

Propuesta de Tesis:

"Diseño y desarrollo de un software lingüístico para apoyar el aprendizaje del Mapuzugun"

Autor: Cristian Ahumada Oliva
Profesor Guía: Claudio Gutierrez

Fecha: 9 de Septiembre de 2020

Wünelu mu chaliwaiñ

Mari mari pu peñi, pu lamgen, pu chillkatufe, pu kimelfe, ka kom tati pu allkütupalu tüfachi zugu. Pewmagen mai kom kümelekaymün kizu tamün pu che eymün.

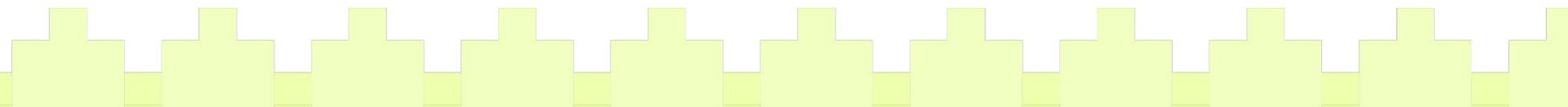
Hola hermanos, hermanas, estudiantes, profesores y toda la gente que vino a escuchar este tema. Espero que estén todos bien con sus Familias



Wünelu mu chaliwaiñ

Iñche Cristian Ahumada Oliva pigen, tuwün Santiaw waria mu , ka tañi chaw ka tañi ñuke ka tuwüigu tachi waria mu.

Me llamo Cristian Ahumada Oliva, soy de Santiago y mis padres también son de esta ciudad.



Wünelu mu chaliwaiñ

**Chillkatun Ingeniería Eléctrica tüfachi chillkatuwe ruka mu,
chillkatuken Magister en Ciencias, Mención computación.**

**Estudié Ingeniería Eléctrica en esta Universidad y estudio Magister
en Ciencias Mención Computación.**

Wünelu mu chaliwaiñ

Ka femgechi, chillkatuken Mapuzugun ka kimeltuken Mapuzugun ta mu, ETHICs mu ñi witxampüramafiel tüfachi zugun.

Además, estudio Mapuzugun y enseño Mapuzugun aquí, en ETHICs para revitalizar esta lengua.

Wünelu mu chaliwaiñ

Fewla mülepan ñi pegelafiel ñi Tesis küzaw:

“Diseño y desarrollo de un software lingüístico para apoyar el aprendizaje del Mapuzugun”

Ahora vine aquí a mostrar mi trabajo de tesis:

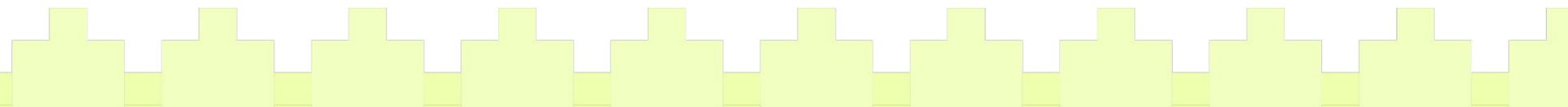
“Diseño y desarrollo de un software lingüístico para apoyar el aprendizaje del Mapuzugun”



Wünelu mu chaliwaiñ

Pewmagen gütxamkawaiñ afle tüfachi presentacion ñi zoy
kümekünual tüfachi küzaw.

Espero que podamos conversar al terminar esta presentación para
mejorar este trabajo.



Agenda

- **Motivación**
- **Antecedentes / Desafíos**
 - **Procesamiento Lenguaje Natural**
 - **Mapuzugun**
 - **Interacción Humano-Computador**
- **Problema a resolver y Objetivos**
- **Metodología**
- **Trabajo Futuro**

Motivación

Motivación

Unesco:

aché (PRY) ↓

avá-guaraní (ARG) ↓

chamacoco (BRA; PRY) ↓

chaná (ARG) ↓

chorote iyojwa'ja (ARG; PRY) ↓

chulupí (PRY) ↓

guaná (PRY) ↓

guaraní boliviano (ARG; BOL) ↓

guaraní kaiowá (BRA) ↓

guaraní mbyá (ARG; BRA; PRY;

