

## FRECUENCIA DE FONEMAS EN LAS NARRACIONES TRADICIONALES EN MALECU

*Frequency of occurrence of phonemes in traditional tales in Maleku*

*Haakon S. Krohn\**

### RESUMEN

Este trabajo constituye un análisis estadístico de la frecuencia de los fonemas en los textos del tipo *marácutúca* 'lo que se ha escuchado' (o 'pláticas tradicionales') en el idioma chibchense malecu hablado en Palenque Margarita. Los resultados muestran que las vocales son más frecuentes que las consonantes. La vocal más común es /a/, seguida por /i/, /e/, /u/ y /o/ (largas o breves). En cuanto a las consonantes, las más usadas son /k/, /t/, /n/ y /m/, mientras que /ʎ/ y /s/ son las menos frecuentes.

**Palabras clave:** Lengua malecu, fonología, frecuencia de fonemas, lingüística de corpus, lingüística cuantitativa.

### ABSTRACT

This work is a statistical analysis of the frequency of occurrence of the phonemes in texts of the type *marácutúca* 'what has been heard' (or 'traditional talks') in the Chibchan language Maleku as spoken in Palenque Margarita. The results show that the vowels are more frequent than the consonants. The most common vowel is (long or short) /a/, followed by /i/, /e/, /u/ and /o/. Regarding the consonants, the most used are /k/, /t/, /n/ and /m/, while /ʎ/ and /s/ are the least frequent ones.

**Key Words:** Maleku language, phonology, frequency of occurrence of phonemes, corpus linguistics, quantitative linguistics.

---

\* Universidad de Costa Rica. Escuela de Filología, Lingüística y Literatura. Costa Rica.  
Correo electrónico: hkrohn@gmail.com  
Recepción: 6/3/16. Aceptación: 11/10/16.

## 1. Introducción

Los análisis estadísticos de la frecuencia de fonemas constituyen parte de la descripción fonológica de una lengua. Los primeros estudios de esta índole se remontan hasta los años 30 del siglo pasado, de ahí que haya numerosos datos disponibles, especialmente de las lenguas más habladas del mundo. Por ejemplo, en el caso del español, se pueden destacar las obras de Zipf y Rogers (1939), Navarro (1946), Quilis y Esgueva (1980), Alameda y Cuetos (1995), y Moreno-Sandoval *et al.* (2008), entre muchos otros. En cambio, por lo que respecta a las lenguas de la familia chibchense, trabajos de este tipo son prácticamente inexistentes; la única excepción es uno de Guillermo González sobre el cabécar que aún no ha sido publicado.

Esta situación es el acicate para el presente estudio, el cual consiste en un análisis cuantitativo de la frecuencia con la que se emplea cada fonema en los textos orales pertenecientes al subgénero *marácutica* ‘lo que se ha escuchado’, también referidos como ‘pláticas tradicionales’ o ‘narraciones tradicionales’, en idioma malecu,<sup>1</sup> específicamente de la variedad diatópica del Palenque Margarita. El corpus analizado consiste en un total de más de 166 mil fonemas. Los resultados proporcionan información acerca de la frecuencia textual de los fonemas, es decir, su distribución ponderada por la frecuencia del uso de las palabras en el corpus. De esta manera, se refleja el uso real de los fonemas; no en la lengua malecu en su totalidad, sino en los textos del subgénero analizado producidos por hablantes del Palenque Margarita de una edad similar a la de los narradores que proporcionaron el material. El corpus puede considerarse una muestra aleatoria del universo estadístico consistente en la totalidad de textos del mismo tipo, por lo que las proporciones calculadas también constituyen

estimaciones de la frecuencia de cada fonema en las narraciones de este tipo que no se hayan incorporado en el corpus. Como indicación de la fiabilidad de las estimaciones, incluimos los errores estándar y los márgenes de error de todos los valores porcentuales presentados.

Nuestra investigación tiene, principalmente, fines teóricos. Los datos obtenidos servirán, en primer lugar, para enriquecer el conocimiento sobre la importancia fonémica de cada fonema de esta lengua. En segundo lugar, constituirá una contribución por parte de la familia chibchense a los estudios fonológicos interlingüísticos. Además, servirá como ejemplo del procedimiento en un análisis de corpus de una lengua de pocos hablantes. Áreas más prácticas en las que esta clase de información suele ser útil, como en la síntesis y el reconocimiento automáticos del habla, o en las aplicaciones clínicas, parecen menos relevantes, dado que se trata de una lengua en estado de declinación (Sánchez, 2009) con un bajo número de hablantes.

## 2. Inventario fonológico del malecu

En este apartado se describe la base cualitativa para nuestro análisis cuantitativo, a saber, el inventario fonológico del malecu. Este sistema consiste en quince fonemas consonánticos, y cinco o diez vocálicos (dependiendo de la interpretación del estatus del alargamiento vocálico). La descripción parte de Constenla (1983; 1998: 7-22), pero hace uso de una terminología más acorde con la empleada en el Alfabeto Fonético Internacional (AFI).

### 2.1. Fonemas consonánticos

El inventario consonántico patrimonial del malecu se presenta en la tabla 1 (Constenla, 1998: 8).

TABLA 1.

