

RAÍCES DEL MAPUDUNGUN

IDENTIFICACIÓN DE RAÍCES Y AFIJOS EN CORPUS SIN ANOTAR A TRAVÉS DE LOS ALGORITMOS DE ANTONI OLIVER

Resumen

Este trabajo trata de la preparación de un corpus del mapudungun para la búsqueda automatizada de entidades morfológicas, y de fenómenos fonológicos reflejados en la escritura. La preparación del corpus tiene mucha importancia en el caso del mapudungun, porque, como veremos, existen ciertas circunstancias que generan una gran diversidad de maneras de escribir esta lengua, y justamente es el soporte escrito al cual nos acercamos para llevar a cabo nuestro propósito.

En segundo término, y una vez superada la primera etapa -o casi- intentaremos aplicar los algoritmos que Antoni Oliver utilizó en su tesis doctoral (Oliver, 2004), para la identificación de raíces y afijos del mapudungun; con esto pretendemos hacer una primea aproximación a la clasificación de los verbos del mapudungun. De momento sólo intentaremos identificar raíces y afijos para ver si el algoritmo funciona como esperamos y, en caso contrario, modificarlo para que nos de los resultados esperados, de esta manera podremos continuar con la labor de etiquetar un corpus para poder utilizarlo en un futuro analizador morfológico. Uno de los aportes de un trabajo como este es que quizás podamos medir la frecuencia de coaparición entre raíces y afijos, y de esta manera hacer una clasificación preliminar de las raíces verbales; en este sentido, podremos observar si existen raíces netamente verbales o netamente nominales, o si por el contrario, siempre su comportamiento depende del afijo que las complementa.

ÍNDICE DE CONTENIDO

Resumen.....	1
Introducción.....	3
Preparación del corpus.....	4
Norwirin Mapudungun Trapümf.....	6
La morfología verbal del mapudungun.....	7
Algunos ejemplos.....	7
Otros afijos del mapudungun.....	9
Negación.....	9
Futuro.....	9
Implicatura rota.....	10
Mirativo, reportativo.....	10
Continuativo.....	10
El corpus.....	10
Planificación para el objetivo.....	11
Paradigma no supervisado.....	11
Formario incompleto.....	11
Proceso.....	12
Bibliografía.....	13
Anexo 1.....	14
Lista de fuentes del corpus.....	14

Introducción

El mapudungun es una lengua polisintética con incorporación a las raíces que pueden generar tanto nombres como verbos y adjetivos, se calcula que actualmente hay unos 900.000 mapuche en Chile y unos 80.000 en Argentina, y aunque no todos ellos hablan su lengua nativa, la gran mayoría expresa un gran interés por recuperar y utilizar el mapudungun. En cuanto a la morfología, podemos decir que tiene dos caras, una relativamente sencilla que corresponde a las partes del discurso que no incluyen al verbo, que es la otra cara de la moneda, la compleja...

"Echando una mirada de conjunto sobre las clases de palabras mapuche, y sus funciones dentro de la oración, el espíritu se halla de repente delante del cuadro de una antigua ciudad; la fantasía le aparenta una multitud de casas humildes, cada cual con su condición propia, pero de impresión increíblemente concorde, todas situadas alrededor de la gruesa mole de una catedral majestuosa, la que domina por completo y unifica la multiforme variedad circundante. Es la imagen de la lengua mapuche. El verbo con su inmensa riqueza de formas y su potencia expresiva casi sin límites, sobrepuja en mucho la posible aplicación de las demás partes de la oración. El idioma mapuche por antonomasia es la lengua del verbo." (Moesbach, 1963).

El verbo mapuche es sufijal, aglutinante y complejo, los análisis morfológicos establecen aproximadamente unas 36 posiciones de sufijación al verbo (Smeets, 1989). Pero lo normal en el discurso es que el verbo se realice con cuatro, cinco o seis morfemas.