URY) ↓

guaraní ñandeva (BRA) ↓

huilliche (CHL) ↓

kadiwéu (BRA) ↓

kahuéskar (CHL) ↓

kaingán (BRA) ↓

kinikinau (BRA) ↓

kunza (ARG; BOL; CHL) ↓

lengua (PRY) ↓

macá (PRY) ↓

manjui (ARG; BOL; PRY) ↓

mapuche (ARG; CHL) ↓

mocoví (ARG) ↓

ona (ARG) ↓

opayé (BRA) ↓

pilagá (ARG) ↓

puelche (ARG) ↓

quechua de Santiago del

Estero (ARG) ↓

sanapaná (PRY) ↓

tapieté (ARG; BOL; PRY) ↓

tehuelche (ARG) ↓

tereno (BRA) ↓

toba (ARG) ↓

toba-mascói (PRY) ↓

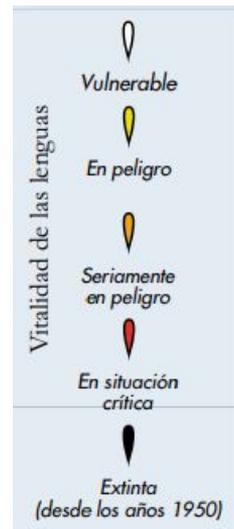
vilela (ARG) ↓

wichí (ARG; BOL) ↓

xetá (BRA) ↓

xokleng (BRA) ↓

yámana (CHL) ↓



*“90% de todas las lenguas podrían ser sustituidas por las dominantes de aquí a finales del Siglo XXI”**

*Fuente: “Vitalidad y peligro de desaparición de las lenguas”. Unesco 2003.

Motivación

Importancia de la lengua



“Nuestros idiomas son fundamentales para la afirmación de nuestras identidades, para la continuidad de nuestras culturas y cosmovisiones y para la transmisión intergeneracional de nuestros conocimientos. Son una expresión de libre determinación, diversidad y resistencia”

Tarcila

Rivera

Zea

Camacho, L., Zevallos, R. (2019) Siminchikkunarayku. Lingüística computacional para la revitalización y el poliglotismo. Hoja de Ruta. Fundación Siminchikkunarayku, Pontificia Universidad Católica del Perú.

*Autor Fotografía: UN Photo/Eskinder Debebe



Lonco Pascual Coña

Testimonio de un cacique mapuche

Bilingüe mapudungun / castellano

Biblioteca del Bicentenario

pehuén®

Kuifike mapuche yem:
chumwechi ñi admo:efel eḡn.

Kiñe dēḡu pian:

Deu fūchan iñche, doi pu-
ra mari tripantu nielu tro-
kiuwn. Fei tēfachi fentren
ñi moḡenum kimyepañiñ ñi
chumkefel kuifike che yem,
kom ñi kakeume admoḡefel
kinnieken; kümeke adḡe-
fuiḡn, weshake ad ká
:efuiḡn.

Itrokom tēfá nütramafiñ
feula: konümpaiafiñ chum-
ḡechi ñi tremmum iñche ka
chumwechi ñi moḡefum fū-
chake l'aku eni.

Feula kalewetui moḡen;
tēfachi weche mētewe wiḡ-
katuiḡn; allwe ḡoimarpuiḡn
taiñ küpal ñi ḡülam ka ñi
dēḡu; kalli rupape kiñe muḡü
tripantu, fei meu epe kim-
werpulaiai ñi mapu dēḡun
eḡn.

Fei meu, dēḡulpe tēfachi
lifro eḡn kiñeke naḡ rume!

Piken mai ta tēfá.

Pascual Coña.

Fewla kalewetui mogen; tūfachi
weche mütewe wigkatuigün; allwe
goirmarpuigün taiñ küpal ñi gülam
ka ñi zugu; kalli rupape kiñe muḡü
txipantu, fei meu epe kimwerpulaiai
ñi mapu zugun egün.

Fei meu, zugulpe tūfachi lifro egün
kiñeke naḡ rume!
Piken mai ta tūfa.

Pacual Coña.

Extensión histórica

Según fuentes históricas, en el territorio existían 1.000.000 de hablantes en el siglo XVI, en todo el territorio entre Coquimbo y Chiloé.

Hoy se estima que existen entre 200.000 y 300.000 hablantes del Mapuzugun.



