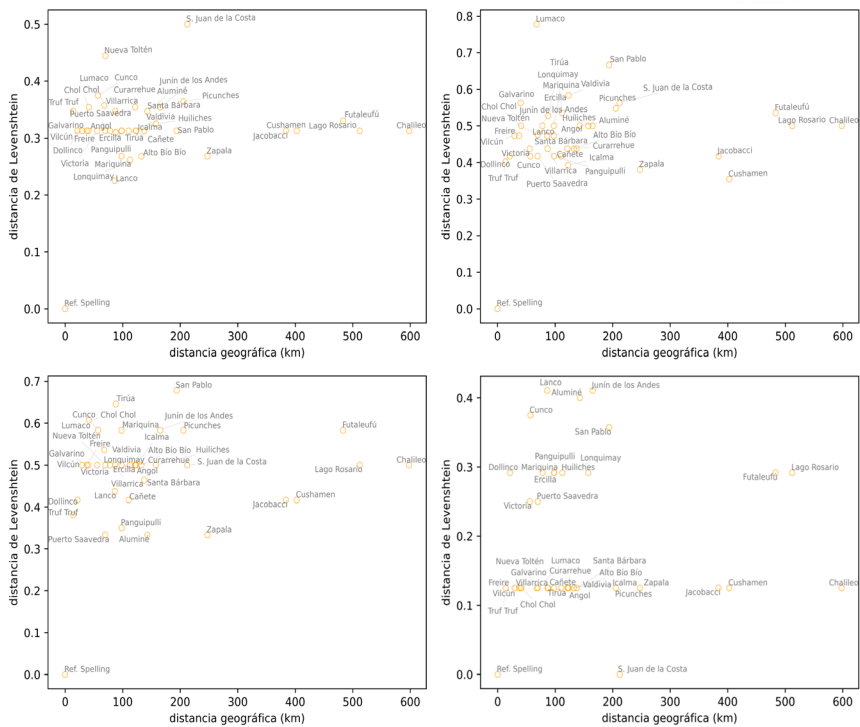
Localidades en SC

- Chile: Biobío
- Chile: Araucanía Norte
- Chile: Araucanía Centro
- Chile: Araucanía Sur
- Chile: Los Ríos
- Chile: Los Lagos
- Argentina

4. RESULTADOS

4.1. Relación global entre las distancias fonéticas y distancias geográficas

La Figura 4 muestra, mediante un gráfico de puntos, la comparación entre las distancias de los cuatro grupos de palabras, es decir, bisílaba A, B, C y monosílaba D.



**Figura 4.** Gráficos de puntos para la relación entre distancias fonéticas y geográficas de palabras con <ü> en cuatro contextos de aparición: arriba a la izquierda, la relación para las palabras del grupo bisílaba A; arriba a la derecha, para bisílaba B; abajo a la izquierda, bisílaba C y, abajo a la derecha, monosílaba D.

Se puede observar que la distribución de los grupos A y B difiere tanto en la distancia al punto de referencia como entre las mismas localidades. Cuando <ü> se encuentra en la primera sílaba existe menos variación que cuando se encuentra al final de la palabra. En ninguno de ambos casos hay una relación geográfica aparente. En el grupo C hay unanimidad en pronunciaciones con 0.5 puntos de distancia al punto de referencia en relación con [ɐ], aunque hay siete localidades con más de 0.2 puntos de diferencia a esta unanimidad, las cuales tampoco están relacionadas geográficamente. Las palabras monosilábicas se pronuncian igual en casi todas las zonas, con algunas excepciones de la palabra *iy* en la zona centro-norte de la Araucanía<sup>9</sup>, en donde no varía la vocal en estudio, sino las consonantes que le rodean, principalmente, en la coda silábica: [ɐj] vs. [ɐjç].

Un primer análisis indica que las palabras bisílabas con <ü> en la segunda, presentan notoria variación, aunque sin correlación geográfica, en cambio cuando la palabra es monosilábica o tiene la vocal en la primera sílaba, el consenso es mayor. Esto sugiere que la combinación de sílabas con vocales distintas a <ü> influye en la variación y que esta tiene, aparentemente, una distribución aleatoria entre las localidades presentadas en *SC*.

