

FRECUENCIA DE FONEMAS¹

Hernán Emilio Pérez

heperez@udec.cl

Laboratorio de Fonética - Universidad de Concepción – Chile

Voicenet Speech - Chile

Resumen

En este artículo se exponen los resultados de un recuento de frecuencia de fonemas en un corpus oral de lectura de noticias en la televisión chilena. Además dichos resultados se comparan con cuatro investigaciones del mismo tipo realizadas para otros dialectos del español. El corpus está constituido por 15.649 palabras (3.281 diferentes), lo que equivale a un total de 75.269 realizaciones de fonemas.

Desde el punto de vista de la teoría fonológica, Emilio Alarcos Llorach expone:

La frecuencia con que los sonidos se repiten en la cadena hablada caracteriza, sin duda, fonéticamente una lengua dada. Igualmente, la frecuencia con que los fonemas aparecen en la elocución con valor distintivo contribuye a caracterizar fonológicamente tal lengua, mostrando el rendimiento funcional de cada uno de ellos. (Alarcos Llorach, 1965: 197)

Desde una perspectiva aplicada, Guirao y García Jurado afirman que este tipo de información interesa también a los investigadores de la percepción del habla, de la psicoacústica, de los procesos cognitivos, de la neurolingüística y de la inteligencia artificial, esto último, principalmente en el desarrollo de interfaces hombre-máquina y en el procesamiento de lenguaje natural:

En este caso el material estadístico brinda la oportunidad de almacenar en la memoria del sistema [de reconocimiento o generación automática del habla, o bien, de procesamiento de lenguaje natural] un inventario de sonidos clasificados por unidades (fonemas, sílabas y palabras) y según su frecuencia, distribución y posibles combinaciones. (Guirao & García Jurado, 1993:7)

Así también estos datos resultan útiles desde el punto de vista clínico y forense:

El recuento estadístico, por otra parte, provee criterios para predecir las unidades lingüísticas en mensajes o discursos distorsionados por anomalías del habla o por interferencias del medio en que se transmiten. [...] Las pruebas de sorderas, patologías del habla, reconocimiento de palabras y discursos en ambientes ruidosos, se van

modificando y perfeccionando con la incorporación de nuevos datos estadísticos acerca de la frecuencia con que determinadas emisiones sonoras aparecen en la lengua. (Guirao & García Jurado, 1993:7)

La frecuencia de los fonemas del español ha sido ya estudiada por Zipf y Rogers (1939), Navarro Tomás (1946), Alarcos Llorach (1965) y Guirao y García Jurado (1993).

Considerando que dentro de nuestro trabajo sobre la lectura de noticias en la televisión chilena contamos con una cantidad pertinente de datos para este tipo de análisis y considerando también que actualmente contamos con las herramientas necesarias para hacerlo, hemos realizado un recuento de la frecuencia de fonemas en nuestro corpus de noticias y compararlo con los resultados de las investigaciones anteriores.

Nuestra muestra está compuesta por dos horas de grabación de los noticieros de los dos canales de mayor sintonía en Chile, lo que constituye un total de 64 notas periodísticas en las que participan como hablantes 136 sujetos diferentes entre locutores, periodistas, entrevistados y comentaristas. El corpus está constituido por 15.649 palabras (3.821 diferentes), lo que equivale a un total de 75.269 realizaciones de fonemas.

Asumiendo que los fonemas son unidades abstractas y debido a que existen diversas maneras de representarlos, ya sea como complejos de rasgos o asociándolos con alguna de sus realizaciones fonéticas o bien mediante alguna notación convencional, hemos optado por este último criterio, de manera operativa y

¹ Todo lo expuesto en este artículo corresponde a una parte de los resultados de mi tesis doctoral "Fonética segmental del español público de Chile: un estudio de variación estilística de la pronunciación en los noticieros de la televisión chilena", proyecto FONDECYT N° 2000142.

con fines meramente expositivos, utilizando algunos de los símbolos que en el alfabeto corriente más o menos se corresponden con dichas entidades fonológicas. De este modo podemos postular el siguiente cuadro fonológico para el español de Chile:

Consonantes		
p	t	k
b	d	g
	f	s
		x
		y
		ch
m	n	ñ
	l	
	r	rr
Vocales		
	i	u
	e	o
	a	

Se puede notar que la distribución espacial del cuadro propuesto es relativamente cercana a la clasificación de segmentos que se hace en fonética articuladora (Navarro Tomás, 1957; Quilis & Fernández, 1969) usando los criterios de zona de articulación, modo de articulación y sonoridad para las consonantes y localización y altura para las vocales, pero esta disposición es más bien accesoria en términos de proponer algún orden más allá del meramente lineal, el cual permite además apreciar la naturalidad de ciertas clases y las correlaciones que, entre éstas, se dan.