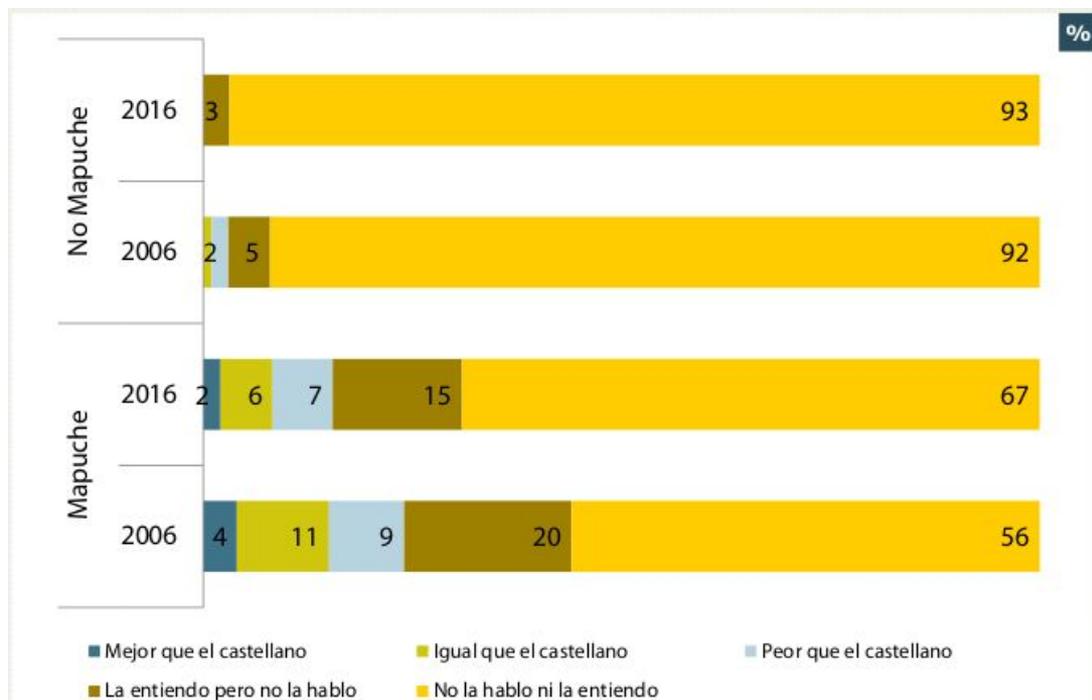
Motivación

Estado Actual

Gráfico extraído de:

“Los mapuche rurales y urbanos hoy” Marzo -Mayo 2016

Centro de Estudios Públicos, CEP



Motivación

No fue un proceso “Natural” de pérdida

- **Invasiones**

- La campaña del desierto (Argentina)
- La “pacificación” (ocupación) de la Araucanía (Chile)

-

- **Exclusión de Espacios Educativos**

- No existen muchas instancias oficiales de aprendizaje y enseñanza **del** Mapuzugun o no tienen suficiente apoyo.
- No hay aprendizaje y enseñanza **en** Mapuzugun



Motivación

El Mapuzugun en Peligro

¿Que se puede hacer?:

Programas de revitalización

- **Estatus:**
 - Dónde se usa
 - Quién lo usa
 - Para qué se usa
- **Espacios educativos:**
 - Aprendizaje y enseñanza **del** Mapuzugun
 - Aprendizaje y enseñanza **en** Mapuzugun





Antecedentes / Desafíos

Disciplinas que involucra este trabajo

- Procesamiento de Lenguaje Natural (PLN).
- Interacción Humano-Computador (HCI).
- Educación por medio de computadores o Softwares educativos.



Antecedentes / Desafíos

Tipología del Mapuzugun

Es una lengua aglutinante y polisintética:

Kim Mapuzuguyekümeaigün

Kim Mapu / zugu / yekü / me/ a / igün

“Iran aprendiendo a hablar Mapuzugun”

Antecedentes /Desafíos

No hay una estandarización del Mapuzugun

Ngütramkan

Gütxamkan

Gvxamkan

(Conversación)



Antecedentes / Desafíos

No hay una estandarización del Mapuzugun

LLampüdken

LLampüzken

Jampvzken



(Mariposa)

Antecedentes / Desafíos

Grafemarios más utilizados del Mapuzugun

Alfabeto Mapuche Unificado				
A a	Chch	D d	E e	F f
G g	I i	K k	L l	Ḷ ḷ
Ll ll	M m	N n	Ñ ñ	Ṇ ṇ
Ng ng	O o	P p	R r	S s
T t	Tr tr	Ṫ ṫ	U u	Ü ü
W w	Y y			