## Fonemas consonánticos del sistema patrimonial malecu

	Bilabial		Coronal		Velar	
	sor.	son.	sor.	son.	sor.	son.
Oclusivas	p		t		k	
Africadas			tʃ	ɟʒ		
Fricativas	ɸ		s		x	
Fricativas lat.			ʃ			
Aprox. lat.				l		
Vibr. simples				r		
Vibr. múltipl.				r		
Nasales		m		n		ŋ

Los fonemas africados, /tʃ/ y /ɟʒ/, son los únicos de este sistema que se distinguen solamente por el rasgo [sonoro] y, además, los únicos articulados parcialmente en la región palatal. El sordo se realiza como prepalatal (Constenla, 1983: 9), mientras que el sonoro es clasificado por Constenla (1998: 11) como mediopalatal y “se realiza facultativamente como africana [tʃ] o espirante débilmente fricativa [ɟʒ]”.<sup>2</sup> Dada la oposición de este fonema con /tʃ/, nosotros hemos optado por el símbolo /ɟʒ/. Cabe subrayar la situación del fonema /tʃ/, el cual es sujeto a un cambio fonemático que, según Constenla (1983: 3), se encuentra en proceso desde principios del siglo XX y que consiste en un reemplazo de dicho fonema por /x/ en ciertos entornos. La fusión se da con mayor frecuencia entre los hablantes de Tonjibe, que en los otros dos palenques o aldeas donde habitan los malecus, Margarita y El Sol, pero debe considerarse un fenómeno de variación libre (Constenla, 1998: 21). Tal hecho origina cierta inseguridad en la interpretación de los resultados del presente análisis, ya que no se

puede determinar si la distribución de estos dos fonemas en el corpus es representativa para todos los hablantes.

Debido a la considerable cantidad de préstamos léxicos del español, algunos hablantes del malecu emplean un sistema ampliado con fonemas de dicha lengua. De acuerdo con Constenla (1998: 19), el sistema máximamente innovador incluye las tres oclusivas sonoras /b/ (bilabial), /d/ (alveolar) y /g/ (velar), así como la nasal palatal /ɲ/. No obstante, estos fonemas no aparecen en nuestro corpus.

## 2.2. Fonemas vocálicos

Constenla (1998: 8) alega que el malecu posee cinco fonemas vocálicos, los cuales se presentan en la tabla 2.<sup>3</sup>

TABLA 2.

## Fonemas vocálicos del malecu de acuerdo con Constenla (1998).

	Anterior	Central	Posterior
Alta	i		u
Media	e		o
Baja		a	

Como se puede apreciar, se trata de un sistema simple de tres alturas. Las dos vocales posteriores son redondeadas, mientras que las otras tres son no redondeadas.

Además, cada vocal puede presentarse como breve o alargada. Respecto a esta distinción, Constenla (1998: 12) afirma lo siguiente:

Las secuencias de vocal anterior o posterior más /:/ se manifiestan como vocales largas diptongantes en cuya producción se pasa de un inicio intermedio o relativamente abierto a un final muy cerrado [...] La secuencia de /a/ más /:/ se diferencia del fonema /a/ sólo por lo que atañe a su mayor duración.

Es decir, dicho lingüista interpreta las vocales alargadas como una combinación de uno de los cinco fonemas vocálicos y el fonema

suprasegmental de alargamiento /:/. De acuerdo con Fox (2000: 77), esta puede considerarse una aproximación “prosódica”:

Implicit within some of these approaches is the view that length is not only to be seen *in relation to* its content, but that some, perhaps all, of the durational properties of individual segments should be *derived from* this context. Length, in other words, is primarily a property of more inclusive, or ‘higher’, units of speech, above all of the syllable, and is only secondarily to be attributed to individual sounds. Such an approach can be called ‘prosodic’, in the sense that length is regarded as distinct from the segments themselves.

Dada la existencia de pares mínimos que difieren únicamente en la cantidad vocálica (Constenla, 1998: 10), las vocales largas también podrían interpretarse como fonemas distintos de las breves, lo cual resultaría en un total de diez fonemas vocálicos en malecu. En este trabajo tomaremos en cuenta las dos posibilidades en la presentación de los resultados, pero nos apegaremos en mayor grado al análisis de Constenla.

### 3. Metodología

#### 3.1. Delimitación y compilación del corpus

Como bien es sabido, lo que es considerado una sola lengua puede presentar grandes variaciones en las dimensiones diafásica, diastrática, diatópica y diacrónica. En un estudio de esta naturaleza es imprescindible definir claramente la variedad lingüística que se va a analizar, para poder compilar un corpus representativo para todos los textos que cumplan con tal definición. A este respecto, McEnery y Wilson (2001: 75) señalan que el corpus es una muestra finita máximamente representativa, y que “[t]he corpus thus stands in contrast to other empirical data sets which have not been sampled to be maximally representative and from which broader conclusions cannot therefore be extrapolated.”

En primer lugar, decidimos delimitar el análisis a los textos orales tradicionales

clasificados como *marácuníca* ‘lo que se ha escuchado’, ya que la gran mayoría de los textos transcritos publicados en malecu (todos por parte de Adolfo Constenla Umaña y sus coautores) pertenecen a este subgénero. De otro género no sería posible armar un corpus lo suficientemente grande. Según indican Constenla, Castro y Blanco (1993: 10), estos textos forman parte del género literario oral *mausírrajáca maráma* ‘pláticas’, que también engloba las narraciones de carácter no tradicional. Por tal razón, el tipo de hecho del habla que constituye nuestro corpus también puede denominarse ‘pláticas tradicionales’ o ‘narraciones tradicionales’.

Como se explica en Constenla, Castro y Blanco (1993: 19), las pláticas tradicionales tienen una función fundamentalmente didáctica e “incluyen de manera muy directa toda clase de informaciones acerca de las creencias, los valores y la historia de la tribu [...] No obstante lo antes dicho, las pláticas tradicionales cumplían también una función importante como entretenimiento.”