A la raíz verbal le siguen los morfemas que expresan aspecto, tiempo, inversión, voz, dirección y concordancia con el objeto. También puede ocurrir la incorporación de otra raíz. El orden de los morfemas suele ser fijo, y se producen cambios morfemáticos sencillos y lógicos en los límites de morfema.

(1) *rüngkükonfemtwaymi*

inmediatamente saltarás hacia adentro, de vuelta al punto de origen

<i>rüngkü</i>	raíz de saltar	<i>a</i>	futuro
<i>kon</i>	adentro	<i>i</i>	enunciado de hecho
<i>fem</i>	inmediatamente		(no hipótesis)
<i>tu</i>	hubo un salto previo	<i>m</i>	segunda persona
	hacia afuera	<i>i</i>	singular

Obviamente la idea sería etiquetar todas las raíces y todos los afijos posibles, entre otras cosas porque en el caso de las lenguas aglutinantes, y en el del mapudungun especialmente, los afijos proveen de mucha información semántica, lo que podría ayudar en los procesos de desambiguación a la hora de la traducción, pero eso ya son palabras mayores, de momento. Mientras tanto intentaremos llevar a cabo la tarea comentada en el resumen, intentar aislar raíces y sufijos, ver si existen raíces que sólo incorporan sufijos relacionados con la categoría verbal, o al contrario, identificar aquellas raíces que no admiten este tipo de afijos.

Preparación del corpus

El mapudungun es una lengua que comenzó a ser escrita con el alfabeto latino por los jesuitas que fueron a evangelizar a los mapuche. Se encontraron con la dificultad de que ciertos sonidos del mapudungun no tenían una correspondencia con las grafías del castellano e introdujeron algunas variaciones en los grafemas. El mapudungun aún no tiene una grafía fijada o normativizada y desde la primera gramática del siglo XVII hasta nuestros días se ha propuesto distintas soluciones gráficas u ortográficas para esta lengua.

Los sonidos del mapudungun cuentan con seis vocales, cinco iguales al castellano y una sexta que se realiza de dos maneras según el contexto, la primera corresponde al símbolo / i / del AFI, y la segunda al / ə /. Las diferentes propuestas grafemáticas que ha recibido este sonido, a veces una para cada realización, han sido variadas: ə, ú, ü, ĩ, v, y, ö, etc.

En cuanto a los diptongos que se forman en la pronunciación y su correspondiente representación gráfica, es más una cuestión de normas ortográficas, que al no estar fijadas, permiten que la representación escrita sea de libre aplicación, por lo tanto, diversa. Los sonidos semi vocálicos representados según el AFI son / w, y, g / (este último es velar similar al que se pronuncia en la palabra castellana 'agudo'). En la representación de estos diptongos encontramos -we, ue-; -ao, aw-; -uí, uy-; -ie, ye-; etc.

La mayor variación de la representación gráfica de los sonidos consonánticos del mapudungun, proviene de aquella que refleja los sonidos que no existen distintivamente en el castellano, como los interdental: / ɭ, ɲ, ʈ /, para estos sonidos tenemos por ejemplo: ɭ-, ɭ, 'ɭ, ɭ', ɭh; ɲ-, ɲ, 'ɲ, h, nd; ʈ, 'ʈ, ʈ-, ʈ, td; etc.

Presentamos aquí una tabla con las principales propuestas grafemáticas para el mapudungun. Sólo representamos los grafemas que difieren entre las propuestas y del castellano.