Grafemario Azümchefe				
A a	Z z	Ü ü	M m	Chch
E e	F f	I i	K k	T t
Nhnh	Tx tx	O o	Y y	Q q
G g	Lh lh	Ñ ñ	R r	S s
Ll ll	P p	U u	W w	L l
N n	Sh sh	T' t'		

Grafemario Mapuche Raguileo				
A a	C c	Z z	E e	F f
I i	K k	L l	B b	J j
M m	N n	H h	G g	Ñ ñ
O o	P p	R r	S s	T t
U u	V v	X x	W w	Y y
Q q				

* A este trabajo no le corresponde tomar una decisión con respecto al estándar.

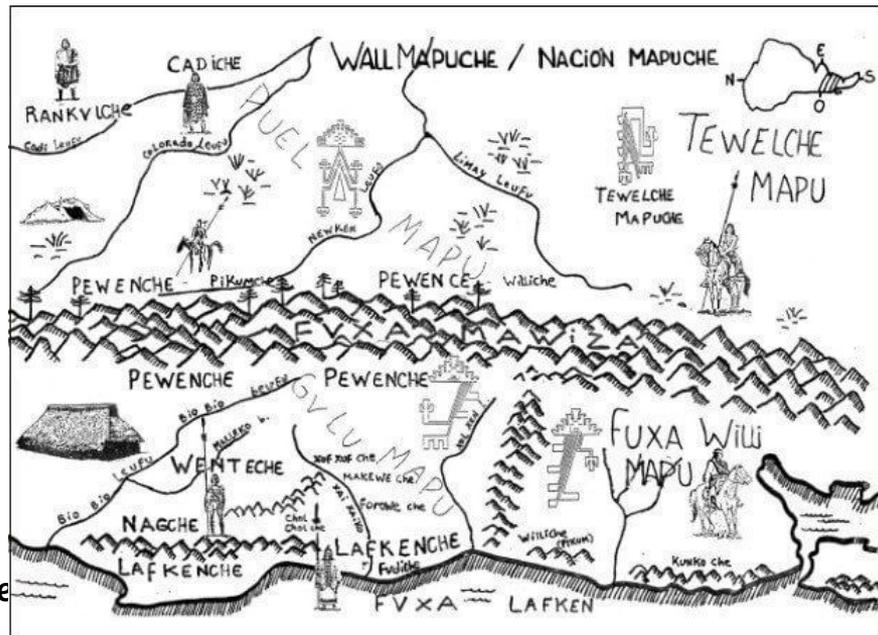
Antecedentes / Desafíos

Diversidad Territorial

Existe una gran diversidad de hablantes dependiente del territorio de origen:

- Pronunciación
- Léxico
- Escrituras más usadas

*Sigue siendo la misma lengua y por lo tanto no hay problemas de entendimiento entre las variantes territoriales



Antecedentes / Desafíos

Falta de corpus y herramientas

Bajo corpus (no estandarizado):

- **Libro, informes y publicaciones disponibles en internet.**
- **Corpus más grande 170 hrs anotadas de Mapuzugun (CMU)**

Existen algunos softwares y herramientas básicas:

- **Analizador morfológico y Normalizador* (Andrés Chandía)**
- **Mapudungun Mew (CEDETI UC)**
- **Mapu (PUCV)**

Otras lenguas minorizadas y trabajos sobre ellas

No solo el mapuzugun:

Compendio de trabajos sobre Quechua, Nahuatl, Wixarika, Shipibo Konibo, Guaraní, Mohawk, Plains Cree, Yorem Nokki, entre otras.

<https://github.com/pywirrarika/naki>

- Desafíos sobre los corpus insuficientes, identificadores de lenguas, traductores, desambiguadores y analizadores morfológicos.
- Hoja de ruta sobre la “Lingüística computacional para la revitalización y el poliglotismo” (BLARK en el caso de Perú)
- BLARK para lenguas minorizadas (Euskal Herria)



Problema a resolver y Objetivos

Preguntas de investigación

Dada la no estandarización, diversidad actual y falta de corpus del Mapuzugun:

- ¿Es posible desarrollar herramientas lingüísticas computacionales para esta lengua?
- ¿Qué herramientas básicas de procesamiento de lenguaje natural se necesitan para el apoyo a su aprendizaje y enseñanza?
- ¿Cuáles de estas herramientas se pueden desarrollar?