En cuanto a los rasgos formales del subgénero abarcado, el más relevante para este análisis es la gran frecuencia con la que se emplea el verbo *quí* /ki:/ ‘decir’. También son notables el paralelismo (que consiste en que un mismo pensamiento se repite numerosas veces) y el uso de ideófonos.

El restringir el corpus a las pláticas tradicionales, todas recogidas durante las décadas recientes de hablantes de edades similares, debe asegurar la uniformidad diafásica, diastrática y diacrónica. En lo que concierne a la dimensión diatópica, las variaciones en este idioma son pequeñas (Constenla, 1983: 3), como consecuencia de su escaso número de hablantes y su limitada extensión geográfica. Sin embargo, la tendencia mencionada en el apartado 2.1. que consiste en el reemplazo del fonema /h/ por /x/ sí parece depender de esta variable, ya que es descrita como más fuerte en Tonjibe (*ibid.*). De todos modos, como se explicará en breve, el corpus analizado proviene principalmente de solo dos hablantes, ambos oriundos del Palenque Margarita, por lo que la uniformidad diatópica también estaría asegurada, aunque, desde luego,

no se puede afirmar que las frecuencias de los fonemas /h/ y /x/ (ni de los demás fonemas) calculadas en este estudio sean las mismas en el habla de todos los narradores de pláticas tradicionales del mismo palenque.

No obstante, sí nos vimos obligados a definir una delimitación más, dado que un preanálisis reveló que las dos narraciones tradicionales en Betancourt y Constenla (1981) son las únicas de todas las publicadas que incluyen un fonema no patrimonial: /d/. Esto se debe al tema tratado en dichos textos, el asesinato a un sacerdote, el cual requiere el empleo de los préstamos *pádre* ‘padre’ y *saserdóte* ‘sacerdote’. A juzgar por lo expuesto en Constenla, Castro y Blanco (1993: 21-22), esta temática es poco frecuente en los textos del tipo *marácunúca*, por lo que el impacto estadístico de las nueve instancias del fonema /d/ que se presentan aquí, probablemente, es mucho mayor de lo que sería el caso si se analizara la totalidad de pláticas tradicionales en malecu. Por lo tanto, tomamos la decisión de excluir ambas narraciones del corpus, de modo que el análisis abarca únicamente los fonemas del sistema patrimonial. En consecuencia, los resultados no son representativos para los textos que incluyan otros fonemas.

Otro hecho problemático es que la diversidad de hablas de las que procede el corpus es muy baja; las narraciones fueron recogidas (por parte de Adolfo Constenla Umaña) principalmente de dos narradores: Eustaquio

Castro y Antonio Blanco, ambos provenientes del Palenque Margarita y nacidos en el 1942. De ellos, Castro es la fuente de la gran mayoría de los textos. Evidentemente, este inconveniente reduce la representatividad de los números. Sin embargo, se trata de una debilidad inevitable en un análisis de corpus de una lengua de tan pocos hablantes, sobre todo si se imponen otras restricciones al corpus, como en nuestra investigación. En efecto, pocos malecuhablantes actuales dominan el subgénero al que se limita el análisis, por lo que, si se toman en cuenta todas las restricciones descritas en el presente apartado, el corpus podría considerarse lo suficientemente representativo para la variedad malecu que corresponde a la intersección de todos los dominios delimitados, aunque el hecho de que Castro ha aportado una parte más grande del corpus que las otras fuentes sí constituye un sesgo que podría ser significativo.

Para cerrar este apartado, los textos incluidos en el corpus se presentan en la tabla 3. Aquí se aprecia que se analizó un total de 65 textos provenientes de nueve publicaciones distintas, y que la cantidad total de fonemas en el corpus es 166 691. Por las razones señaladas arriba, los únicos textos malecus publicados que no se analizaron son los de Betancourt y Constenla (1981) y los de Constenla (1996); los últimos porque no son *marácunúca*, sino que pertenecen a la clase de comunicación lingüística *maulácu maráma* ‘las cosas que se dicen cantadas o en estilo recitativo’.

TABLA 3.  
Textos incluidos en el corpus

Publicación	Cantidad de textos	Cantidad de fonemas
Constenla (1983). “Descripción del sistema fonemático del guatuso”.	1	449
Constenla (1986). “La función de una alternancia gramatical guatusa en el discurso narrativo tradicional”.	1	594
Constenla (1991). “Tres textos guatusos del ciclo narrativo de las uniones con los animales”.	3	5586
Constenla (1992). “Hagiografía y antihagiografía en la tradición oral guatusa”.	2	8874
Constenla, Castro y Blanco (1993). <i>Lacá Majifijica. La transformación de la tierra.</i>	11	31651
Constenla (1999). “El respeto a la vida animal en la tradición oral guatusa”.	2	4569
Constenla (2003). “Dos textos guatusos sobre los profetas del cataclismo”.	2	17568
Constenla y Castro (2011). <i>Pláticas sobre felinos.</i>	25	50274
Constenla y Castro (2014). <i>Muérrojá mausírrajáca. Pláticas sobre ogros.</i>	18	47126
Total	65	166691

### 3.2. Adecuación del corpus

En primera instancia, todos los textos del corpus fueron convertidos en formato digital de texto. Este paso requirió de mucho tiempo, sobre todo porque los textos de los libros de Constenla, Castro y Blanco (1993) y Constenla y Castro (2011, 2014) tuvieron que ser escaneados previo a la conversión. Las demás obras están disponibles en archivos PDF en línea, pero aun así, todos los textos debieron ser revisados manualmente para eliminar los errores que surgieron al pasarlos de un formato a otro.

