



# Dungu Peyem: Analizador y generador morfológico para mapudungun (FST)

Autor: Andrés Chandía. Director de tesis: Toni Badía. [Universitat Pompeu Fabra]

**Dungu:** palabra, habla  
**pe:** proximidad temporal y/o física con los hechos  
**ye:** rasgo constante  
**m:** marca de verbo nominal instrumental instrumento o locación.  
"herramienta del lenguaje".

Los mapuche habitan al sur de Argentina y Chile, en donde son más numerosos, cerca de 700 mil personas, de ellos sólo un 30% son hablantes de mapudungun. Es una lengua polisintética aglutinante, tiene una estructura interna compleja, compuesta por una serie de morfemas que siguen a la raíz. La composición e incorporación son productivas en la morfología verbal.

Nuestro propósito es crear herramientas para dotar al mapudungun de corpórea lingüísticamente anotado. El sistema principal es un analizador morfológico capaz de procesar los inputs lingüísticos, y de generar unidades idiomáticas bien formadas. Algunos problemas son p. ej. la falta de un sistema ortográfico estandarizado. Creemos que las técnicas de FST son las apropiadas para enfrentar la morfofonología del mapudungun.

## Análisis

Reconocimiento de la raíz y de cada uno de los afijos que conforman la estructura de la palabra. En la literatura lingüística se suele representar de la siguiente manera:

Ule i -dengül-a -ymün  
mañana comer-judías -NRLD-2p  
'Mañana comeréis judías'

El dungu Peyem entrega los resultados de esta forma:

**Input:** kimpalay

**Output:**

kimpalay ==> [@TV][=conocer, aprender, saber, recordar][+HH@pa17][+NEG@la10][+IND@üy4][+3@Ø3]

Palabra analizada seguida de flecha al análisis, cada componente encerrado entre paréntesis cuadrados, primer componente categoría g. de la raíz. Significado de la raíz en castellano. Sufijos introducidos por +. Cada sufijo: abreviatura del tipo de sufijo. Forma del sufijo introducida por @. Número de slot, en formas verbales.

## Complejidad de las formas verbales del mapudungun

Lengua no normativizada → variación escrita.

Variación fonológica, intercambiabilidad dentro de la misma variante (mismo hablante):

dunguwekieli ~ dunguwekeeli ~ dunguwekeli  
pichin ~ pichiñ | arkün ~ arken  
ponwi ~ ponui

Inserción de fonemas ü, g, ñ  
kimle → kimüle | leliliñ → leliliñi  
motrike → motriñke

Geminación → simplificación  
konnuli → konuli

Movilidad de sufijos:

Sufijos comprendidos entre slots 16 y 27.

Reduplicación de raíces.

Incorporación en las raíces verbales.

Composición nominal.

De la descripción lingüística a la implementación computacional



Dungu Peyem: algoritmo de procesamiento de cadenas textuales mediante transductores de estados finitos, se procesa un input con ciertas restricciones formales introducidas a través de expresiones regulares que reflejan la morfo-táctica del mapudungun. Se produce un output, análisis morfológico, o se compone una forma léxica a través de la generación, proceso inverso al análisis.

## Generación

Es el proceso inverso al análisis, se introducen en la interfaz los componentes (la glosa) de una palabra y el sistema retorna la palabra:

**Input:** [@N][=sueño][+VRB@tu36][+CF@ke14][+IPD@fu8][+IND@üy4][+1@Ø3][+PL@iñ2]

**Output:**

[@N][=sueño][+VRB@tu36][+CF@ke14][+IPD@fu8][+IND@üy4][+1@Ø3][+PL@iñ2] ==> umawtukefuyiñ



<http://www.chandia.net/content/dungu Peyem>

## Proceso de implementación

Tablas de restricciones: Obligatoriedad

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1			AH	AH	23	23	23		20	17	17		20	10	6	6	4	4	4	3P	3P
2	Slot		Ø-e	Ø-0	nge	uw	mu	me-e	pa	pu	me-e	nu-	e	fi	üy	ül	Ø	ün	i	chi	
3			vt																		
4			vi																		
5	36V	nge	(i)																		
6	36V	l	(t)																		
7	36S	nge	(l)																		
8	36S	ye	(t)																		
9	36S	Ø	(i)																		
10	35	ñma	-0																		
11	34	üm	-0																		
12	23	uw																			
13	23	mu																			
14	M-18	ür																			
15	M-18	yekü																			
16	10	la																			
17	10	ki																			
18	10	nu-																			
19	8	fu																			
20	6	e																			
21	6	fi																			
22	3P	ün																			
23	3	Ø3																			
24	3	ng																			
25	3T	y3																			

Tablas para la aparición obligatoria de sufijos, para la aparición condicionada opcional, para la coaparición, exclusión.

