**Análisis diacrónico de un corpus en mapudungun**

**Felipe Neira Osses**

Típicamente en los estudios de la lengua mapuche se suele partir de la idea que esta es una lengua relativamente estable diacrónicamente y con pocos cambios importantes, principalmente, en su paradigma verbal (Lenz 1863-1938, VI; Salas 1990, s/p;).

Con esto en mente, nuestra investigación busca detectar y evidenciar cambios en la lengua mapuche utilizando metodologías computacionales, tratando de explicitar los problemas que puede tener la aplicación de dichos mecanismos en un estudio de corpus en la lengua mapuche. De forma más específica, el trabajo tiene por objetivo en rastrear la partícula -em en varios textos en mapudungun, realizando un trabajo diacrónico desde 1606 hasta el 2007. La elección de esta partícula nace de sus características morfológicas y semánticas, pues esta ha experimentado muchos cambios a lo largo del tiempo y resulta particularmente interesante analizar con herramientas computacionales dicho camino. Para poder lograr nuestro cometido es necesario, en primer lugar, formar un corpus de mapudungun con sus respectiva caracterización y limpieza. Luego, necesitamos pesquisar las partículas y analizar sus contextos de uso para poder establecer los significados registrados y si se presentan cambios a nivel diacrónico.

El carácter de la investigación es exploratorio, debido a los problemas ya mencionados de conseguir y trabajar con el corpus de mapudungun. Trataremos de evidenciar y proponer soluciones a las dificultades de realizar un análisis de este tipo en la lengua mapuche y buscaremos aprender de los errores, debido a que en el trabajo buscamos trazar el camino a seguir para realizar una investigación más rigurosa a futuro y complementar los resultados.

**El mapudungun**

El mapudungun es una lengua aislada del sur de los Andes que es hablada por aproximadamente 300 mil personas en territorio chileno y argentino (Smeets 2007). Tipológicamente es definida como una lengua aglutinante, polisintética, incorporante y predominantemente sufijadora (Salas 1992, 68).

Debido a la propuesta de Chamoreau (2008) consideramos que es necesario analizar los cambios en la lengua teniendo en cuenta también las variantes geográficas (cuando es posible) y también las situaciones de contacto de la lengua, puesto que estos factores nos permiten tener una visión más completa de su evolución.

Esta lengua posee varias zonas dialectales que son habladas en territorio chileno y argentino (Salas 1992, 59). A grandes rasgos, podemos reconocer cuatro variantes de la lengua, las que fueron propuestas por Lenz basándose en su distribución geográfica (1895-1897, XXII):

Sin embargo, las diferencias entre ellas son elementalmente superficiales y la ininteligibilidad es prácticamente inexistente (Salas 1992, 62). A pesar de esto, consideraremos también las diferencias sintácticas y léxicas en la codificación de este dominio, pues en algunos casos se pueden evidenciar diferencias estilísticas entre una variante y otra (más detalles en la metodología).

Además, en nuestra investigación utilizaremos la propuesta de periodización de Villena (2017), quien divide la historia de la lengua mapuche en tres períodos marcados por el nivel de contacto entre el mapudungun y el castellano: el período misional (1606-1894) que comienza con la primera gramática de Valdivia (1606) y se caracteriza por el poco contacto de la lengua española con el mapudungun; el período etnográfico (1895-1980), enmarcado luego del proceso de conquista de la Araucanía y caracterizado por los intentos de castellanización de la población del territorio; y finalmente el período institucional (1980-Actualidad), vigente hasta el día de hoy, donde la lengua cuenta con cierto nivel de institucionalidad y reconocimiento estatal debido a la castellanización de la población mapuche.

**3.3-Partícula -em**

En primer lugar, es necesario definir el significado actual de la partícula estudiada, esto para poder tener un punto de partida de las relaciones sintácticas y semánticas del morfema. De esta forma, la gramática básica de la lengua mapuche (2006) define -em dependiendo de si se trata de una persona o un objeto:

‘Afijado a un sustantivo de persona, señala que se trata de alguien que ya no existe, es decir un difunto. Mientras que con un sustantivo inanimado otorga el significado de algo sin vigencia’ Por ejemplo

Chawem ‘difunto padre’

Kuifiem ‘tiempos pasados’

Así, podemos ver que su significado actual está ligado a aquellas cosas que ya no existen y se utiliza frecuentemente junto a sustantivos. Además, cabe destacar que actualmente solo sobrevive el uso de -yem (o bien -em cuando se encuentra junto a una consonante).