Tradicional	Académico	Raguileo	Nhewenh	Azümchefe	chandia.net
Vocales					
ũ, ə	ũ, ĩ	v	v	ü	ü
Semivocales					
w, u	w	w	w	w	w
y, i	y	y	y	y	y
q	g	q	q	q	g
Consonantes					
ch	ch	c	ch	ch	ch
ñ	ñ	ñ	nh	ñ	ñ
tr	tr	x	tr	tx	tr
t·	t̪	(t)	td	(t)	t'
n·	<u>n</u>	h	nd	nh	n'
l·	l̪	b	ld	lh	l'
d	d	z	sd	z	d
sh, ʃ	s, (sh)	s	sh	sh	sh
ll	ll, (lh)	j	lh	ll	ll
ŋ	ng	g	g	g	ng

Tabla 1: propuestas grafemáticas

Como ya hemos visto, no son estas las únicas maneras de escribir el mapudungun, a lo que tenemos que añadir lo que podríamos llamar una ortografía castellanizante; es el caso de palabras que habitualmente se escriben en mapudungun “según la ortografía mapuche”, de esta manera: *ruka*, *kom*, *kasike*, *linge*, *wapi*, *wemul*; pero que no es raro encontrar escritas así: *ruca*, *com*, *cacique*, *lingue*, *huapi*, *huemul*, de mano de personas mapuche, incluso coexistiendo la grafía castellanizante con dos maneras de escribir una misma palabra según las propuestas mapuche, por ejemplo, en un mismo escrito, hecho por la misma persona, podríamos encontrar la palabra *wagün*, *huagún* o *wagvn*, para transcribir la forma sonora mapuche del verbo 'gotear'.

Norwirin Mapudungun Trapümfē

Todo esto ha hecho que tengamos que desarrollar una herramienta para la preparación de un corpus con una grafía unificada, que nos pueda generar un buen soporte para la posterior investigación a través de medios informáticos. La grafía que hemos elegido, tomando en cuenta otras propuestas que no están reflejadas en la Tabla 1, son las de su última columna: “chandia.net”. Para las interdentales hemos preferido la representación con el apóstrofe por los siguientes motivos: “Pensemos primero en los escolares mapuche que tienen que escribir en mapudungun mientras les dicta el profesor; por ejemplo, si el alumno quiere subrayar un título o una palabra importante, luego tendrá dificultades para saber si determinada letra era subrayada o no, porque probablemente se solapará con el subrayado para resaltar. Pensemos ahora en los textos informáticos, es más rápido y fácil escribir un apóstrofe junto a la letra que un punto medio o subrayarla, ambas acciones requieren de la interacción de dos teclas a la vez, el apóstrofe, sólo de una. El apóstrofe no necesita de códigos especiales para representarse en formato electrónico, como el subrayado. Tampoco se confundirá en la escritura manual, como el punto medio”.

La herramienta informática que hemos desarrollado es el NMT, *Norwirin Mapudungun Trapümfē*, 'Unificador Ortográfico del Mapudungun'. Esta herramienta está disponible por web en: <http://www.chandia.net/herramientas>. Tiene una interfaz gráfica desplegada mediante java y php, a través de la cual se adquieren los datos y se entregan los resultados; internamente cuenta con un script que corre sobre linux aprovechando los programas, muy eficientes para el tratamiento de texto, que este entorno ofrece, awk, sed, cat y perl, entre otros. Básicamente, se hace un cambio de caracteres indicado a través de expresiones regulares contextuales y no contextuales; aprovechando este mismo sistema, se corrigen generaciones erróneas provenientes de las primeras reglas, o provocadas por el propio texto original. También se elimina el código informático que pueda tener un documento, sea html, xml o los códigos de los procesadores de texto.

Hablábamos más arriba de la representación de los diptongos, también hemos incluido este aspecto a través del script de transformación del NMT. En este sentido, creemos que la representación de diptongos y hiatos debería formar parte de una gramática normativa del mapudungun, mediante la cual se resolvería si es correcto escribir *troikug* o *troykug*, por ejemplo; en nuestro caso hemos optado por representar diptongos y hiatos según si se producen o no las semivocales, reflejando esto en la escritura, para nosotros la forma correcta es *troykug*.