## 4.2. Mapas de calor

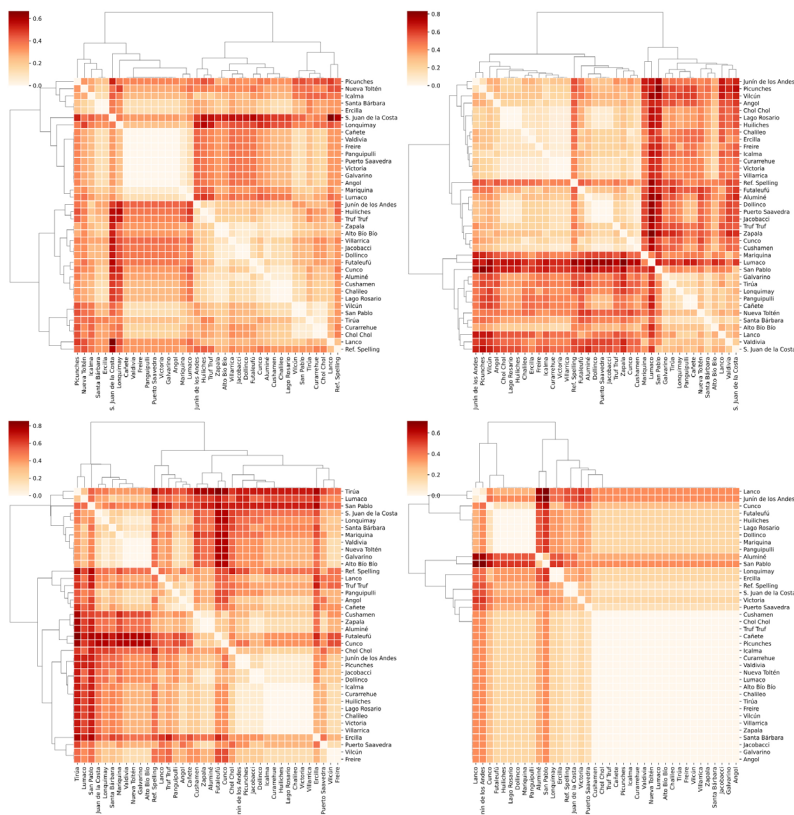
Como se mencionó en la sección de metodología, los mapas de calor y los dendrogramas permiten graficar la jerarquía de las similitudes de los datos, donde se dan relaciones bidimensionales, en este caso, distancias fonéticas y geográficas. La Figura 5 muestra cuatro mapas de calor, correspondientes a los cuatro grupos de palabras: bisílaba A, B, C y monosílaba D.

El gráfico del grupo B, a diferencia del grupo A, muestra más contornos cálidos en casi todas las localidades del mapa, excepto en algunas localidades de Argentina, como Lago Rosario y Huiliches, y en la zona centro-norte de La Araucanía, incluyendo Freire, Icalma, Victoria y Chol-Chol. Todas estas localidades, tanto en Argentina como en La Araucanía, presentan una tendencia hacia el sonido medio central [ɐ]. En el grupo C también se evidencia variación alofónica, pero está dispersa en diferentes sectores, haciendo menos visible su relación geográfica. Sin embargo, nuevamente se observa una agrupación en Argentina (Huiliche, Lago Rosario, Chalileo y Picunches) y, en menor medida, en la zona centro-norte de La Araucanía (Icalma, Victoria, Galvarino y Dollinco). Estas localidades se agrupan en torno al alófono [ɐ] en ambas sílabas. En el grupo D, se observa un consenso aun mayor en cuanto a [ɐ], con algunas excepciones en

<sup>9</sup> Para una descripción fonética específica del mapudungun hablado en Victoria ver Lagos (1981 y 1984).

las localidades de Lanco, Aluminé, Junín de los Andes y San Pablo, que presentan la vocal, pero con variación consonántica en la coda en ambos casos, como en *üy* [ɣ.ˈgɔj] y *wün* [m.ˈbən].

Se analizaron individualmente algunas palabras para observar diferencias específicas y, en particular, las palabras *kütral* y *küül'en* (no seleccionadas del grupo A) muestran muy poco consenso alofónico. Por el contrario, las palabras *miit'a* y *küila* (seleccionadas del grupo A) presentan más similitudes que diferencias. En el primer caso, se observa una pequeña tendencia de pronunciación en la zona de Argentina en torno a los sonidos [ʃ] y [ʒ], aunque existe una notable variación en los márgenes consonánticos. En el segundo caso, la variación en torno a [ʒ] es menos marcada y no está relacionada geográficamente con un sector determinado, ya que es un alófono general en las palabras con sílabas abiertas.



**Figura 5.** Mapas de calor y dendrogramas para la relación entre distancias fonéticas y geográficas de palabras con <ü> en cuatro contextos de aparición. Arriba a la izquierda, la relación para las palabras del grupo bisílabo A; arriba a la derecha, para bisílabo B; abajo a la izquierda, bisílabo C y, abajo a la derecha, monosílabo D.

Esta relación puede estar sujeta al contexto acentual. Mientras que las palabras más diferenciadas tienen el acento mayoritariamente en la primera sílaba, como en *keüla* ['k<sup>h</sup>.lɐ], las que tienen más similitudes lo tienen en la segunda, como en *kelü* [k<sup>h</sup>.lɐ]. Es decir, cuando la vocal en cuestión no lleva el acento, se produce una mayor variación. Aunque en la segunda sílaba exista una mayor fluctuación que cuando está en la primera, la sexta vocal muestra resistencia a convertirse en [ə] cuando está acentuada. Este resultado contradice lo propuesto por Salas (1976), quien afirma que el contexto inicial de sílaba favorece a [u], mientras que este análisis muestra un uso notorio de [ə] en ese contexto.

Es importante destacar que *keülen* y *keütral*, además de tener el acento en la sílaba que no contiene <ü>, poseen una sílaba compleja, a diferencia de las palabras *miüta* y *keüla* con ambas sílabas abiertas. Esto también podría indicar que la mayor variación alofónica de la palabra corresponde a una mayor complejidad articulatoria de las sílabas, además de la configuración prosódica.