**Metodología**

Como fue dicho anteriormente, en nuestro trabajo utilizaremos la periodización propuesta por Villena (2017) como delimitación de los estadios sincrónicos de nuestro estudio, de esta forma analizaremos por separado dichos períodos y luego trataremos de compararlos entre sí buscando diferencias significativas entre estos.

El punto de partida para la realización de este corpus está en los trabajos históricos realizados por Benjamin Molineaux (2020), en ellos el autor prepara un corpus con el objetivo de facilitar los trabajos diacrónicos en la lengua mapuche y nos permite acceder de forma más simplificada a textos complejos.

El corpus que buscamos construir consta de una serie de textos variado, que va desde calepinos y gramáticas en el periodo misional, hasta libros de poesía y prosa en el período institucional. Esta variedad se debe a que se busca abarcar la mayor cantidad de palabras que es posible obtener en la lengua. Posteriormente es necesario borrar manualmente los pies de página y palabras en español que puedan interferir en futuros análisis, dejando únicamente las construcciones hechas en mapudungun.

Una vez preparados los textos, estos se abrirán usando Python y se desplegará una lista de los documentos utilizados por periodo. Cuando estos estén abiertos se le removerá la puntuación, los espacios en blanco y todas las mayúsculas, dejando únicamente las palabras en minúsculas. Luego de esto se preparará una lista de palabras que contenga todas las oraciones del corpus y un string de la totalidad del corpus para usos posteriores.

En segundo lugar, se contabiliza la cantidad de palabras totales, la cantidad de palabras únicas y la relación entre estas. Esto se hace con el objetivo de cuantificar las frecuencias de uso y variación lingüística.

Finalmente, se despliega la cantidad de usos de las partículas estudiadas, esto debido a que nos permite analizar la frecuencia de uso y cómo se contrastan las apariciones de cada forma con otros períodos. También, de forma paralela se despliegan todas las apariciones y las cuatro palabras que la anteceden, así es posible analizar junto a qué tipo de lexemas esta suele ser anexada y contrastarlos también de forma diacrónica.

Este proceso será también realizado con una muestra del período institucional, que nos servirá de contrapunto para el análisis cuantitativo del trabajo. El corpus actual consiste principalmente en una serie de entrevistas realizadas por el profesor Felipe Hasler (2015) y consiste solamente en conversaciones y epew contadas por los hablantes de la zona de Kurako Rangkül.

**Análisis**

Para efectos de esta entrega, utilizaremos como ejemplo gran parte del primer periodo, que se compone de los textos registrados por Valdivia (1606, 1626) y Febrés (1765) y consiste principalmente en textos religiosos traducidos por intérpretes de la época que trabajaban junto a los padres. Además, se encuentra escrito de forma muy españolizada por los autores debido al poco contacto y conocimiento que estos poseían de la lengua.

1. **Ta mn quiñe duam meu va ta piabimn ta Dios**
2. Ta mün kiñe duam mew va ta

DET POS.2PL uno deseo con DEM DET

piafimün ta Dios

decir-FUT-3.P-2PL.A DET Dios

Con todo vuestro coraçon dezilde a Dios (Valdivia 1626)

En el ejemplo anterior podemos ver diferencias entre la escritura de los padres (1) y lo que podría considerarse una escritura actual en el grafemario unificado, propuesto por María Catrileo.

Esto corresponde el primer problema que presenta el análisis de las partículas estudiadas y en general para la aplicación de metodologías computacionales, pues es necesario hacerse responsable de distintas formas de escritura y análisis propuesto por los padres, apareciendo muchas veces algunas palabras juntas, o bien, mezcladas con otras. Sin embargo, en esta versión de nuestro trabajo se busca evitar dicha dificultad y partir del supuesto que las formas encontradas corresponden a la totalidad (o al menos a una parte representativa) de la muestra.