A pesar de las muchas pruebas que hemos hecho con el NMT, no es posible llegar a un resultado perfecto porque, primero, la parte del input es muy variada, entiéndase las distintas grafías, incluyendo la castellanizante, la indistinta utilización de ellas por la misma persona en un mismo escrito y la utilización frecuente de soluciones personales. En segundo término está la coincidencia de algunas grafías de distintas propuestas para distintos sonidos, por ejemplo, el dígrafo *nh* representa al sonido que el AFI representa con el símbolo / ɲ / según la propuesta grafemática *Nhewenh* (*Ñewehñ*), sin embargo el mismo dígrafo, la propuesta *Azümchefe* (*Adümchefe*) lo utiliza para el sonido que el AFI representa con / ŋ /. En casos como este no hay

posibilidad de generar una regla contextual ya que ambos sonidos se dan en los mismos contextos, tendríamos que tener un corrector ortográfico con la forma que consideremos correcta en su base de datos, para poder solucionarlo.

Otro gran problema es la extensa utilización de palabras castellanas que hace el hablante (escribiente) mapuche, y como es de suponer, con las más variadas formas de escritura, desde la totalmente mapuchizante hasta el calco lingüístico, esto, sin embargo, es tema de otro estudio.

La morfología verbal del mapudungun

En mapudungun la forma verbal finita mínima contiene los morfemas de modo (real, hipotético, volitivo) y persona (1ª, 2ª, 3ª), además, si la persona es 1ª o 2ª, contiene la información de número (singular, dual, plural). Este conjunto de marcadores se llama “flexión verbal obligatoria finita” (Salas, 1992), cuya persona es la focal.

(2) a. El hombre vio a la mujer en la casa

b. <i>wentru</i>	<i>pefi</i>	<i>domo</i>	<i>ruka mew</i>
hombre	vio	mujer	en casa

Wentru es la persona focal y *domo*, la persona satélite, en este caso la persona focal corresponde al sujeto del castellano y la satélite al complemento directo, pero la frase mapuche es sólo superficialmente análoga a la del castellano (Zúñiga 2006).

(3) <i>domo</i>	<i>peeyew</i>	<i>wentru</i>	<i>ruka mew</i>
mujer	vio	hombre	en casa

La frase (3) podría traducirse como 'la mujer fue vista por el hombre en la casa' o 'a la mujer el hombre vio en la casa'. Sin embargo vemos que la forma verbal es diferente, pero en ambas frases corresponde a 'vio', la diferencia es que en la frase (2) la forma verbal indica que *wentru* es la persona focal y en (3) lo es *domo*.

Algunos ejemplos

(4) *lelin* 'yo miré', *leli-* es la raíz y *-n* expresa modo real, persona focal 1ª singular.

(5) *lelifiñ* 'yo lo(s)/la(s) miré (a él, a ella, a ellos, a ellas)', el morfema *-fi* significa persona satélite 3ª determinada paciente y *-ñ* es una variante de *n* que significa modo real, persona focal 1ª singular agente.

- (6) *lelyenew*, *-ew* significa persona satélite 3ª determinada agente; y *-n* significa modo real, persona focal 1ª singular paciente, se traduciría como 'él/ella me miró' o 'ellos/ellas me miraron'.
- (7) *lelingen*, el morfema *-nge* significa persona satélite 3ª indeterminada agente; y *-n* significa modo real, persona focal 1ª singular paciente, 'fui mirado', aunque la traducción sugiere una voz pasiva, éste no es el caso, ya que si el sufijo de persona satélite es *-nge*, el agente es indeterminado, lo que significa que no puede estar auto referido por una frase sustantiva en la cláusula; así
- (8) **nguru lelingen*, 'zorro fui mirado', o sea, 'fui mirado por el zorro'.
- (9) **kiñe che lelingen*, 'una persona fui mirado', o sea, 'alguien me miró' son agramaticales. Las formas aceptables llevan el sufijo *-e...ew*, como en
- (10) *nguru lelyenew*, 'zorro fui mirado por él', o sea, 'el zorro me miró'.
- (11) *kiñe che lelyenew*, 'una persona fui mirado por ella', o sea, 'alguien me miró', ya que *-e...ew* significa persona satélite 3ª determinada agente, y como tal puede ser auto referencial con una frase sustantiva de la cláusula.