Preguntas de Investigación

- **¿Qué recursos se necesitan para poder desarrollar las herramientas que todavía no son alcanzables en un corto plazo para contribuir al desarrollo de la lingüística computacional del Mapuzugun?**
- **¿Es posible desarrollar un software amigable y útil para los estudiantes y profesores de Mapuzugun?**

Hipótesis

Es posible, dado el estado del Mapuzugun, desarrollar y disponibilizar un software simple para el Mapuzugun que:

- a. aporte a la revitalización del Mapuzugun impactando en el ámbito de la enseñanza de esta como segunda lengua, mejorando el aprendizaje de los estudiantes de esta lengua.**
- b. contribuya a futuros investigadores y desarrolladores de aplicaciones de PLN de Mapuzugun.**

Objetivo General

Desarrollar un software simple que permitirá al estudiante inicial de Mapuzugun consultar palabras y frases, tanto para obtener su significado como para su análisis morfológico. Estará basado en herramientas computacionales para el procesamiento lingüístico de esta lengua. Este software estará orientado a estudiantes (particularmente iniciales) de la lengua.

Se espera además que, parte de este desarrollo, sirva de base para el desarrollo de herramientas más complejas, tanto para el ámbito de la lingüística computacional del Mapuzugun, como el de la revitalización lingüística referido al aprendizaje y enseñanza de esta lengua.

Objetivos específicos:

1. Diseñar e implementar un traductor inter-grafemarios del Mapuzugun, para los tres grafemarios más reconocidos, tomando en cuenta la inconsistencia de los textos.
2. Adaptar el analizador morfológico para el Mapuzugun existente, que pueda segmentar las palabras, de forma lingüística, para su uso en espacios de enseñanzas, filtrando aspectos para el entendimiento de personas no expertas en lingüística.
3. Desarrollar e implementar un traductor informal a partir del analizador morfológico existente, esto es, que dé al estudiante nociones e intuiciones de cómo se podría traducir formalmente una cierta palabra y/ o frase del Mapuzugun.

Objetivos específicos:

4. Validar las herramientas básicas (1)-(3) desarrolladas, con expertos en la lengua y con estudiantes de Mapuzugun.
5. Integrar el traductor inter-grafemarios, el analizador morfológico modificado y el traductor informal en un proceso que muestre el análisis de palabras y frases.
6. Generar, a partir de las herramientas validadas, un software simple que pueda ser usado por quienes enseñan y/o estudian el Mapuzugun, junto con una prueba de usabilidad y un estudio de su impacto en la enseñanza y aprendizaje.

Metodología

Metodología.

Recolección de datos

- **Corpus escrito del CMU (170 horas de audio transcritas en Mapuzugun con distintas variantes territoriales)**
- **Textos de instituciones en Mapuzugun (Conadi, INDH, Amnistía Internacional, entre otros)**
- **Libros y publicaciones disponibles en internet en Mapuzugun.**

Metodología

Recolección de datos

41 archivos	UN:9,	AZ:11,	RA:10,	KW:11	
1.139.365 palabras	UN:1.031.737,	AZ:54.610,	RA:28.217,	KW:24.801	
149.320 pal. únicas	UN:130.826,	AZ:8.378,	RA:5.454,	KW:4.662	

UN: Grafemario Unificado.

AZ: Grafemario Azümchefe.

RA: Grafemario Ragileo.

KW: Otro grafemario.

*Las palabras de distinto grafemario se consideran como distintas en esta cuenta.

Metodología

Desarrollo e Integración de tres herramientas.

- **Traductor de Grafemarios**
- **Segmentador/Analizador Morfológico**
- **Traductor informal**