Dicho esto, nuestro corpus contiene 39.900 palabras aprox y cerca de 8000 palabras únicas, lo que nos entrega una relación de 0.2. A pesar de estos resultados preliminares, aun sigue siendo necesario limpiar aún más el corpus de construcciones en español.

Dentro del primer período es posible contabilizar 19 usos de yem, 61 usos de ema y 33 usos de em, siendo este segundo el más utilizado y frecuente. Además, cuando analizamos los resultados de forma más exhaustiva podemos ver que en su mayoría estos aparecen junto a sustantivos y verbos nominalizados, es interesante constatar que estas partículas se encuentran anexadas con mayor frecuencia a figuras religiosas y a miembros de la familia o de la comunidad.

yem ['ga', 'maria', 'ñuque', 'mo', 'christo', 'may', 'piuque', 'ñahue', 'esposa', 'mlehue']

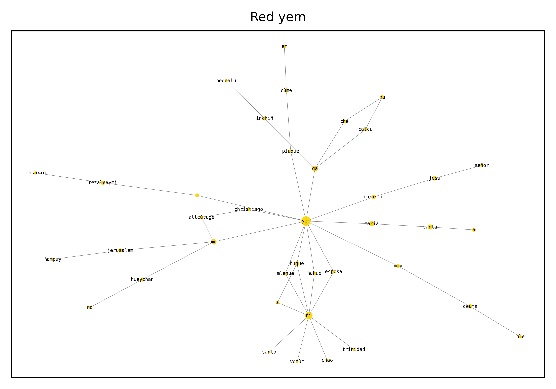
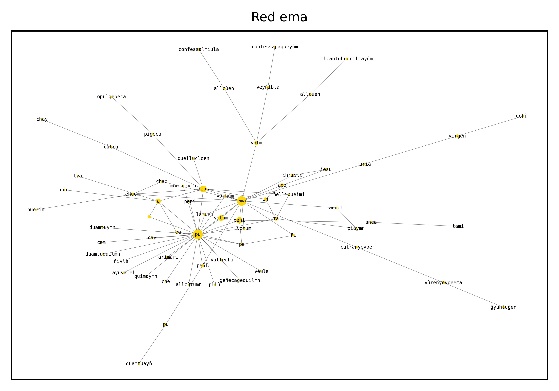
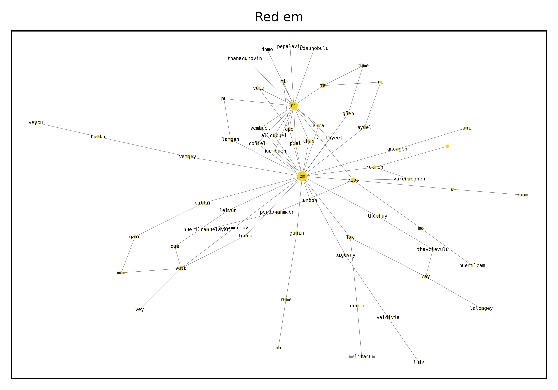
ema ['dios', 'peñi', 'votùm', 'apo', 'votm', 'votum', '', 'lamuen', 'christo', 'cutranyevoe']

em ['dios', 'apo', 'piel', 'chao', 'lay', 'rulmen', 'gùñin', 'huerquen', 'anca', 'coñiel']

Esto se condice con las características misionales del corpus estudiado. Sin embargo, a pesar de esta naturaleza misionera, podemos desprender dos observaciones interesantes para el estudio de la partícula:

1. La alta frecuencia de la utilización de la interjección *nga* antes de la partícula *yem,* lo que nos puede indicar un posible significado interjectivo mirativo de la partícula, el que estaría actualmente en desuso.
2. Condiciéndose con lo anterior podemos ver una alta aparición de la partícula *yem* en forma posterior a la posposición mew (mo en el corpus) que tiende a utilizarse en posición en extremo final.

