

Phoneme frequency in Mapudungun

Aldo Berríos Castillo (University of Edinburgh)

Berríos (2023) and Berríos & Salamanca (2024) report the phoneme frequency of Mapudungun, a language spoken in the Southern Cone of South America, using a lexicon of mono-morphemic units – considered as types– prepared from a Mapudungun-Spanish Dictionary published in 1916 by Augusta and re-edited in 2017. This is an exceptional lexicographical reference material (Salas 1985, Molineaux 2016). The main dialect used in this work is called Central Mapudungun, and it is spoken in the Araucanía Region, Chile.

Phoneme frequency is a descriptive index that can characterise a particular language by providing a phonological profile (Tambovtsev 2010). Within typological studies, phoneme frequency can be used to establish general trends within languages (Gordon 2016) and make comparisons among world languages (Everett 2018). Similarly, it has been used in several fields, including the study of language development, language teaching, clinical studies of language-related diseases, diachronic studies, etc.

In this quantitative descriptive research of a non-experimental cross-sectional type, a methodology was followed to account for some grammatical phenomena that may skew the phonemic count due to their high frequency, especially in dictionary materials. Thus, inflectional and derivational suffixes, syntactic reduplication, compound words, and nominal incorporation were not included in the final count. For example, in (1), a dictionary entry such as *kimeltuchefe* ‘teacher’ is not included as such. First, the word is segmented, and only the main lexical item is considered for the final count, i.e. the root *kim* ‘know’. If this mono-morphemic unit is already accounted for in another entry (e.g. *kim* ‘adj. wise, knowledgeable’), all those tokens or instances are considered under a single type for counting effects.

- (1) kimeltuchefe
 kim -el -tu -che -fe
 know -CAUS -APL -person -AG
 ‘teacher, lit. that who makes people know’

Other sociolinguistic and linguistic aspects were considered, such as the exclusion of loanwords from Spanish and Quechua, as well as phonemes in interjections, due to their phonotactic anomaly. A list of 2914 mono-morphemic lexical units was gathered, from which the frequency rate of the 28¹ phonemes in Mapudungun was counted, using language software R. The results are shown in Table 1.

Table 1: Phoneme (Phon.) frequency rank (n=13287), including absolute values (Abs.), relative values (Rel.) normalized to 100%, and the cumulative percentage (Acc.).

Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Phon	/ə/	/a/	/k/	/e/	/i/	/o/	/u/	/w/	/l/	/p/	/m/	/ŋ/	/ʎ/	/n/
Abs.	1527	1319	973	870	758	631	618	617	610	548	525	439	433	429
Rel.	11,5	9,9	7,3	6,5	5,7	4,7	4,7	4,6	4,6	4,1	4,0	3,3	3,3	3,2
Acc.	11,5	21,4	28,7	35,3	41,0	45,7	50,4	55,0	59,6	63,8	67,7	71,0	74,3	77,5
Rank	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Phon	/f/	/tʃ/	/j/	/ɟ/	/ɲ/	/tʃ/	/θ/	/t/	/ʃ/	/ɬ/	/ɰ/	/ɲ/	/tʃ/	/s/
Abs.	419	414	379	373	336	303	302	116	110	107	67	49	9	6
Rel.	3,2	3,1	2,9	2,8	2,5	2,3	2,3	0,9	0,8	0,8	0,5	0,4	0,1	0,0
Acc.	80,7	83,8	86,6	89,4	92,0	94,2	96,5	97,4	98,2	99,0	99,5	99,9	100	100

Lexicographic source: Augusta, F. de. (1916, 2017)

¹ Mapudungun phonemes are / p t tʃ tʃ k / m n ɲ ɲ / f θ s ʃ / l ʎ ʎ / ɟ j w ɰ / i e i a o u /

The phoneme frequency shows a typical shaped distribution for this type of phonostatistical study (Martin, 2007:4). Some phonemes account for a high percentage of the whole distribution. The first seven phonemes, for example, correspond to 50% of the total count, while the remaining 19 phonemes have gradually less frequency, with some segments showing relatively imperceptible rates. To report the results, articulatory variables were also considered, such as the frequency of groups of phonemes according to its mode and place of articulation (in vowels, the corresponding articulatory correlates were considered). Similarly, the sonorant/obstruent ratio was also reported (i.e., 3.15:1).