En el gráfico del grupo B hay una gran distancia entre las pronunciaciones, especialmente del tipo acentual: *kurü* tiene 25 pronunciaciones de oxítonas y 12 de proparoxítona; *kuvü* 27 y 10; y *kelü* 24 y 13, respectivamente<sup>10</sup>. En *kelü* la variante [ə], muy frecuente en *keüla*, se alterna con [ʊ], lo que podría sugerir que la posición de la vocal en la palabra motiva su variación, en especial, cuando <ü> está en la última sílaba de la palabra, como se menciona en el párrafo precedente. No hay otra observación informativa en el análisis de estas palabras.

La palabra *keniün* muestra muchas similitudes y nuevamente están relacionadas con la alta frecuencia del acento oxítono (33 y 4)<sup>11</sup>. Sin embargo, este caso representa una excepción en cuanto a las notorias diferencias que presentan las palabras con una sílaba cerrada mencionadas anteriormente. La tendencia alofónica de [ə] se corresponde con los dialectos del centro en Croese, es decir, la palabra *keniün* en particular se pronuncia de manera similar en las zonas III, IV, V y VI, lo que en Salas podría ser mapuche, pehuenche o lafkenche, con 0.55 puntos desde la referencia. Aunque una vez más se aclara que estas tendencias geográficas no son suficientemente concluyentes para determinar un comportamiento dialectológico.

Las palabras *riipü* y *ngürü* del grupo C (gráfico abajo a la izquierda) muestran una notable diversidad, sin embargo, existen zonas delimitadas de similitud en torno al alófono [ə] en la Araucanía y en la Argentina. Esto es consistente con el supuesto de que sin contacto con otras vocales hay menor variación en comparación a cuando hay diversidad de vocales en la palabra, sin embargo, la diferencia entre una y otra no es sustantiva en este sentido: aproximadamente 0.1 punto de distancia. Por el contrario, la gráfica de las palabras *nüün* y *iij* (abajo a la

<sup>10</sup>Para ver la totalidad de la distribución alofónica de la vocal <ü> ver los Anexos 2 y 3.

<sup>11</sup>Para ver la totalidad de la distribución prosódica de las palabras en estudio ver Anexo 4.

derecha) presenta una evidente uniformidad en casi todas las zonas, con excepción de las localidades de San Pablo, Aluminé, Lanco y Junín de los Andes donde hay una mayor distancia, a más de 0.4 puntos desde la referencia. En comparación con los otros grupos de palabras, se evidencia una correlación entre el número de sílabas y la baja variación alofónica entre las áreas donde se usan estas palabras.

En resumen, los aspectos de la distribución geográfica son los siguientes:

- a) Las zonas de la Araucanía centro-norte, como Freire, Icalma, Victoria y Cholchol, y ciertas localidades de Argentina, como Lago Rosario y Huilliches, muestran una notable consistencia en torno al alófono [ɟ].
- b) En la Araucanía y en Argentina, específicamente en zonas como Huilliche, Lago Rosario, Chalileo y Picunches, se observa una variación alofónica dispersa pero con algunas agrupaciones notables.
- c) Las localidades de Lanco, Aluminé, Junín de los Andes y San Pablo presentan una mayor distancia fonética respecto al punto de referencia, especialmente en palabras como *iɟy* y *wiin*, mostrando variación consonántica en la coda.
- d) Las palabras *kütral* y *kül'en* presentan poca cohesión alofónica, mientras que *miit'a* y *küla* muestran más similitudes, con la variación geográfica menos pronunciada.
- e) Las palabras monosilábicas, como *iɟy*, tienen una pronunciación más homogénea en casi todas las zonas, con algunas excepciones en la Araucanía centro-norte.
- f) La variación de la vocal <ü> está influenciada por su posición en la palabra y por la complejidad de las sílabas, siendo menos variable en palabras oxítonas y con sílabas abiertas.
- g) Existe una tendencia alofónica en las zonas centro-norte de la Araucanía, Los Ríos y en Argentina, aunque no se evidencian patrones consistentes a lo largo de todas las localidades.

En resumen, los aspectos de variación alofónica interna son los siguientes:

- a) La posición de la vocal en la palabra no presentó ningún tipo de tendencia geográfica, pero sí se observó más variación cuando la vocal está en la segunda sílaba.
- b) Sobre el número de vocales en la palabra, se podría suponer una relación uniforme entre las palabras que solo tienen sílabas con <ü>, aunque no se evidencia en las gráficas. A pesar de ser dos vocales las

- que se estudian, en estas palabras rara vez el sonido es diferente de [ɟ].
- c) La posición del acento es más informativa, ya que no hay variación de la vocal cuando las palabras que la contienen son oxítonas, sin importar su posición en la palabra. Este contexto es el más influyente de los presentados.
  - d) El número de sílabas es determinante. Se calculó la distancia de otras palabras monosilábicas en la base de datos (no asociadas a todas las localidades) y casi siempre hay una tendencia a la homogeneidad de pronunciación, principalmente, de los alófonos [ɟ] y [ʈ].
  - e) Las formas silábicas, en particular la coda, están influenciadas por la variación de la vocal, es decir, si existe variación de <ü> también de sus márgenes.
  - f) Por último, existe tendencia alofónica de [ɟ] sostenida en la zona centro-norte de la Araucanía, en Argentina y, en menor medida, en Los Ríos.