Seguidamente, el corpus compilado fue convertido en una representación fonológica. Como la ortografía práctica del malecu empleada en todos los textos es fonemática, esto se pudo llevar a cabo automáticamente. También se eliminaron todos los signos de puntuación, para que cada caracter presente en el corpus representara un fonema. Las correspondencias entre los grafemas y los fonemas (del sistema patrimonial) se indican en la tabla 4. Además, cabe aclarar que las vocales largas son marcadas por medio de un acento agudo ortográfico: < í, é, á, ó, ú >. Las breves, por su lado, no llevan acento gráfico.

TABLA 4.

Correspondencias entre las letras de la ortografía práctica del malecu y los fonemas.

Fonemas consonánt.	Letras	Fonemas vocálicos	Letras
/m/	m	/i/	i
/n/	n	/e/	e
/ŋ/	nh	/a/	a
/p/	p	/o/	o
/t/	t	/u/	u
/k/	c, qu		
/tʃ/	ch		
/j/	y		
/f/	f		
/s/	s		
/x/	j		
/h/	lh		
/l/	l		
/r/	r		
/r/	rr		

Un inconveniente hallado durante la adecuación del corpus es el hecho de que algunos textos incluyen todas las vacilaciones producidas por el narrador a la hora de ser grabado, mientras que en otros, están eliminadas de la transcripción.<sup>4</sup> Específicamente, las vacilaciones consisten en palabras incompletas, o bien, en repeticiones del principio de palabras que eventualmente se pronuncian completas. En el último caso, se trata, en la mayoría de las ocasiones, de fonemas vocálicos reiterados numerosas veces, como en el ejemplo 1 (Constenla y Castro, 2014: 80, 148, 213):

(1) Níni nij i... i... iquíca ó... ó  
muérra tócuftátócu fá maráma.  
esto=LIG este=(3)-sobre 3-...  
3-... 3-decir-INF que... que ogro AUM-RED PL  
‘Sucedió que esto decía sobre  
los que son ogros grandes.’

Esto constituía un problema para el corpus, ya que si la vocal /i/ de la secuencia *i... i... iquíca* se contara cuatro veces, y la *ó* repetida se tomara en cuenta como dos instancias del fonema /o:/, la frecuencia de fonemas vocálicos resultaría considerablemente mayor en los

textos que incluyen este tipo de vacilación en la transcripción que en los demás. Por lo tanto, para reforzar la uniformidad del corpus, decidimos considerar el sonido repetido como una sola instancia del fonema, por lo que la transcripción fonológica del ejemplo 1 es la que se da en 2:

(2) /ni:ni nix iki:ka o: mue:ra  
to:kuɸa:to:kuɸa: mara:ma/

En cambio, cuando la transcripción revela que la pronunciación de una palabra fue interrumpida sin llegar a producirse de manera completa, la palabra incompleta no fue eliminada del corpus, ya que la parte pronunciada puede consistir en casi cualquier secuencia de fonemas.

En cuanto a los ideófonos, son abundantes en algunos de los textos y sí se tomaron en cuenta en el análisis. Sin embargo, tuvieron que ser modificados cuando incluían una repetición de una misma letra como representación de una vocal muy alargada, por ejemplo en la imitación del sonido del diablo transcrito en Constenla y Castro (2011: 68) como *lhéec*. Aquí, los dos últimos grafemas <e> no representan dos instancias del fonema vocálico breve /e/, sino un alargamiento adicional de /e:/. Por lo tanto, los grafemas repetidos en estos casos fueron eliminados, de modo que la transcripción fonológica del ejemplo citado es /ɛ:k/.

Por último, los textos incluyen algunas palabras provenientes del español. La mayoría de estas cumplen, claramente, una función en el discurso, por ejemplo *porque*, *por*, *pero* y *este*, por lo que las consideramos préstamos y sí las tomamos en cuenta en el análisis. No obstante,

en un caso se trata de una alternancia de código no perteneciente a la narración: *Eh, perdón* (Constenla, 2003: 110). Estas dos palabras fueron eliminadas del corpus.

#### 4. Resultados

Los resultados del análisis se presentarán en este apartado. Primeramente, la frecuencia relativa entre los fonemas consonánticos y vocálicos se muestra en la tabla 5. Aquí se observa que, a pesar de que el inventario consonántico sea más grande, las vocales son más comunes que las consonantes en el corpus, con una frecuencia de 52,98 %.

Al utilizar valores porcentuales para expresar la frecuencia de una forma lingüística, también debe haber una indicación de la fiabilidad de ese número. Por tanto, también hemos incluido el error estándar y el margen de error con un nivel de confianza (NC) del 99 % de todos los porcentajes presentados en este artículo. De esta manera, los datos constituyen una estimación de la frecuencia de los fonemas del sistema patrimonial en todas las pláticas tradicionales en malecu, no solo en las que se incluyeron en el corpus. El error estándar (EE) es calculado mediante la fórmula  $\sqrt{pq}$ , donde  $p$  es la proporción del fonema o del conjunto de fonemas en cuestión,  $q$  es la proporción de todos los demás fonemas y  $N$  es el número total de fonemas (en este caso, 166 691). El margen de error con un nivel de confianza del 99 % se obtiene multiplicando el error estándar por el valor crítico (valor  $z^*$ ) 2,58.

TABLA 5.