En la *Tabla 1* se puede apreciar el detalle de los porcentajes de frecuencia de aparición en nuestro corpus para cada uno de los fonemas.

Tabla 1. Distribución porcentual de la frecuencia de los fonemas del español

/ i /	7.46	/ f /	0.75
/ e /	14.13	/ s /	9.61
/ a /	12.31	/ x /	0.74
/ o /	9.28	/ ch /	0.32
/ u /	3.05	/ y /	0.69
/ p /	2.58	/ m /	2.62
/ t /	4.92	/ n /	7.78
/ k /	3.94	/ ñ /	0.24
/ b /	1.92	/ r /	6.19
/ d /	4.84	/ rr /	0.64
/ g /	0.94	/ l /	5.05

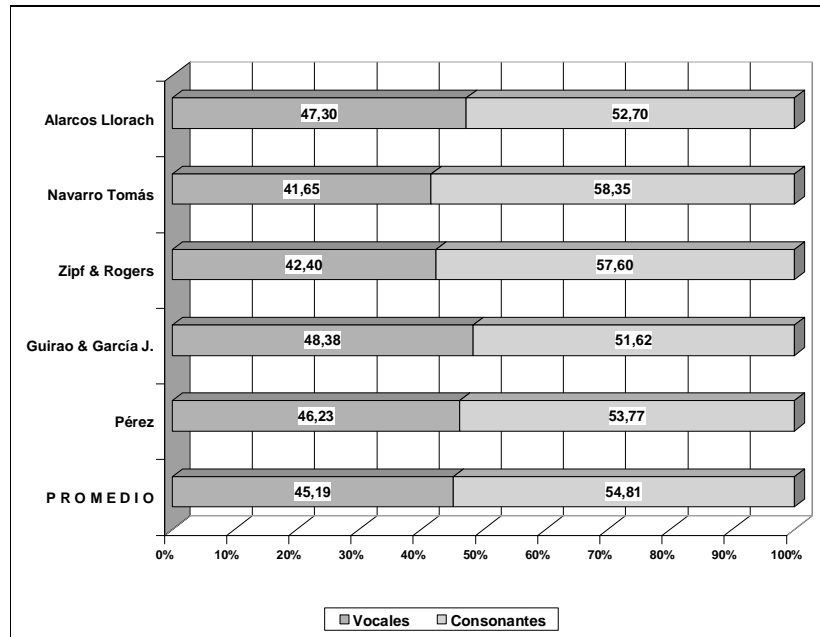
En la *Tabla 2* se puede apreciar el detalle de los resultados porcentuales para cada uno de los fonemas de cada uno de los estudios en cuestión, más los resultados de nuestro estudio, más el promedio de las distribuciones de todos los estudios.

Tabla 2. Comparación de la distribución porcentual de la frecuencia de los fonemas del español

	Alarcos Llorach	Navarro Tomás	Zipf y Rogers	Guirao y García J.	Pérez	Prom.
/ i /	8.60	4.76	4.20	6.59	7.46	6.32
/ e /	12.60	11.75	12.20	14.99	14.13	13.13
/ a /	13.70	13.00	14.06	13.27	12.31	13.27
/ o /	10.30	8.90	9.32	10.75	9.28	9.71
/ u /	2.10	1.92	1.76	2.79	3.05	2.32
/ p /	2.10	3.04	2.92	2.68	2.58	2.66
/ t /	4.60	4.82	4.46	4.48	4.92	4.66
/ k /	3.80	4.23	3.84	4.31	3.94	4.02
/ b /	2.50	2.54	3.26	3.08	1.92	2.66
/ d /	4.00	5.00	5.06	4.00	4.84	4.58
/ g /	1.00	1.04	1.02	1.11	0.94	1.02
/ B /	0.10					
/ D /	0.25					
/ G /	0.25					
/ f /	1.00	0.72	0.72	0.53	0.75	0.74
/ z /	1.70	2.23	1.74			1.89
/ s /	8.00	8.50	8.12	9.39	9.61	8.72
/ x /	0.70	0.51	0.58	0.70	0.74	0.65
/ ch /	0.40	0.30	0.30	0.40	0.32	0.34
/ y /	0.40	0.40	2.40	0.72	0.69	0.92
/ m /	2.50	2.40	2.98	3.17	2.62	2.73
/ n /	2.70	2.94	5.94	7.14	7.78	5.30
/ ñ /	0.20	0.36	0.36	0.24	0.24	0.28
/ N /	3.70	4.69				4.20
/ r /	2.50	2.40	5.90	5.40	6.19	4.48
/ rr /	0.60	0.80	1.04	0.39	0.64	0.69
/ R /	4.50	3.51				4.01
/ l /	4.70	5.46	5.20	3.88	5.05	4.86
/ ll /	0.50	0.60	0.60			0.57

En nuestro análisis, las vocales representan un 46.23% del total de los fonemas y las consonantes el 53.77% restante, proporción muy semejante a los resultados de cada uno de los estudios aquí comparados y al promedio de todos los recuentos; o sea, siempre un poco menos de la mitad del universo para las vocales (*Figura 1*).