## 5. DISCUSIÓN

Los resultados obtenidos mediante el análisis de los mapas de calor y los dendrogramas revelan algunos patrones en la distribución geográfica y fonética de la sexta vocal del mapudungun. En particular, se identificaron diferencias en la variación alofónica según la posición de la vocal <ü> en la palabra, la cantidad de sílabas y la ubicación del acento. Se observó una mayor consistencia en la pronunciación de palabras monosílabas y aquellas con el acento en la última sílaba, mientras que las palabras con acento en la primera sílaba o con sílabas complejas mostraron una mayor variabilidad. Además, las similitudes fonéticas en ciertas macrozonas, como la Araucanía Norte, Los Ríos y Argentina, sugieren posibles influencias dialectales, aunque no de manera concluyente. A partir de estos hallazgos, se evidencia que la sexta vocal <ü> no se comporta de manera isoglósica clara, lo que plantea interrogantes sobre la existencia de una distribución dialectal definida del mapudungun para esta vocal. Este estudio se enmarca en un contexto sincrónico, explorando las variaciones fonéticas actuales y comparándolas con las propuestas dialectales anteriores de Salas (1978) y Croese (1980).

No existe aún un estudio histórico, etnográfico y lingüístico completo y panorámico que permita afirmar con certeza la existencia de una dialectología del mapudungun. Los procesos históricos subyacentes aún no son completamente conocidos, lo que limita nuestra capacidad para determinar si existe una verdadera dialectología en esta lengua. Este trabajo se basa en datos empíricos extraídos de SC y trata de interpretarlos dentro del marco sincrónico de las variedades fonéticas



del mapudungun. A diferencia de estudios como los de Cerrón-Palomino (2003) para la lengua quechua, no existen datos concluyentes sobre los movimientos poblacionales específicos e históricamente situados que podrían explicar la diversidad de variaciones fonéticas en las hablas mapuche. Por lo tanto, se observa el objeto de estudio desde una perspectiva sincrónica, basada en las variedades existentes, la información disponible en estudios filológicos y la literatura general al respecto.

Es fundamental explicar que la distribución geográfica no equivale a dialecto. Aunque se han observado tendencias en la pronunciación de <ü> entre diferentes localidades, esto no implica necesariamente la existencia de dialectos claramente definidos en el sentido tradicional. La variación fonética puede ser resultado de factores geográficos, históricos y sociales que no necesariamente coinciden con los límites isoglósicos propuestos para el mapudungun. Por lo tanto, la distribución geográfica que se ha observado puede reflejar un continuo dialectal o variabilidad interna dentro de la lengua, más que una relación geográfica concluyente, como lo propone Álvarez Santullano *et al.* (2016).

La relevancia de realizar una exploración de esta naturaleza radica en su potencial para evaluar las etiquetas dialectológicas tradicionales y ofrecer una comprensión más profunda de la variabilidad alofónica de la sexta vocal del mapudungun, siempre controversial para hispanohablantes. Este estudio proporciona datos empíricos que pueden utilizarse para reevaluar estas etiquetas y abrir nuevas perspectivas para futuras investigaciones en fonética dialectal del mapudungun. Además, contribuye a la innovación de los estudio que preservan y fomentan el mapudungun, al posicionar esta lengua dentro de los métodos computacionales en el estudio de las formas lingüísticas internas.

En resumen, este trabajo no solo ofrece nuevas observaciones sobre la variación de <ü> en el mapudungun, sino que también invita a reflexionar sobre las dinámicas digitales de la diversidad lingüística en contextos indígenas. Al evaluar las etiquetas tradicionales y explorar la variabilidad fonética, abre nuevas posibilidades para comprender mejor la complejidad lingüística y cultural del mapudungun.

## 6. CONCLUSIONES

El análisis de los datos en *SC* sobre la distribución geográfica de las palabras relacionadas con <ü> revela variaciones significativas según los criterios fonéticos de observación propuestos. Sin embargo, no se evidenció una correlación directa entre las distancias de las localidades estudiadas y sus pronunciaciones, lo que implica que la sexta vocal en *SC*, en particular, no presenta un comportamiento

isoglósico ni resulta informativa respecto a las etiquetas dialectológicas tradicionales del mapudungun.

Se observaron casos de similitud en lugares geográficamente distantes. Por ejemplo, *kuri* se pronuncia de la misma manera [kũ. 'zə] en Dollinco y Cushamen, a más de 400 kilómetros de distancia. Por el contrario, hay pronunciaciones muy diferentes en localidades vecinas, como entre Lumaco [xwə] y Galvarino ['kũ. ə], a menos de 40 km. No obstante, se identificaron agrupaciones menores en algunas localidades de la Araucanía, especialmente del centro-norte, como en Galvarino, Freire, Victoria, Angol y Chol-Chol. También se encontró una relación menor en la macrozona de Los Ríos, específicamente, entre Valdivia, Panguipulli y Mariquina; así como entre Puerto Saavedra, Tirúa y Cañete. En la macrozona de Argentina, Chalileo se diferenció notablemente del resto de localidades, donde también se evidenció un consenso notorio en torno alófono más frecuente en general: [ə]. Estas observaciones sugieren que, aunque existen algunas pequeñas relaciones entre la pronunciación de la vocal y ciertas macrozonas con localidades cercanas en SC, estas no son suficientes para justificar las etiquetas dialectológicas tradicionales propuestas por Croese y Salas, discutidas durante todo el trabajo.