Si analizamos la red lingüística de la aparición de las partículas, utilizando las 4 palabras anteriores a cada uso podemos ver una clara cercanía a las formas nominales, siendo las centrales en los gráficos, mientras que las formas verbales tienen a ser más exteriores, además la red de yem/em parece tener menos ramificaciones que la forma *ema*, lo que nos puede indicar una menor cantidad de significados, o bien, de usos, siendo ambas teorías que deben ser probadas posteriormente



En cuanto al período actual, este contiene 31147 palabras en total y 5369 palabras únicas, dándonos una relación de 0,17. De estos resultados podemos desprender dos diferencias:

En primer lugar, y debido a la estandarización de la escritura, es más fácil dar certezas acerca de la cantidad exacta de palabras encontradas. Y, en segundo lugar, parece ser que en el primer período existía una mayor cantidad de diferencias entre los usos de palabras, siendo la variante actual más simplificada. Sin embargo, esto puede deberse también a la falta de estandarización del corpus misional, o bien, de la diferencia entre los géneros discursivos analizados.

También, en este corpus podemos encontrar 56 usos de *yem* y 26 usos de *em*, desapareciendo la forma *ema*, pues de esta no aparece absolutamente ningún resultado. Esto nos indica un cambio importante, pues de ser la forma más extendida resulta extraño que perdiera lugar con las formas menos utilizadas. Sin embargo, esto puede deberse a una simplificación fonética que fusionó ambas utilizaciones.

En este corpus, también podemos observar una mayor presencia de figuras familiares que rodean a las partículas, donde ya no aparecen figuras religiosas:

yem ['che', 'chuchu', 'laku', 'ñuke', 'füchakeche', 'kuku', 'palu', 'mama', 'weychafe']

em ['kuku', 'laku', 'chaw', 'ñuke', 'wenüy', 'wentru', 'ñukentu', 'che']

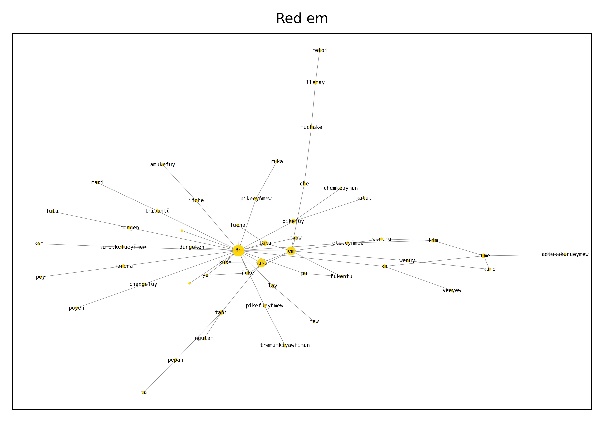
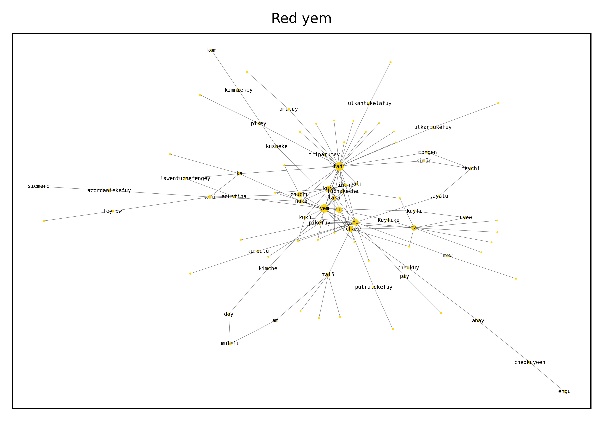
Otro análisis importante viene de los contextos anteriores, pues si analizamos las redes lingüísticas podemos encontrar una aparición muchísimo mayor de verbos utilizados con el sufijo *-fu*, el que indica no vigencia:

1. Amu-ke-fu-y ñi laku yem

Ir-HAB-NV-3S POS abuelo.paterno yem

Mi abuelo solía ir (pero ya no/pero algo ocurrió)

Esto nos puede indicar un cambio en el significado de *yem*, pues parece estar más ligado con la temporalidad y la defunción, pues las construcciones con *-fu* + *yem* son típicamente utilizadas cuando se habla de un finado del cuál se tenía cierto cariño y respeto. Así, la oración en (3) puede significar también, “él solía ir cuando estaba vivo”, pues se sobre entiende el deceso.