Frecuencia de los fonemas vocálicos y consonánticos

Fonemas	Número de apariciones	Frecuencia (%)	Error estándar	Margen de error (NC = 99 %)
Vocálicos	88311	52,98	0,12	±0,32
Consonánticos	78380	47,02	0,12	±0,32
Total	166691	100,00	-	-



La diferencia entre las dos proporciones es de  $52,98\% - 47,02\% = 5,96\%$ . Por consiguiente, el error estándar de esta diferencia es  $EE_{DIF} = \frac{EE_1 + EE_2}{2}$ , donde  $EE_1$  es el error estándar de la primera proporción, y  $EE_2$ , el de la otra. Esto nos da 0,17. Si se divide la brecha de  $5,96\%$  entre el error estándar de la diferencia, es decir,  $\frac{5,96}{0,17}$ , queda claro que la diferencia corresponde a 35,06 veces el  $EE_{DIF}$ . Por lo tanto, podemos concluir que las vocales, en efecto, son más frecuentes que las consonantes.

Ahora se presentarán los datos de las consonantes con mayor detalle; estos se aprecian en la tabla 6. Los fonemas están ordenados según su frecuencia de aparición, y el porcentaje indicado alude a su proporción entre todas las consonantes (78 380 en total). Al igual que en la tabla anterior, se incluyen el error estándar y el margen de error con un nivel de confianza del 99 %.

TABLA 6.

## Frecuencia de los fonemas consonánticos

Fonema	Número de apariciones	Frecuencia (% de cons.)	Error estándar	Margen de error (NC = 99 %)
/k/	15384	19,63	0,14	$\pm 0,37$
/t/	9769	12,46	0,12	$\pm 0,30$
/n/	9392	11,98	0,12	$\pm 0,30$
/m/	9011	11,50	0,11	$\pm 0,29$
/r/	6269	8,00	0,10	$\pm 0,25$
/ŋ/	5722	7,30	0,09	$\pm 0,24$
/r/	4515	5,76	0,08	$\pm 0,21$
/p/	4155	5,30	0,08	$\pm 0,21$
/x/	3587	4,58	0,07	$\pm 0,19$
/ϕ/	2791	3,56	0,07	$\pm 0,17$
/ʃ/	1993	2,54	0,06	$\pm 0,15$
/dʒ/	1881	2,40	0,05	$\pm 0,14$
/l/	1769	2,26	0,05	$\pm 0,14$
/s/	1166	1,49	0,04	$\pm 0,11$
/ʎ/	976	1,25	0,04	$\pm 0,10$
Total	78380	100,00	-	-

Aquí se observa que /k/ es el fonema consonántico claramente más usado. Este es seguido por /t/, /n/ y /m/, los cuales presentan frecuencias muy similares. El menos usado es /h/, mientras el segundo menos frecuente es /s/. En general, los fonemas fricativos, africados y laterales son mucho menos empleados que los

demás, y los oclusivos, nasales y vibrantes son los modos de articulación más frecuentes. Dicha distribución se aprecia con mayor claridad en la tabla 7. La inclusión del fonema fricativo lateral /h/ en el rubro de fricativos elimina la inseguridad causada por la oscilación entre /h/ y /x/.

TABLA 7.

Frecuencia de los fonemas consonánticos según su modo de articulación.

Modo de articulación	Número de apariciones	Frecuencia (% de cons.)	Error estándar	Margen de error (NC = 99 %)
Oclusivo	29308	37,39	0,17	±0,45
Nasal	24125	30,78	0,16	±0,43
Vibrante	10784	13,76	0,12	±0,32
Fricativo	8520	10,87	0,11	±0,29
Africado	3874	4,94	0,08	±0,20
Aprox. lat.	1769	2,26	0,05	±0,14
Total	78380	100,00	-	-

En la tabla 8, las consonantes están agrupadas según su punto de articulación. Aquí se evidencia que la mayoría de las articulaciones consonánticas tienen lugar en la zona coronal;

cerca de un 50 %. El segundo punto de articulación más común es el velar, mientras que el bilabial es el menos frecuente.

TABLA 8.

Frecuencia de los fonemas consonánticos según su punto de articulación.

Punto de articulación	Número de apariciones	Frecuencia (% de cons.)	Error estándar	Margen de error (NC = 99 %)
Coronal	37730	48,14	0,18	±0,46
Velar	24693	31,50	0,17	±0,43
Bilabial	15957	20,36	0,14	±0,37
Total	78380	100,00	-	-

En lo que concierne a las vocales, los datos serán presentados de dos maneras distintas, con el fin de abarcar las dos percepciones contrastantes del estatus del alargamiento. Primero, la tabla 9 muestra las proporciones de los cinco fonemas vocálicos si el alargamiento es considerado un rasgo suprasegmental. Esta tabla incluye las mismas columnas que las anteriores, con la adición de tres más en el extremo derecho. La

denominada “Frecuencia de alarg.” indica la frecuencia con la que cada fonema se asocia con el rasgo de alargamiento. De esta cifra también se presenta el error estándar y el margen de error, en este caso con un nivel de confianza de 95 %, dado que está basada en una menor cantidad de datos (el número de apariciones del fonema en cuestión).

TABLA 9.

Frecuencia de los fonemas vocálicos si el alargamiento se considera suprasegmental.

Fonema	Número de apariciones	Frecuencia (% de voc.)	Error estándar	Margen de error (NC = 99 %)	Frecuencia de alarg. (%)	Error estándar	Margen de error (NC = 95 %)
/a/	33950	38,44	0,14	±0,36	41,08	0,27	±0,52
/i/	25084	28,40	0,12	±0,32	45,80	0,31	±0,62
/e/	11982	13,57	0,12	±0,31	44,84	0,45	±0,89
/u/	9435	10,68	0,11	±0,29	28,10	0,46	±0,91
/o/	7860	8,90	0,09	±0,23	55,62	0,56	±1,10
Total	88311	100,00	-	-	<b>42,84</b>	<b>0,17</b>	<b>±0,33</b>

A partir de estos números se puede concluir que el fonema /a/ es el más común, seguido por los dos anteriores /i/, /e/ y, por último, los dos posteriores /u/ y /o/. Además, la vocal que se combina con el suprasegmental /:/ con mayor frecuencia es /o/, mientras la que presenta alargamiento en la menor parte

de los casos es /u/. Sin embargo, la frecuencia de alargamiento es relativamente parecida para todas las vocales, salvo /u/.