En interacciones entre la 1ª y la 2ª personas, la lengua expresa además la composición del diálogo, como diálogo mínimo versus diálogo expandido. En el diálogo mínimo hay dos personas, un hablante y un oyente; en el diálogo expandido participan más de dos personas. Así,

- (12) *enlelyen*, el marcador *-e* significa persona satélite 2ª agente en diálogo mínimo; y *-n* significa modo real, persona focal 1ª singular paciente, luego, la forma verbal entera sólo puede significar 'fui mirado por ti', pero en
- (13) *lelimun*, *-mu* significa persona satélite 2ª agente en diálogo expandido y *-n*, modo real, persona focal 1ª singular paciente, luego, la forma entera sólo puede significar 'fui mirado por ustedes (vosotros)'.

Del mismo modo en

- (14) *lelye(i)yu*, *-e* significa persona satélite 2ª paciente; *-y* la persona focal es 1ª agente y *-u*, número dual, que incluye al paciente y al agente, luego, el diálogo es mínimo y el significado de la forma verbal entera sólo puede ser 'yo te miré', pero en
- (15) *leliwiyiñ*, *-w* significa persona satélite 2ª paciente; *-i* es la persona focal es 1ª agente y *-ñ*, el número plural que incluye agente(s) y paciente(s), la *-y* es epentética, sin valor, luego la forma verbal entera significa 'yo los miré a ustedes' ('yo os miré a vosotros', en castellano peninsular), nosotros te miramos, nosotros los miramos a ustedes.

Existe también una clase de sufijo, llamado “indirectizante” por Salas (Salas, 1992), que transforma al paciente en paciente indirecto o “ético”, como en

(16) *lelyen*, 'fui mirado por ti'

(17) *leliñmaen ñi mamüll*, 'fui mirado eso por ti mi leña', o sea 'me miraste mi leña'.

Hay indirectizantes que separan real o metafóricamente al paciente directo o “foco” de la acción del paciente indirecto o ético (1ª persona singular en el ejemplo); hay otros indirectizantes que implican el movimiento (real o metafórico) contrario, son aproximativos. En el ejemplo, el sufijo *-ñma*, 'indirectizante separativo' implica la idea de que miraste mi leña con codicia, para apoderarte de ella, alejándola de mí.

Las personas gramaticales están organizadas en una jerarquía de focalización

1ª --> 2ª --> 3ª determinada --> 3ª indeterminada

presente en una interacción, si la persona situada más alta en la jerarquía será focal y la más baja será satélite, sin incidencia del rol de agente o paciente de los participantes en la interacción.

La ausencia de marcas de número en la 3ª persona, la cuidadosa expresión del número de los participantes en el acto de habla, y la jerarquía interpersonal de focalización, revelan que el egocentrismo y el dialogismo son los principios organizadores básicos del sistema mapuche de marcadores de persona. (Salas, 1992).

Otros afijos del mapudungun

Negación

(18) *tripa-y*, 'salió'; *tripa-la-y*, 'no salió'.

(19) *tripa-lmi*, 'si sales'; *tripa-no-lmi*, 'si no sales'.

(20) *tripa-nge*, '¡sal!'; *tripa-ki(no)-l-nge*, 'no salgas'.

Futuro

(21) *wiya amu-y*, 'ayer fue'; *wüle amw-a-y*, 'mañana irá'.

(22) *wiya amu-la-y*, 'ayer no fue'; *wüle amu-la-ya-y*, 'mañana no irá'.

Implicatura rota

(23) *dewma mawün'-y*, 'ya llovió (acaba de llover)'; *dewma mawün'-fu-y*, 'ya llovió (y no sirvió de nada)'.

(24) *umawtuy*, 'durmió'; *umawtu-fu-y*, 'quería dormir (pero no pudo)'.