De esta forma, podemos encontrar un cambio en el significado analizando el contexto en que esta partícula aparece.

**Conclusiones**

En base a estos resultados podemos concluir, en primera instancia, tres diferencias con el período actual:

En primer lugar, es posible encontrar una mayor aparición de *ema* en el período misional por sobre las otras dos variantes. Sin embargo, en la actualidad esta parece encontrarse en desuso.

En segundo lugar, podemos ver una mayor variedad de significados en las apariciones de la partícula. Además, los significados actualmente registrados no aparecen en la época, ni siquiera en contextos donde típicamente suelen aparecer. Esto nos puede delimitar un camino semántico que va desde el cariño o conmiseración hasta defunción, o bien, tiempo nominal, un path posible y estudiado en otras lenguas (Camargo 2008; Aikhenvald 2012; Mihas 2013).

Finalmente, podemos encontrar usos no anexados a sustantivos tanto en Febrés como en Valdivia. Lo que nos podría indicar una gramaticalización de una interjección, un uso arcaico, o bien, un registro erróneo por parte de los sacerdotes. Estas tres posibilidades nos presentan nuevas ramas de investigación sobre los textos del período.

Campbell, L.. (1998). Historical Linguistics: An introduction. Edinburgo: Edinburgh University Press

Chamoreau, C. (2008). Grammatical borrowing in Purepecha. In Grammatical borrowing in cross-linguistic perspective (pp. 465-480). De Gruyter Mouton.

De Valdivia, L. (1561-1642). Sermón en lengva de chile, de los mysterios de nvestra santa fe catholica, para predicarla a los indios infieles del reyno de Chile, dividido en nveve partes pequeñas acomodadas a su capacidad. Santiago de Chile.

De Valdivia, L. (1606) Arte, y gramatica general de la lengva que corre en todo el Reyno de Chile, con vn Vocabulario, y Confeſſionario. Compueſtos por el Padre Luys de Valdivia, de la Compañía de Jeſus, en la Prouincia del Piru. Ivntamente con la Doctrina Chrisſtiana y Catheciſmo del Concilio de Lima en Eſpañol, y dos traducciones del en la lengua de Chile, que examinaron y aprobaron los dos Reverendiſsimos ſeñores de Chile, cada qual la de ſu Obiſpado, Lima: Francisco del Canto.

Febrés, A. (1765). Arte de la Lengua General del Reyno de Chile. Lima: Universidad Nacional del Sur.

Havestadt, B. (1777). Chilidungu sive res chilenses vel descriptio status tum na-turalis, tum civilis, cum moralis regni populique chilensis, inserta suis locis perfectae ad chilensem linguam manudictioni. Westfalia: Permissu superiorum ac Rmi. & Eximii D. Ordinarii Coloniensis facultate speciali. Typis Aschen-dorfianis.

Molineaux, Benjamin (2020) Prolegomena to a Corpus of Historical Mapduungun, (c) The University of Edinburgh, URL: https://benmolineaux.github.io/. Accessed DD/MM/YEAR.

Salas, A. (1992). El Mapuche o Araucano: Fonología, Gramática y Antología de Cuentos. Madrid: MAPFRE.

Smeets, I. (2007). A Grammar of Mapuche. Berlín: Walter de Gruyter.

Villena, B. (2017). Fuentes para el estudio del mapudungún. Propuesta de periodización. Revista de Lenguas y Literatura Indoamericanas–antes Lengua y Literatura Mapuche–, 19, 141-167.

Zuñiga, F. (2006). El Mapudungun: El Habla Mapuche. Santiago: Centro de Estudios Públicos.

Zwartjes, O., Zimmermann, K., & Schrader-Kniffki, M. (Eds.). (2014). Missionary Linguistics V/Lingüística Misionera V: Translation theories and practices. Selected papers from the Seventh International Conference on Missionary Linguistics, Bremen, 28 February-2 March 2012 (Vol. 122). John Benjamins Publishing Company.