En cuanto a la variación interna de la vocal, una de las relaciones más evidentes fue entre la posición del acento y la posición de la vocal. No hay variación de <ü> cuando esta se encuentra en la primera sílaba, ya sea oxítone o paroxítone. Por otro lado, cuando <ü> se encuentra en la última sílaba, la vocal y el acento varían considerablemente entre las localidades de SC, exceptuando las ya identificadas de la Araucanía, Los Ríos y Argentina. Como mencionan Salamanca y Quintrileo (2009), la posición de la vocal influye en su variabilidad alofónica: “en posición inicial no varía su realización [i], pero sí lo hace cuando se encuentra en la última, alternando con [ə] y, en diferentes contextos, con [ʊ]” (19), tal como se evidencia en la distribución de los datos presentados aquí.

El número de sílabas de la palabra, también se presentó como un notorio contexto de variación. Mientras que para las palabras bisílabas la variación era más compleja de interpretar, para las palabras monosilábicas la relación fue casi total. Este resultado también podría estar relacionado con la cantidad de vocales involucradas en la pronunciación de la palabra. Como se mencionó anteriormente, la vocal entera puede influir en la relación alofónica de <ü>. Sin embargo, la palabra *riipü* tiene dos vocales idénticas y presenta gran variabilidad, por lo tanto, es imposible por el momento afirmar que este contacto motive heterogeneidad. Un análisis computacional más detallado permitiría aclarar el contexto de estas variantes en relación con su posición léxica, observando por ejemplo, cómo se comportan las otras palabras con dos vocales <ü> no seleccionadas de los datos.

Todo indica que los contextos prosódicos son el principal factor de la variabilidad observada y que esta variación no se corresponde geográficamente

con las macrozonas propuestas en SC. En futuros estudios, resulta crucial considerar estos contextos de manera sistemática, así como los tipos de consonantes en las posiciones de coda y ataque dentro de las sílabas en relación con la vocal. Además, estudiar otras vocales del mapudungun mediante técnicas computacionales similares podría proporcionar una comprensión más amplia de los comportamientos dialectológicos relacionados con <ü>, mediante la observación de la distribución geográfica de datos cuantitativos en propuestas como SC. Esto permitiría, a su vez, determinar si el *continuum* explicado por Álvarez Santullano et al. (2016) también existe a nivel fonético en el sistema vocálico del mapudungun presentado en esta base de datos.

Por otro lado, este estudio exploratorio sugiere que, al estudiar un solo rasgo y aislar muchas variables del sistema fonológico del mapudungun, es posible evaluar las isoglosas dialectales con datos cuantitativos. Estos datos empíricos de SC, a su vez, permiten evaluar las etiquetas tradicionales propuestas por los autores estudiados, las cuales pueden estar influenciadas por diversos factores más allá de los lingüísticos. Por lo tanto, es importante reiterar que, aunque se identificaron algunas relaciones menores, estas no alcanzan un nivel de abstracción suficiente para definir claramente si la variación alofónica de <ü> tiene alguna relación con la dialectología del mapudungun, no al menos en el marco de datos proporcionado por SC.

En cuanto a sus fortalezas, el uso de técnicas computacionales avanzadas y el análisis cuantitativo proporcionan una base sólida para conocer la distribución geográfica exacta de la vocal en estudio y su relación dialectal. Entre las debilidades, se reconoce la necesidad de considerar más variables fonéticas y contextos prosódicos para obtener una comprensión más completa. Los resultados generales indican que <ü> no presenta un comportamiento isoglósico claro y no es un marcador confiable para la diferenciación dialectal en el mapudungun de SC. Las proyecciones futuras incluyen la aplicación de metodologías similares a otras vocales y rasgos fonéticos, así como la integración de análisis más detallados que consideren el contexto prosódico y la interacción con otros segmentos. Los aportes de este estudio destacan la importancia de reevaluar las etiquetas dialectológicas tradicionales a la luz de datos empíricos y cuantitativos, promoviendo una visión más matizada y precisa de la variabilidad lingüística.

Finalmente, este estudio abre nuevas oportunidades para futuras investigaciones que integren técnicas computacionales y análisis cuantitativos. La metodología empleada puede aplicarse a otras características lingüísticas del mapudungun y otras lenguas indígenas, contribuyendo a una comprensión más profunda de la variabilidad dialectal y a la preservación y revitalización de estas lenguas. Además, la colaboración interdisciplinaria entre lingüistas, informáticos y las comunidades hablantes es fundamental para desarrollar herramientas y recursos que reflejen de

manera más precisa la diversidad lingüística y cultural de la nación mapuche.