In the total count, it is noted that vowels /i/ and /a/ have the highest count, followed by /k/ and the remaining vowels. 75% of the phonological inventory with the highest frequency correspond to sonorant segments, save two non-coronal stops (/k/ and /p/). The remaining eight coronal obstruents represent much lower frequencies in comparison, with alveolars being in the lowest end. In typological terms, it is noteworthy that /t/ has a low frequency, since it is an unmarked segment in world inventories (Gordon 2016).

The results obtained in this study were compared with another study carried out on a smaller sample of dictionaries published in the 1990s, with a similar methodology (i.e., Márquez 2016). Both studies considered the same amount of phonemes $n = 28$, which coincides with Sadowsky et al (2013). Equivalent results were obtained when comparing the relative frequencies in both studies.

Interestingly, the largest difference between both phonostatistical studies corresponded to the relatively higher frequency of the mid central vowel /i/ in this study, compared to the previous study, in which it occupied the second place. This may be related to sociolinguistic phenomena, given that the materials used in Márquez (2016) and in our study cover different time periods. Thus, it is possible that language change may have occurred between those time periods, and it is expressed in both phonemic frequency distributions. An interesting notion to further explore is that of 'lexical impoverishment' (Choidi & Loncon 1999), in which native lexical items are lost due to contact with a majority language. It could be hypothesised that vocabulary with the /i/ segment has become lost in time, while being replaced with loanwords from Spanish, which usually do not have that segment. Similarly, the productive derivation system of the language allows to coin novel words, which may amplify the effect if items without /i/ are used to create new lexicon. With all, a more in-depth analysis is needed to identify this situation.

Another aspect, among many others, arising from this study is the low frequency rates of dental consonants in the corpus, which follows the relationship identified by Gordon (2016), in which intra-language frequencies correlate with cross-language frequencies. Given that the phonological contrast between dental and alveolar counterparts is typologically marked (Molineaux 2018), it was expected that the four dental phonemes in Mapudungun would have low frequency.

References

- Augusta, F. de. (1916). *Diccionario araucano-español y español-araucano*. Imprenta Universitaria.
- Augusta, F. de. (2017[1916]). *Diccionario mapudungún-español* [...] Edited by B. Villena. Ediciones UCT.
- Berrios, A. (2023). *Frecuencia fonemática del mapudungun*. Master's Thesis. <http://repositorio.udec.cl/handle/11594/10582>
- Berrios, A., & Salamanca, G. (2024). Frecuencia fonemática en mapudungun. *Literatura y Lingüística*, (50).
- Chiodi, F. & Loncon, E. (1999). *Crear nuevas palabras*. UFRO-CONADI.
- Everett, C. (2018). The similar rates of occurrence of consonants across the world [...] *Language Sciences*, 69, 125–135
- Gordon, M. (2016). *Phonological Typology*. Oxford University Press.
- Márquez, C. (2016). *Frecuencia de fonemas y consideraciones teórico-metodológicas en un corpus del mapudungun*. Thesis.
- Martin, A. T. (2007). *The evolving lexicon*. Dissertation.
- Molineaux, B. (2016). Vigencia del Diccionario Araucano de Félix de Augusta [...] *Boletín de Filología*, 51(1), 187–209.
- Molineaux, B. (2018). Pertinacity and change in Mapudungun stress assignment. *IJAL*, 84(4), 513–558.
- Sadowsky, S., Painequeo, J. H., Salamanca, G., & Avelino, H. (2013). Mapudungun. *JIPA*, 43(01), 87–96.
- Salas, A. (1985). Fray Feliz José de Augusta. Su aporte [...] *CUHSO*, 2(2), 197–272.
- Tambovtsev, Y. (2010). Euphony in world languages. *Lingua Posnaniensis*, 52(1), 99–111.