Mirativo, reportativo

(25) *aku-y*, 'llegó'; *aku-rke-y*, 'cuentan que llegó'.

(26) *trewa*, 'perro'; *trewa-rke*, ¡un perro!

(27) *domo kalku*, 'bruja'; *domo kalku-rke*, 'bruja resultó ser'.

Continuativo

(28) *müle-y*, 'está'; *müle-ka-y*, 'aún está'.

(29) *katrü-n*, 'cortar'; *katrü-ka-n*, 'hacer cortes (cortar y cortar)'.

(30) *ramtu-n*, 'preguntar'; *ramtu-ka-n*, 'interrogar (preguntar y preguntar)'.

(31) *nütram-y*, 'narró, contó'; *nütram-ka-y*, 'conversó'.

El corpus

El corpus que hemos obtenido está extraído desde diferentes fuentes (ver [Anexo 1](#)). Una vez pasado por la fase de preparación, lo sometimos a un proceso de tokenización y ordenamiento, finalmente obtuvimos 38.374 formas distintas sin ningún tipo de anotación. El proceso de tokenización en un comienzo fue automático, pero obtenido el primer resultado tuvimos que hacer una revisión y luego un proceso de limpieza asistido, es decir, utilizando herramientas informáticas pero texto por texto. finalmente unificamos todos los resultados en un solo archivo “02.completo.nmt.tok” que se puede encontrar en “*raices_mapudungun/corpus/textos_mapuche_nmt_tokenizados*”.

Planificación para el objetivo

Como ya comentamos, nuestro objetivo es aislar las raíces de los afijos y verificar si todas las raíces combinan con todos los afijos o si hay algunas restricciones que nos lleven a una primera clasificación de las raíces, en un primer momento de las eminentemente verbales, si dichas restricciones existen. Para alcanzar este propósito nos basaremos en gran medida en lo que se explica en el capítulo 4 de la tesis de Oliver (Oliver, 2004). Nuestro trabajo es el nivel más básico dentro del proceso que Oliver designa como “Descubrimient de la morfologia. Paradigma no supervisat. Formari incomplet”.

Tal como hizo el autor en su trabajo, nosotros utilizaremos el formalismo de descomposición morfológica, el cual supone que toda forma se puede descomponer en una pseudobase y una pseudoterminación y que la pseudobase se mantiene invariable para todos sus afijos. Una pseudobase es la parte que consideramos raíz de la forma al cabo de la aplicación del algoritmo de descomposición, y se le aplica este nombre porque puede coincidir o no, con la raíz real, lo mismo pasa con la pseudoterminación, puede coincidir o no con el o los sufijos; esto es así porque hay más de una división posible para cada forma, de esta manera podríamos generar un pseudo paradigma, el conjunto de pseudoterminaciones comunes a una o más pseudobases.

Paradigma no supervisado

Recordemos que el corpus que hemos generado no tiene ningún tipo de información adicional más que las propias formas. Oliver trabaja con el ruso y el croata, y en vista de ello su planteamiento consiste en completar el formario de entrada hipotetizando las formas a partir del proceso inicial de descubrimiento de la morfología utilizando el formario incompleto. Las formas hipotetizadas se verifican a través de una serie de herramientas informáticas como analizadores morfológicos y consultas a internet.

En nuestro caso, no contamos con ningún analizador morfológico (es lo que pretendemos llegar a construir), tampoco hay buscadores que den la posibilidad de seleccionar el idioma mapuche. Además, nosotros aún no queremos obtener los paradigmas del mapudungun, sólo queremos encontrar raíces y afijos, y ver cómo se interrelacionan. En este sentido es que nuestro trabajo está en un nivel inicial.