## REFERENCIAS BIBLIOGRÁFICAS

- Alvar, Manuel; Moreno, Francisco y Ueda, Hiroto. (1997). La sexta vocal del mapuche. *Thesaurus* LII, n°1, 2 y 3, 183-189. Centro Virtual Cervantes.
- Álvarez, Giannina; Ruiz, Magaly; Arias, Alain; Lezcano, María y Fuentes, Ramón. (2020). Description of Acoustic and Articulatory Parameters of Vowels in Mapudungun Speakers. Pilot Study. *International Journal Odontostomatology*, 14(2), 205-212.
- Álvarez Santullano, Pilar. (1986). Descripción fonemática del Huilliche: Estudio comparativo. *ALPHA: Revista de Artes, Letras y Filosofía*, 1(2), 45-50.
- Álvarez Santullano, Pilar; Risco Del Valle, Eduardo y Forno, Amilcar. (2016). Descripción fonético-fonológica del sistema consonántico del mapuche hablado en territorio huilliche en los albores del siglo XXI: a propósito de la noción de continuum, *Revista de Lingüística Teórica y Aplicada*, 54, 101-127.
- Augusta, Feliz. (1903). *Gramática Araucana*. Valdivia, Chile: Imprenta Central J. Lampert.
- Catrileo, María. (2014). La variación estilística en el nivel fonológico del mapudungun. *Lenguas y Literaturas Indoamericanas*, 2, 1-19.
- Cerrón-Palomino, Rodolfo. (2003). *Lingüística Quechua*. Centro de Estudios Regionales Andinos Bartolomé de las Casas, 2da Ed. Cuzco, Perú.
- Croese, Robert. (1980). Estudio dialectológico del mapuche. *Estudios Filológicos*, 15, 7-38.
- Croese, Robert; Salas, Adalberto y Sepúlveda, Gastón. (1978). Proposición de un sistema unificado de transcripción fonémica para el mapudungu. *Revista de Lingüística Teórica y Aplicada*, 16, 151-160.
- Dehnhardt, Macarena; Valenzuela, Adrián y Villarroel, Natalia. (2015). *Adaptación de la sexta vocal /ü/ del mapudungun al español de Chile. Una revisión del Diccionario Etimológico de Rodolfo Lenz*. Tesis para optar al grado de Licenciado en Lengua y Literatura Hispánica con Mención en Lingüística. Facultad de Filosofía y Humanidades, Departamento de Lingüística, Universidad de Chile.
- Echeverría, Sergio. (1964). Descripción fonológica del mapuche actual. *Boletín del Instituto de Filología de la Universidad de Chile*, 16, 13-59.
- Erize, Esteban. (1960). *Diccionario comentado Mapuche-Español: Araucano Pebuenche Pampa Picunche Racülche Huilliche*. Buenos Aires, Argentina: Universidad Nacional del Sur.
- Febrés, Andrés. (1764). *Arte de la lengua general del Reyno de Chile*. Misionero de la Compañía de Jesús. Lima, Calle de la Encarnación.

- Google Maps. (2024). *Mapa del sur de Chile* [Mapa]. Recuperado de <https://www.google.com/maps/@-38.8708622,-72.7769887,8.47z?entry=ttu>
- Havestadt, Bernardo. (1777 [1883]). *Chilidúgu sive Res Chilenses vel Descriptio Status tum Naturalis, tum Civilis, cum Moralis Regni Populique Chilensis, inserta suis locis perfectae ad Chilensem Linguam Manudictioni*, Deo O. M. Monasteri Westphaliae Typis Aschendorffianis.
- Heggarty, P., Shimelman, A., Abete, G., Anderson, C., Sadowsky, S., Paschen, L., Maguire, W., Jocz, L., Aninao, M. J., Wägerle, L., Appelgan, D., Pheula do Couto e Silva, A., Lawyer, L., Suelly Arruda Câmara Cabral, A., Walworth, M., Michalsky, J., Koile, E., Runge, J., y Bibiko, H.-J. (2019). Sound comparisons: A new online database and resource for research in phonetic diversity. *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia, 19, 280-284.
- International Phonetic Association. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.
- Lagos, Daniel. (1981). El estrato fónico del mapudungu(n). *Nueva Revista del Pacífico*, 19/20, 42-66.
- Lagos, Daniel. (1984). Fonología del mapuche hablado en Victoria. *Actas Jornadas de Lengua y Literatura Mapuche*, 1, 41-50. Universidad de la Frontera e Instituto lingüístico de Verano.
- Lagos, Daniel. (2001). *Fonología de las lenguas indígenas en Chile*. Valparaíso: Celto, Universidad de Playa Ancha.
- Lenz, Rodolfo. (1910). *Diccionario etimológico de las voces chilenas derivadas de lenguas indígenas americanas* (Vol. 1). Imprenta Cervantes.
- Levenshtein, Vladimir. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Union*, 10(8), 707-710.
- Mena, Daniela. (2020). *Resolución de aspectos controversiales de la fonética y fonología del mapudungun mediante métodos de fonética acústica y estadística inferencial*. Universidad de Concepción, Facultad de Humanidades y Arte. Programa de Doctorado en Lingüística.
- Mena, Daniela; Figueroa, Mauricio; Rogers, Brandon y Salamanca, Gastón. (2019). Losing one allophone at a time: an acoustic and statistical study on mapudungun's sixth vowel. In *Actas del XIX Congreso Internacional de Ciencias Fonéticas (ICPhS)*. Melbourne, Australia.
- Moreno Fernández, Francisco. (2018). *Dialectología: (Dialectology)*. In *The Routledge handbook of Spanish language teaching* (pp. 377-390). Routledge.
- Mösbach, Ernesto de. (1962). *Idioma Mapuche*. Santiago de Chile: Imprenta San Francisco.
- Müller, Andreas y Guido, Sara. (2016). *Introduction to machine learning with Python: a*