Metodología

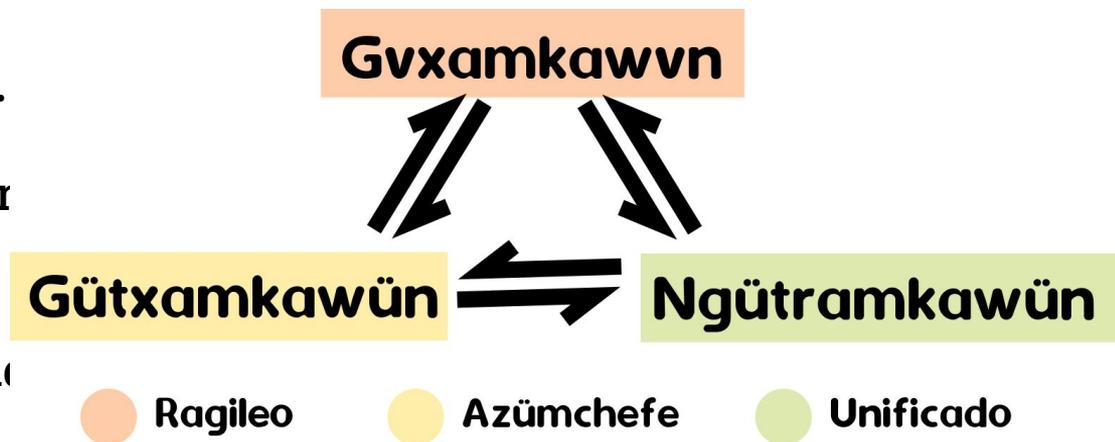
Traductor de Grafemarios

Conversión entre 3 Grafemarios.

Toma en cuenta ciertas variaciones comunes.

Base para identificar los grafemarios.

Se valida con los datos disponibles de cada grafemario.



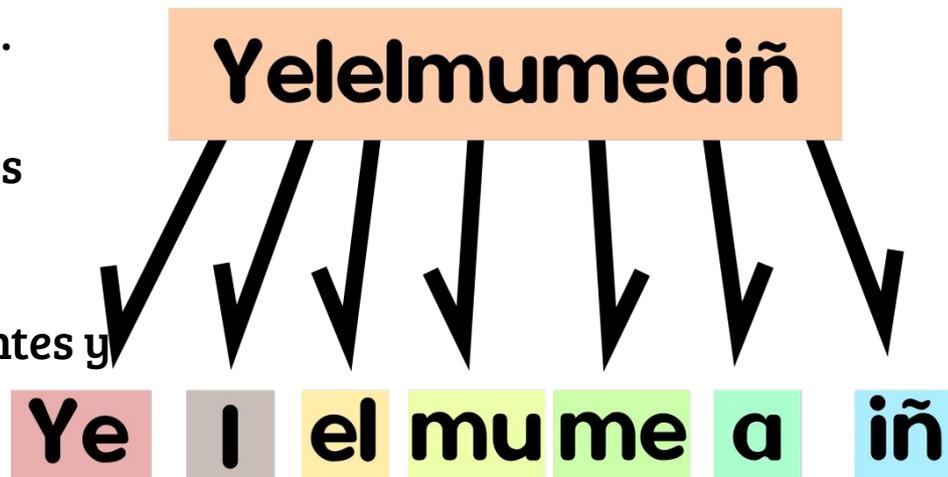
Metodología

Segmentador / Analizador Morfológico

Adaptación de un analizador morfológico existente y su definición morfemas para personas sin conocimientos lingüísticos.

Segmentación se valida a partir de datos conocidos.

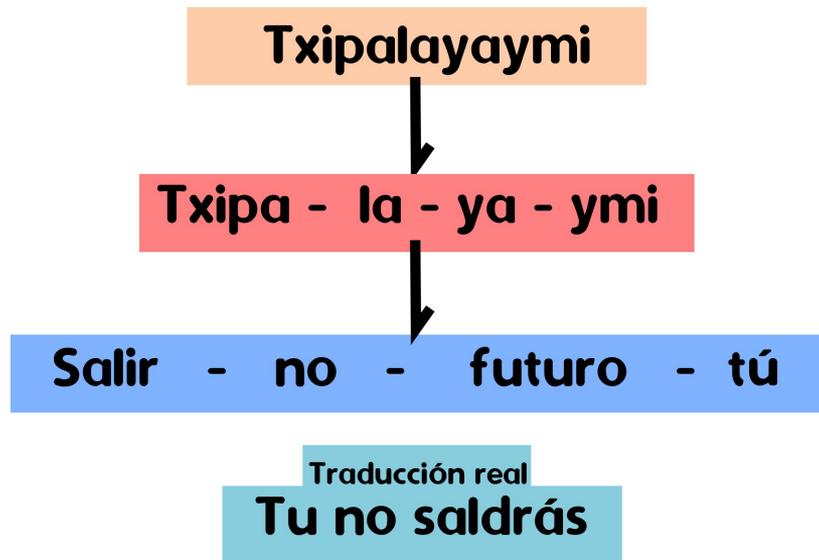
Interpretabilidad se valida con estudiantes y profesores de Mapuzugun



Metodología

Traductor Informal

A partir de la segmentación se genera una Traducción Informal, que se acerca más a la Traducción Real, pero que no necesariamente está correcta gramaticalmente.