Ahora bien, en la tabla 10 se exponen las frecuencias de los diez fonemas vocálicos con la que se debe contar si la cantidad se considera parte del segmento.

TABLA 10.

Frecuencia de los fonemas vocálicos si el alargamiento se considera parte del segmento.

Fonema	Número de apariciones	Frecuencia (% de voc.)	Error estándar	Margen de error (NC = 99 %)
/a/	20002	22,65	0,14	±0,36
/a:/	13948	15,79	0,12	±0,32
/i/	13596	15,40	0,12	±0,31
/i:/	11488	13,01	0,11	±0,29
/u/	6784	7,68	0,09	±0,23
/e/	6609	7,48	0,09	±0,23
/e:/	5373	6,08	0,08	±0,21
/o/	4372	4,95	0,07	±0,19
/o/	3488	3,95	0,07	±0,17
/u:/	2651	3,00	0,06	±0,15
Total	88311	100,00	-	-

De acuerdo con estos datos, /a/ es el fonema claramente más frecuente, y /u:/ es el menos recurrente. También se aprecia que —con la excepción de /u/ y /u:/— dentro de cada par de vocales breve y alargada, los dos integrantes presentan frecuencias muy similares. Esto también es indicado por los porcentajes en la columna “Frecuencia de alargamiento” en la tabla 9, los cuales no presentan grandes variaciones entre una vocal y otra. Tal hecho podría considerarse una indicación de que el alargamiento es asignado a las sílabas independientemente de la vocal que constituya el núcleo y, por lo tanto, debe clasificarse como un rasgo suprasegmental separado de los fonemas vocálicos.

Por último, se presentarán los números de todos los fonemas en una sola tabla, la 11, con el fin de determinar el orden de preferencia de los fonemas, desde el más común hasta el menos usado. Aquí presuponemos el sistema de cinco fonemas vocálicos. En esta tabla se incluyen dos columnas nuevas: primeramente, “EE<sub>DIF</sub>” es el error estándar de la diferencia entre la proporción de determinado fonema y el anterior.

Por ejemplo, el valor de 0,13 en esta columna para el fonema /i/ corresponde al error estándar de la diferencia entre la proporción de este fonema y la de /a/. A su vez, el EE<sub>DIF</sub> en la fila de /k/ alude al error estándar de la diferencia entre las distribuciones de /k/ e /i/, **y así sucesivamente**.

La **última** columna de la tabla, “Dif/EE<sub>DIF</sub>”, muestra la brecha entre las proporciones de los dos fonemas comparados (en el caso de /i/ y /a/: 20,37 % – 15,05 % = 5,32 %) dividida entre el error estándar de esta diferencia. El **número** resultante es una medida adicional<sup>5</sup> de cuán confiable es el orden de los fonemas en la tabla: entre más alto sea el número, mayor es la probabilidad de que el orden respectivo verdadero entre los dos fonemas en cuestión sea el mismo que presentan en la tabla. Según Reed (1949: 242-245), una desviación de 1,00 error estándar de la diferencia ocurre en 31,74 % de las proporciones en una distribución normal, mientras que una de 3,00 se da en solamente 0,27 % de los casos. Cuando el valor de “Dif/EE<sub>DIF</sub>” es 2,58 (equivalente al valor crítico utilizado para calcular el margen de error) o más,

la probabilidad de que el orden respectivo de ambos fonemas en la tabla sea pura casualidad es de menos de 1,00 %; es decir, el orden presenta un nivel de confianza del 99 % o más. Si se exige este nivel de confianza, se debe concluir que las diferencias entre /t/ y /u/, /u/ y /n/, /ʃ/ y /dʒ/, y /dʒ/ y /l/ pueden ser atribuidas a la casualidad, ya que presentan valores menores de 2,58. Especialmente la brecha entre las proporciones de /u/ y /n/ es irrelevante, pues el “Dif/EE<sub>DIF</sub>”

de 0,32 indica que la probabilidad de que se trate de una distinción aleatoria es de 74,9 % (Reed, 1949: 244).

Además de que es imprescindible tomar en cuenta los márgenes de error, es necesario recordar el hecho de que de los fonemas /x/ y /ʎ/ se encuentran en proceso de fusión, por lo que estas dos proporciones deben tomarse con aún mayores reservas, tanto aquí como en las otras tablas que incluyen sus frecuencias respectivas.

TABLA 11.

Frecuencia de todos los fonemas.