Formario incompleto

Oliver da este nombre a la lista de formas que no abarca todas las posibilidades de realización, es decir aquella que no contiene todos los paradigmas. De una lengua aglutinante como el mapudungun sería imposible tener todas las combinaciones de afijación, aunque

teóricamente podrían darse todas las formas de los paradigmas básicos, por lo menos para el mapudungun, ya que en este sentido los paradigmas son relativamente pocos y no muy extensos. Si utilizamos el algoritmo de adquisición de la morfología deberíamos poder obtener el número de raíces encontradas, si contáramos con algún tipo de anotación, o de base de datos de categorías gramaticales, deberíamos poder obtener el número de raíces que contiene cada categoría gramatical.

Proceso

En primer lugar se hará la segmentación de todas las formas de la lista, esto nos arrojará muchas pseudobases y pseudoterminaciones, las que tendremos que agrupar; posteriormente hemos de hacer una simplificación que nos permita obtener las raíces y los afijos correctos.

Inicialmente, y siguiendo a Oliver, intentaremos llevar a cabo la tarea con un corpus reducido compuesto por 4 raíces con todas las variantes que de ellas encontremos, por ejemplo *kūdaw* 'trabajo', *kūdawfe* 'trabajador', *kūdawntu* 'lugar de trabajo, taller', *kūdawkawe* 'herramientas', *kūdawūn* 'trabajar', etc. En total, el corpus reducido tiene 1.081 formas cuyas raíces son *ad-*, *dwam-*, *kūdaw-*, *maw-*.

Las heurísticas de simplificación que utilizaremos serán:

1. Número mínimo de raíces asociadas a un sufijo: los sufijos que tengan un número de raíces asociadas inferior a 2 serán eliminados.
2. Simplificación por raíces: si resulta que el algoritmo encuentra más raíces de las cuatro con que contamos en el corpus reducido, estas serán eliminadas porque lo que venga después de la raíz que conocemos debería formar parte de los sufijos.

Bibliografía

Gelbukh^{1,2}, Alexander; Sidorov¹, Grigori; Velásquez¹, Francisco (2004) Análisis morfológico automático del español a través de generación. México D. F.: 1 Centro de Investigación en Computación (CIC), Instituto Politécnico Nacional (IPN). Seúl: 2 Universidad Chung-Ang.

Kañulef Martínez, Elisew (2000) *Estado del arte de la escritura en mapudungun*. <http://educacion.ulagos.cl/inves/doc/resumen.doc>

Monson¹, Christian; Levin¹, Lori; Vega¹, Rodolfo; Brown¹, Ralf; Font Llitjos¹, Ariadna; Lavie¹, Alon; Carbonell¹, Jaime; Kañulef², Elisew; Wiska², Rosendo (2004) *Data Collection and Analysis of Mapudungun Morphology for Spelling Correction*. Pittsburgh: 1 Language Technologies Institute, Carnegie Mellon University. Temuko: 2 Instituto de Estudios Indígenas, Universidad de La Frontera.

Navarro, Dióscoro (1992) *Breve Diccionario Español-Mapuche*. Valdivia: Fértil Provincia.

Oliver Gonzalez, Antoni (2004) *Adquisició d'informació lèxica i morfosintàctica a partir de corpus sense anotar: aplicació al rus i al croat*. Tesis Doctoral. Programa de Doctorat: *Ciència Cognitiva i Llenguatge*. Bini: 1999 – 2001. Departament de Lingüística General. Barcelona: Universitat de Barcelona.

Salas, Adalberto (1992) Lingüística mapuche. Guía bibliográfica. *Revista Andina* 10 (2), Cusco. [También disponible en <http://www.uchile.cl/facultades/csociales/lenguages/guia1.htm>]

_ (2006) *El mapuche o araucano. Fonología, gramática y antología de cuentos*. Edición Revisada. Santiago: Centro de estudios públicos.

Smeets, I. (1989). A Mapuche Grammar. Ph.D. Dissertation. University of Leiden.

Zúñiga, Fernando (2006) *Mapudungun El habla mapuche*. Santiago: Centro de estudios públicos.