## A grammar of Mapuche

(Ineke Smeets)

Matriz de slots para las formas verbales

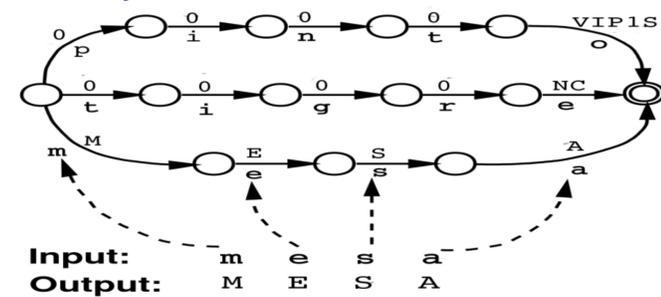
Slots

Raíz + 36 - 35 - 34...26 - 25 - 24...03 - 02 - 01

Exclusión por razones gramaticales o semánticas. Slots 1-15 sufijos flexivos, posición fija, persona, número, modo, aspecto, tiempo, valor de verdad, negación y nominalizadores. Slots 16-27, sufijos derivativos, modificadores semánticos. Slots 28-36, marcadores de aspecto y valencia. Una forma verbal mapuche es una raíz + 1 o más sufijos derivativos opcionales y al menos un sufijo flexivo.

## Finite state morphology

(Beesley & Karttunen)



## Expresiones regulares (regex)

Secuencias de caracteres que forman un patrón de concordancia con otro lenguaje (natural). Las regex tienen una sintaxis y una semántica fijas y estrictas, cada carácter de una regex tiene un sentido meta-lingüístico o literal y puede ser utilizado en la construcción del patrón de concordancia que permite procesar texto mediante diferentes aplicaciones, búsqueda y reemplazo es la más común. Las regex contienen metacaracteres que pueden expresar, por ejemplo, cualquier carácter del lenguaje al que hacen referencia, rangos de caracteres [a-z], [0-9].

## Morfotáctica

El estudio y modelado de la formación de palabras de los lenguajes naturales se llama Morfotáctica o Morfosintaxis. «Los dos problemas fundamentales de cualquier procesador morfológico son:

- Reglas de formación de palabras (o morfo-táctica): los morfemas no se combinan libremente. P. e., una raíz verbal no se puede unir con un sufijo nominal. Es necesario establecer las combinaciones válidas de morfemas...
- Reglas de alomorfía: los morfemas pueden presentar variantes en alternancia para distintos contextos... cont + ar / cuent + o  
A. Moreno. Lingüística computacional. Ed. Síntesis. Madrid 1998.

## Reglas nivel gramatical o abstracto

# RULES for single non verbal roots  
# Defines how stems are formed for Adjective, Adverbial, Noun, Numeral and Interrogative roots  
define RuSTEMWITH ["@AJ" => ?\* \_ ?\*  
[" +VRB""@0"] [" +VRB""@e"] [" +VRB""@ns"] ?\*]  
.o. ["@AV" => ?\* \_ ?\* [" +VRB""@0"] [" +VRB""@l"] ?\*]  
.o. ["@N" => ?\* \_ ?\* [" +VRB""@0"] [" +VRB""@e"] [" +VRB""@ü"] [" +VRB""@l"] ?\*]  
[" +VRB""@i"] ?\*].o. ["@NU" => ?\* \_ ?\*  
[" +VRB""@0"] [" +VRB""@e"] [" +VRB""@l"] ?\*].o.  
["@QC" => ?\* \_ ?\* [" +VRB""@e"] ?\*].o. ["@QT" => ?\* \_ ?\* [" +VRB""@l"] ?\*];

## Conclusión:

Este es un trabajo en desarrollo que hasta ahora ha ido incorporando paulatinamente la morfología verbal del mapudungun y que se encuentra, después de haber incorporado las otras partes de la oración, en un proceso de refinamiento e inclusión de algunos fenómenos más complicados de esta lengua.

La versión en desarrollo del dungu Peyem se puede visitar en la web:

<http://www.chandia.net/content/dungu Peyem>