Fonema	Número de apariciones	Frecuencia (% de todos)	Error estándar	Margen de error (NC = 99 %)	EE <sub>DIF</sub>	Dif. / EE <sub>DIF</sub>
/a/	33950	20,37	0,10	±0,25	-	-
/i/	25084	15,05	0,09	±0,23	0,13	40,32
/k/	15384	9,23	0,07	±0,18	0,11	51,65
/e/	11982	7,19	0,06	±0,16	0,10	21,48
/t/	9769	5,86	0,06	±0,15	0,09	15,53
/u/	9435	5,66	0,06	±0,15	0,08	2,48
/n/	9392	5,63	0,06	±0,15	0,08	0,32
/m/	9011	5,41	0,06	±0,14	0,08	2,89
/o/	7860	4,72	0,05	±0,13	0,08	9,10
/ʎ/	6269	3,76	0,05	±0,12	0,07	13,68
/ŋ/	5722	3,43	0,04	±0,12	0,06	5,09
/r/	4515	2,71	0,04	±0,10	0,06	12,12
/p/	4155	2,49	0,04	±0,10	0,06	3,92
/x/	3587	2,15	0,04	±0,09	0,05	6,53
/ϕ/	2791	1,67	0,03	±0,08	0,05	10,07
/ʃ/	1993	1,20	0,03	±0,07	0,04	11,62
/dʒ/	1881	1,13	0,03	±0,07	0,04	1,81
/l/	1769	1,06	0,03	±0,06	0,04	1,86
/s/	1166	0,70	0,02	±0,05	0,03	11,18
/ʎ/	976	0,59	0,02	±0,05	0,03	4,12
Total	166691	100,00	-	-	-	-

## 5. Conclusiones

En este artículo se ha presentado una serie de datos que constituyen una estimación estadística de la frecuencia de los fonemas patrimoniales en las pláticas tradicionales en malecu del Palenque Margarita hablado por personas nacidas a mediados del siglo XX. A este respecto, cabe resaltar lo señalado por Heylen (2005: 261; citado en Vinogradov, 2016: 137): que los resultados de un análisis de este tipo son “strictly speaking only valid for observations that instantiate a similar type of language use as the one that was represented in the corpus”. No obstante, aunque los números no son representativos para otras variedades del malecu, al menos constituyen una primera indicación de la frecuencia de fonemas en esta lengua.

A estas alturas, consideramos pertinente reiterar que se trata de una distribución textual, ponderada por la frecuencia del uso de cada palabra. Como apunta Herdan (1966: 32), esta

no se debe confundir con la distribución léxica, la cual se obtiene analizando las entradas de un diccionario y que, por consiguiente, es de menor interés lingüístico. Dicho de otra manera, algunos fonemas son más frecuentes que otros porque forman parte de palabras más usadas. Con motivo de ilustrar la gran variación entre las frecuencias de las palabras y cómo este hecho influye fuertemente en las proporciones de los distintos fonemas, se muestran en la tabla 12 las 15 palabras más empleadas en el corpus. No se indican las frecuencias porcentuales de las palabras, dado que, en malecu, “la palabra, si bien puede resultar de alguna utilidad en la descripción morfosintáctica, no tiene ninguna pertinencia desde el punto de vista fonológico” (Constenla, 1983: 16). Por esta razón, tal conteo no es pertinente en un trabajo de este carácter.<sup>6</sup> De todos modos, resulta claro que una de las causas por las que /a/, /i/ y /k/ son los fonemas más usados es su aparición en varias de las palabras más frecuentes.

TABLA 12.

Palabras más frecuentes en el corpus.

Forma ortográfica	Forma fonológica	Significado literal <sup>7</sup>	Número de apariciones
tacá	/taka:/	y= IL	2055
iquí	/iki:/	3-decir	1703
maráma	/mara.ma/	PL	1540
ní	/ni:/	este	983
nán	/na.n/	este=pues	742
maráme	/mara.me/	PL = MR	677
ó	/o:/	que	516
rriquí	/riki:/	(3)-3ERG-decir	472
níni	/ni:ni/	esto= LIG	428
quí	/ki:/	(3)-decir	385
chí	/ʃi:/	todo	377
epéme	/epe.me/	NEG <sub>1</sub> = NEG <sub>2</sub>	346
jué	/xue:/	en.efecto	249
ninháfa	/niŋa:ɸa/	así= ENF	226
malécu	/male.ku/	1l -persona	220

Finalmente, está claro que para complementar los resultados de este estudio, deben llevarse a cabo análisis similares de otras especies de habla malecu, sobre todo las espontáneas y no literarias. Esta tarea implica la grabación y la transcripción de una considerable cantidad de información, y debería realizarse lo antes posible, mientras la lengua siga viva. La importancia de un proyecto de este tipo sería enorme, ya que el corpus sería de gran utilidad para innumerables estudios lingüísticos —no solo fonológicos— del malecu. Asimismo, hace falta desarrollar trabajos similares con las otras lenguas de la familia chibchense.

### Abreviaturas

- linde morfemático  
= el elemento siguiente es un clítico  
. separa elementos españoles que en conjunto corresponden a un solo elemento malecu

(X) el elemento X presenta una realización Ø

1 primera persona inclusiva, absoluto

3 tercera persona, absoluto

3ERG tercera persona, ergativo

AUM aumentativo

ENF enfatizador

IL ilativo

INF infinitivo

LIG enclítico ligativo

MR modo real

NEG<sub>1</sub> partícula negativa, ‘no’

NEG<sub>2</sub> partícula negativa, ‘todavía no’

PLplural

RED reduplicación del aumentativo

### Notas

1. El malecu también se conoce como *guatuso* y pertenece a la rama vótica de la familia chibchense. Es hablado en tres aldeas, llamadas *palenques*, en el cantón de Guatuso en el norte de Costa Rica: Margarita, Tonjibe y El Sol.
2. Cabe advertir que los símbolos empleados por Constenla en esta cita son engañosos, ya que en el

AFI, [j] representa una articulación oclusiva, no africada. Según la descripción de dicho lingüista (1998: 11), la realización africada de este fonema parece ser [j̟]. El símbolo simplificado /j/ utilizado por Constenla para representar el fonema tampoco es afín con el AFI.