Anexo 1

Lista de fuentes del corpus

1. Antología del cuento mapuche, en Salas, 2006.
2. Chalintukun, en “Boletín mapuche de noticias. Comité de apoyo al pueblo mapuche”,
<http://www.earth-treaty.com/mp/i7.htm>
3. Chew amuayñ ñi dungun?, en Periódico Adkintuwe,
http://www.nodo50.org/azkintuwe/feb14_1.htm
4. Chufut: Inauguran Radio Comunitaria Mapuche “Petu Mogeleyñ”, en
<http://vulcano.wordpress.com/2008/03/03/chubut-inauguran-radio-comunitaria-mapuche-%e2%80%9cpetu-mogeleyñ-%e2%80%9d/>
5. Claudia Says, en Campaña contra el Sida hace énfasis en abstinencia, pareja exclusiva y buen uso del condón, <http://www.kelwochiloe.cl/?p=211>
6. Comentarios del pueblo araucano (la faz social) por Manuel Mankilef, en
http://biblioteca.serindigena.org/libros_digitales/comentarios/comentarios-Title.html
7. Cuentos y Leyendas Mapuches 1 parte, en <http://www.rie.cl/?k=132252&j=132302>
8. Cuentos y Leyendas Mapuches 2 Parte, en <http://rie.cl/?a=132254>
9. Cuentos y Leyendas Mapuches 3 Parte, en <http://rie.cl/?a=132258>
10. Dunguyu, en Comentarios a la noticia "Mapuches piden intervención de la Iglesia en huelga de hambre" del diario La Nación,
http://www.lanacion.cl/prontus_noticias_v2/site/artic/20071227/pags/20071227223508.html
11. Extracto de verbos del Breve Diccionario Español-Mapuche, de Navarro, 1992.
12. Extracto de verbos del Breve glosario del mapudungun, en en Zúñiga, 2006.
13. Extractos de algunos corpus del Proyecto Avenue de la Universidad Carnegie Mellon,
<http://www.cs.cmu.edu/~avenue/Chile.html>
14. Kalfü pewma mew y Ñi pewma mew ngüman, de Elikura Chiwaylaf, en
<http://es.geocities.com/kuifike/elicura/elicura.html>
15. Kimlafuy feychi relosh mapuche yem, de Bruno Serrano, en
<http://es.geocities.com/kuifike/bserrano/bserrano.html>

16. Libertad a los presos políticos del pueblo Mapuche, en barcelona independent media center, http://barcelona.indymedia.org/front.php3?article_id=9129
17. Lonko Paskwal Koña ñi tukulpadugun - Testimonio de un cacique mapuche
18. Poemas de Graciela Winaw 1, en <http://es.geocities.com/kuifike/graciela/ghuinao.html>
19. Poemas de Graciela Winaw 2, en http://www.poetasdelmundo.com/verinfo_america.asp?id=2516
20. Poemas de Leonel Lyenlaf, en <http://es.geocities.com/kuifike/lienlaf/lienlaf.html>
21. Poemas de Winaw y Mankepillan, en <http://www.letras.s5.com/mapu.htm>
22. Püchüke che ñi ülkantulelñeal, de Juan Paulo Wirimilla, en <http://www.letras.s5.com/jph010805.htm>
23. Textos en mapudungun, en Zúñiga, 2006.
24. Ul trayenco, de Juan Paulo Wirimilla, en <http://es.geocities.com/kuifike/huirmilla/huirmilla.html>
25. Wallmapuwen presentó programa político y social, en Tribuna del Bío Bío, http://www.tribunadelbiobio.cl/one_news.asp?idnews=1778&newseditions=35
26. Wallorüpan, de Juan Paulo Wirimilla, en <http://www.letras.s5.com/jph030105.htm>
27. Windows® XP Mapudungun Penywün kütrün, en <http://www.microsoft.com/downloads/details.aspx?FamilyId=0E09942E-36C8-4CA8-83F9-9AD87586301A&displaylang=es>