Metodología

Traductor Informal

A partir de la segmentación se genera una Traducción Informal, que se acerca más a la Traducción Real, pero que no necesariamente está correcta gramaticalmente.

Nentumekelafuy

Nentu-me-ke-la-fu-y

Sacar-ir a-hábito-no -pasado-el/ella

Traducción real

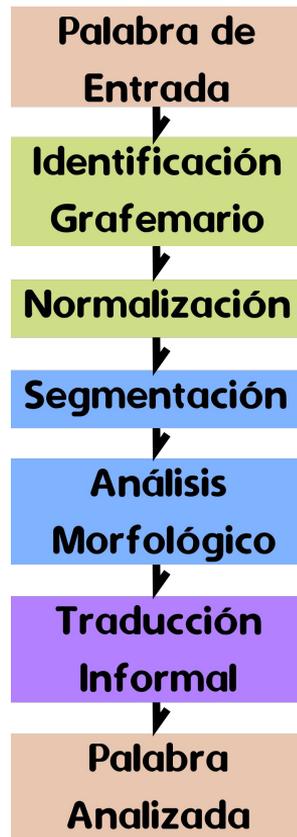
El/ella no iba habitualmente a sacar

Metodología

Integración de las Herramientas.

El usuario ingresa una palabra o frase.

El programa le entrega una palabra o frase analizada.



Metodología

Generación de interfaz básica para el software

Se buscará mantener una interfaz simple. Falta definir:

- Aplicación web , software descargable y/o aplicación móvil.
- Uso de internet o autocontenido (en caso de ser aplicación o software descargable)
- Solo proceso completo o también división por partes.



Metodología

Evaluación y Validación

Se evaluará la herramienta integrada en dos aspectos:

- **Lo técnico: Usabilidad y buen funcionamiento del software.**
 - Pruebas con usuarios, métodos estandarizados.
 - Verificación de criterios de usabilidad.
- **Lo Social: Utilidad percibida por estudiantes y profesores.**
 - Se diseñará una evaluación específica para este punto. Que tome en cuenta como cambia el rendimiento en tareas básicas de la lengua con y sin el uso del software.

Trabajo Futuro

Trabajo Futuro (más allá de este trabajo)

Responder, dado la diversidad, los escasos corpus y la no estandarización:

- ¿Qué herramientas básicas de Procesamiento de Lenguaje Natural, se requerirían para generar aplicaciones más complejas como Traducciones automáticas, NER, etc?
- ¿Qué recursos se necesitan para poder desarrollar las herramientas que todavía no son alcanzables en un corto plazo y que contribuyan al desarrollo de la lingüística computacional del Mapuzugun?

Trabajo Futuro (más allá de este trabajo)

**Generar investigación sobre el Mapuzugun en computación:
Natural Language Processing.**

- Traductor Automático
- Transfer learning
- Reconocimiento de Voz
- **Human Computer Interaction.**
 - Softwares de aprendizaje
 - Herramientas
 - Implementaciones más accesibles de herramientas de NLP.
- **Darle Estatus a la Lengua a través de su inclusión en la Academia, tanto como objeto de estudio como parte de la vida cotidiana (científica).**



Trabajo Futuro (más allá de este trabajo)

Zoy Kim Mapuzugual (Kom iñchiñ)

Aprender más Mapuzugun (Tod@s nosotr@s)

**Chaltu may tamün allkütumufiel.
Mülele ramtun,
Ramtumuchi müten.**

Agradecimientos y Apoyo

Programa de Pueblos Indígenas de la FCFM, Universidad de Chile. Por el apoyo económico a esta tesis.

Además toda la colaboración de:

- **Alberto Huenchumilla y Mapuzuguletuañ**
- **Andrés Chandía, Universidad de Barcelona**
- **Iñaki Alegría, Universidad del País Vasco**
- **Antonis Anastasopoulos, Carnegie Mellon University**
- **Felipe Bravo, Universidad de Chile**
- **Francisco Gutierrez, Universidad de Chile**