- guide for data scientists*. O'Reilly Media, Inc.
- Sadowsky, Scott; María José Aninao y Paul Heggarty. (2019). *Sound Comparisons: Mapudungun*. <https://soundcomparisons.com/Mapudungun>.
- Sadowsky, Scott; Painequeo, Héctor; Salamanca, Gastón y Avelino, Heriberto. (2013). Mapudungun. *Journal of the International Phonetic Association*, 43(1), 87-96.
- Salamanca, Gastón y Quintrileo, Elizabeth. (2009). El mapuche hablado en Tirúa: Fonemas segmentales, fonotaxis y comparación con otras variedades. *Revista de Lingüística Teórica y Aplicada*, 47(1), 13-35.
- Salamanca, Gastón y Urrea, Paulina. (2021). Fonemas segmentales del mapudungun hablado en Icalma y configuración de un perfil fonético-fonológico del cordón cordillerano de habla mapuche-pewenche. *Logos La Serena*, 31(2), 220-236.
- Salas, Adalberto. (1976). Esbozo fonológico del mapudungun, lengua de los mapuches o araucanos de Chile central. *Estudios filológicos*, 11(2), 143-154.
- Salas, Adalberto. (1978). Mapuche-español. Análisis fonológico contrastivo. *VICUS, Cuadernos Lingüística*, 2, 57-86.
- Salas, Adalberto. (1992). El mapuche o araucano de Chile. *Fonología, gramática y antología de cuentos*. Madrid, España: MAPFRE.
- Smeets, Ineke. (2008). *A Mapuche Grammar*. Berlin-New York: Mouton Grammar Library.
- Sociedad Chilena de Lingüística. (1986). *Encuentro para la unificación del alfabeto mapuche: Propositiones y acuerdos*. Temuco, Chile: Universidad Católica de Temuco.
- Soto-Barba, Jaime; Lara, Isolina y Salamanca, Gastón. (2016). Descripción fonético-acústica de la sexta vocal en el chedungun hablado en Alto Bío-Bío. *Onomázein*, 34, 229-241.
- Suárez, Jorge. (1959). The phonemes of an araucanian dialect. *International Journal of American Linguistics*, 25, 177-181.
- The Pandas Development Team. (2020). Pandas development pandas-dev/pandas: Pandas. *Zenodo*, 21, 1-9. <https://doi.org/10.5281/zenodo.3509134>.
- Turk, James y Stephens, Micah. (2015). Jellyfish: approximate and phonetic matching of strings. <https://github.com/jamesturk/jellyfish>.
- Valdivia, Luis. (1606 [1887]). *Arte y gramática general de la lengua que corre en todo el Reyno de Chile*. Leipzig: B. G. Teubner.
- Viegas Barros, Pedro. (2005). *Voces en el viento: raíces lingüísticas de la Patagonia: lingüística comparativa de las lenguas aborígenes del sur del continente americano*. Buenos Aires, Argentina: Mondragón.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S., Brett, M., Wilson, J., Millman, J., Mayorov, N., Nelson A., Jones, E., Kern, R., Larson, E., ... SciPy 1.0. Contributors (2020). SciPy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods*, 17, 3, 261-272.

Zúñiga, Fernando. (2006). *Mapudungun. El habla mapuche. Introducción a la lengua mapuche con notas comparativas y un CD*. Santiago de Chile: Centro de Estudios Públicos.

7. ANEXOS

Anexo 1. Distancias geográficas en kilómetros desde el punto de referencia.

<i>Localidad</i>	<i>Distancia en kilómetros</i>
Punto de referencia (centro de Temuco)	0.0
Truf Truf	13.869427478423592
Dollinco	21.341255293725403
Vilcún	29.469301443358802
Freire	37.31239284377652
Galvarino	40.16584685475334
Chol Chol	41.01917214528286
Victoria	55.54673944641287
Cunco	57.002529245797675
Lumaco	67.9523695246853
Puerto Saavedra	69.60166056761258
Nueva Toltén	69.63124689447221
Ercilla	77.91701397622668
Lanco	86.08996204557057
Villarrica	86.80852559940777
Tirúa	87.80627921134553
Angol	97.77062185788242
Mariquina	97.82393085052496
Panguipulli	98.27575828385896
Cañete	110.34208645293297
Lonquimay	112.32012886835695
Curarrehue	121.28735180669665
Icalma	122.6406973905147
Valdivia	123.21136273588043
Alto Bío Bío	132.50511427646103
Santa Bárbara	137.41699292927456

Continuación Anexo 1.