3. A diferencia del uso en el AFI, empleamos aquí el símbolo /a/ para la vocal central baja.
4. Los textos que incluyen estas vacilaciones son los que aparecen en Constenla (1983, 1992, 1999, 2003), y Constenla y Castro (2011, 2014), mientras que no están presentes en Constenla (1986, 1991), ni en Constenla, Castro y Blanco (1993).
5. Esta columna es redundante en el sentido de que el margen de error, indirectamente, ofrece la misma información. Sin embargo, se incluye para mostrar una forma alternativa de presentar el nivel de confiabilidad de los datos.
6. Además, para proporcionar esta información, también habría que definir cómo tratar las palabras incompletas.
7. Adaptado de Constenla y Castro (2011, 2014).

### Bibliografía

- Alameda, José Ramón y Fernando Cuetos. 1995. *Diccionario de frecuencias de las unidades lingüísticas del castellano*. Oviedo: Servicio de publicaciones de la Universidad de Oviedo.
- Barrantes, Ramiro, María Eugenia Bozzoli y Patricia Gudiño (eds.). 1986. *Memoria del Primer Simposio Científico sobre Pueblos Indígenas de Costa Rica*. San José: MOPT e Instituto Geográfico de Costa Rica.
- Betancourt de Sánchez, Helia y Adolfo Constenla Umaña. 1981. “La expedición al territorio de los guatusos: una crónica colonial hispana y su contraparte en la tradición oral indígena”. En: *Revista de Filología y Lingüística* 7 (1-2): 19-34.

- Constenla Umaña, Adolfo. 1983. "Descripción del sistema fonemático del guatuso". En: *Revista de Filología y Lingüística* 9 (1): 3-20.
- Constenla Umaña, Adolfo. 1986. "La función de una alternancia gramatical guatusa en el discurso narrativo tradicional". En: Barrantes, Bozzoli y Gudiño (comp.): 119-127.
- Constenla Umaña, Adolfo. 1991. "Tres textos guatusos del ciclo narrativo de las uniones con los animales". En: *Estudios de Lingüística Chibcha* 10: 101-119.
- Constenla Umaña, Adolfo. 1992. "Hagiografía y antihagiografía en la tradición oral guatusa". En: *Revista de Filología y Lingüística* XVIII (1): 83-124.
- Constenla Umaña, Adolfo. 1996. *Poesía tradicional indígena costarricense*. San José: Editorial de la Universidad de Costa Rica.
- Constenla Umaña, Adolfo. 1998. *Gramática de la lengua guatusa*. Heredia: Editorial de la Universidad Nacional.
- Constenla Umaña, Adolfo. 1999. "El respeto a la vida animal en la tradición oral guatusa". En: *Revista de Filología y Lingüística* XXV (2): 119-133.
- Constenla Umaña, Adolfo. 2003. "Dos textos guatusos sobre los profetas del cataclismo". En: *Estudios de Lingüística Chibcha* XXII: 61-128.
- Constenla Umaña, Adolfo, Eustaquio Castro y Antonio Blanco. 1993. *Lacá Majifjica. La transformación de la tierra*. Cartago: Editorial de la Universidad de Costa Rica.
- Constenla Umaña, Adolfo y Eustaquio Castro. 2011. *Pláticas sobre felinos*. San José: Editorial Universidad de Costa Rica.
- Constenla Umaña, Adolfo y Eustaquio Castro. 2014. *Muérrajá mausírrajáca. Pláticas sobre ogros*. San José: Editorial Universidad de Costa Rica.
- Fox, Anthony. 2000. *Prosodic features and prosodic structure: the phonology of suprasegmentals*. Nueva York: Oxford University Press.
- Herdan, Gustav. 1966. *The advanced theory of language as choice and chance*. Berlín: Springer.
- Heylen, Kris. 2005. "A quantitative corpus study of German word order variation". En: Kepser y Reis (eds.): 241-263.
- Kepser, Stephan y Marga Reis (eds.). 2005. *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*. Berlín: Mouton de Gruyter.
- McEnery, Tony y Andrew Wilson. 2001. *Corpus Linguistics: An introduction*. 2ª edición. Edimburgo: Edinburgh University Press.
- Moreno-Sandoval, Antonio, Doroteo Torre Toledano, Raúl de la Torre, Marta Garrote Salazar y José María Guirao. 2008. "Developing a phonemic and syllabic frequency inventory for spontaneous spoken Castilian Spanish and their comparison to text-based inventories". International Conference on Language Resources and Evaluation.
- Navarro Tomás, Tomás. 1946. *Estudios de Fonología Española*. Syracuse, Nueva York: University Press.



- Quilis, Antonio y Manuel Esgueva. 1980. "Frecuencia de fonemas en el español hablado". En: *Lingüística Española Actual* 2: 1-25.
- Reed, David W. 1949. "A statistical approach to quantitative linguistic analysis". En: *Word* 5 (3): 235-247.
- Sánchez Avendaño, Carlos. 2009. Situación sociolingüística de las lenguas minoritarias de Costa Rica y censos nacionales de población 1927-2000: Vitalidad, desplazamiento y autofiliación etnolingüística". En: *Revista de Filología y Lingüística* 35 (2): 233-273.
- Vinogradov, Igor. 2016. "Linguistic corpora of understudied languages: Do they make sense?" En: *Káñina* XL (1): 127-141.
- Zipf, George K. y Francis M. Rogers. 1939. "Phonemes and variophones in four present-day romance languages and classical Latin from the viewpoint of dynamic philology". *Archives Néerlandaises de Phonétique Expérimentale*.