Localidad	Distancia en kilómetros
Aluminé	142.86274846204606
Huiliches	157.54074835165122
Junín de los Andes	165.1221402974685
San Pablo	193.76094532963654
Picunches	205.50629997523336
San Juan de la Costa	212.13009374069486
Zapala	247.1569360271878
Jacobacci	383.85360245167334
Cushamen	402.6220273343321
Futaleufú	483.09600916661145
Lago Rosario	512.1135803403132
Chalileo	597.9238254465414

Anexo 2. Distribución alofónica de <ü> en las 15 palabras en estudio.

Palabra	Distribución alofónica de ü
küla	[ɤ]=28, [ɘ]=4, [Ø]=5
kü'en	[ɤ]=22, [ɤ̃]=9, [ø]=1, [Ø]=5
müt'a	[ɤ]=29, [ɤ̃]=6, [Ø]=2
kütral	[ɤ] = 19, [ɘ] = 11, [ɤ̃] = 1, [Ø] = 6
kelü	[ɤ] = 26, [ɪ] = 5, [ʊ] = 5, [ë] = 1
kurü	[ɤ] = 32, [ɪ] = 1, [ʊ] = 3, [Ø] = 1
kewün'	[ɤ] = 37
küwü	[ɤ] = 32, [ʊ] = 4, [Ø] = 1
molljün	[ɤ] = 30, [ɘ] = 1, [wɤ] = 3, [wë] = 1, [wɪ] = 1, [wĩ] = 1
ant'ü	[ɤ] = 29, [ë] = 1, [ɤ̣] = 1, [ɛ̣] = 1, [ʊ] = 5
pulkü	[ɤ] = 1, [ɤ̃] = 1, [ø] = 5, [ʊ] = 28, [ʊ̃] = 2
ngü/rü	[ɤ] = 37/[ɤ] = 34; --/[Ø] = 1; --/[ʊ] = 2
rü/pü	[ɤ] = 30/[ɤ] = 30; [ë] = 2/--; --/[ø] = 1; --/[ɪ] = 1; [ʊ] = 3/[ʊ] = 4; --/[ʊ̃] = 1; [Ø] = 2/--
wün	[ɤ] = 37
üy	[ɤ] = 37



**Anexo 3.** Frecuencias totales de la distribución alofónica de <ü>.

Frecuencias totales					
[ə] = 490	[ə] = 453	[ə̞] = 1	[ʷə] = 3	[ə̠] = 2	[ə̣] = 31
[Ø] = 23	--	--	--	--	--
[ö] = 7	--	--	--	--	--
[ɪ] = 9	[ɪ] = 5	[ɪ] = 2	[ʷɪ] = 1	[wĩ] = 1	--
[ʊ] = 57	[ʊ] = 54	[ʊ̞] = 3	--	--	--
[ë] = 6	[ë] = 4	[ʷë] = 1	[ë̞] = 1	--	--

**Anexo 4.** Distribución acentual de las 14 palabras en estudio.

Palabra	Distribución acentual	
<i>küla</i>	oxítona: 20	paroxítona: 17
<i>kül'en</i>	oxítona: 35	paroxítona: 2
<i>müt'a</i>	oxítona: 20	paroxítona: 17
<i>kütral</i>	oxítona: 36	paroxítona: 1
<i>kelü</i>	oxítona: 24	paroxítona: 13
<i>kurü</i>	oxítona: 25	paroxítona: 12
<i>kenün'</i>	oxítona: 33	paroxítona: 4
<i>kunwü</i>	oxítona: 27	paroxítona: 10
<i>mollfün</i>	oxítona: 33	paroxítona: 4
<i>ant'ü</i>	oxítona: 26	paroxítona: 11
<i>pulkü</i>	oxítona: 17	paroxítona: 20
<i>ngürü</i>	oxítona: 23	paroxítona: 14
<i>rüpiü</i>	oxítona: 21	paroxítona: 16
<i>wiin</i>	oxítona: 37	--
<i>iüy</i>	oxítona: 37	--

**Anexo 5.** Fórmula de Levenshtein (versión simplificada)

$$d(i, j) = \min \begin{cases} d(i - 1, j) + 1 \\ d(i, j - 1) + 1 \\ d(i - 1, j - 1) + 1_{(a_i \neq b_j)} \end{cases}$$

Donde:

- $d(i,j)$  es la distancia de Levenshtein entre los primeros  $i$  caracteres de una cadena  $a$  y los primeros  $j$  caracteres de una cadena  $b$ .
- $1(a_i \neq b_j)$  es un indicador que vale 1 si  $a_i \neq b_j$  y 0 de lo contrario (costo de sustitución).

## Anexo 6. Distribución Código para Calcular la Distancia de Levenshtein

### python

```
import jellyfish
def funcion_distancia(string1, string2):
    return jellyfish.levenshtein_distance(string1,
        string2)/max(len(string1), len(string2))

# Ejemplo de uso:
string1 = "ejemplo"
string2 = "ejemplo"
distancia = funcion_distancia(string1, string2)
print(f "La distancia de Levenshtein es: {0}")
```