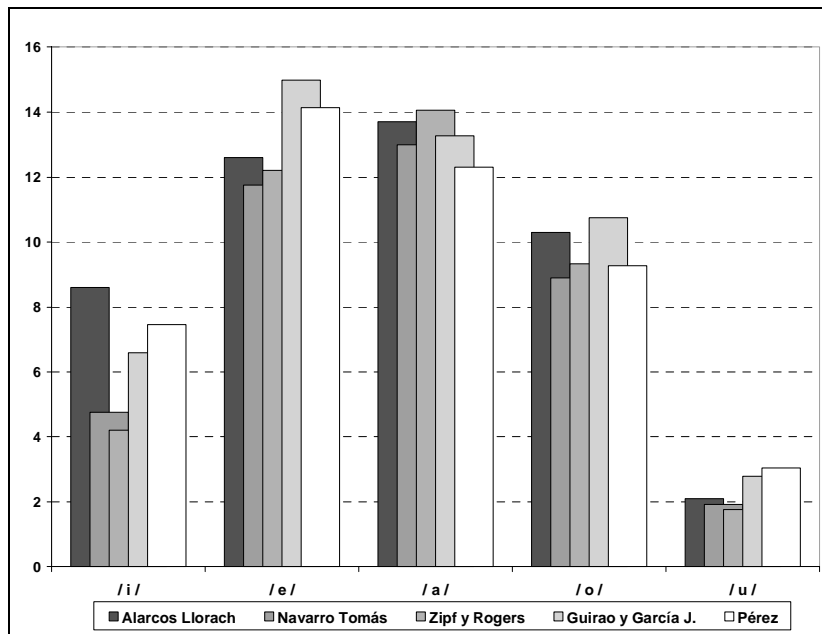
Figura 1. Gráfico comparativo de la distribución porcentual de las vocales y las consonantes



La vocal, y a la vez el fonema, más frecuente, es la /e/ (14.99% y 14.13% en nuestro recuento) en los estudios orientados al español de Latinoamérica y la /a/ (13.7%, 13% y 14.06%) para el caso de los estudios orientados hacia el español peninsular. En nuestro estudio la vocal /a/ alcanzó un índice de

un 12.31% de frecuencia de aparición. Cabe destacar, en este último caso, que alrededor de un 5% de las ocurrencias de la vocal (462 casos de 9.267) y un 0.6% del total de segmentos del corpus corresponden a la realización de la preposición "a".

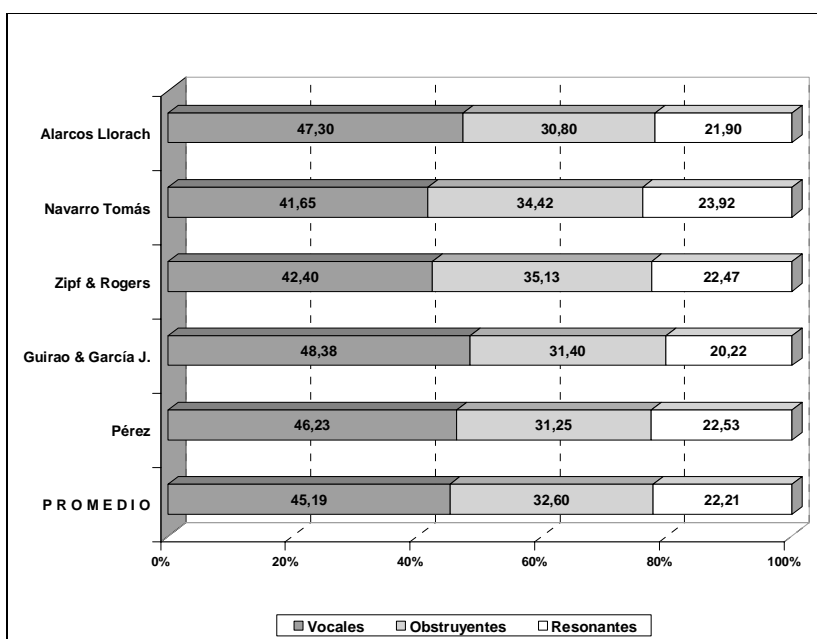
Figura 2. Gráfico comparativo de la distribución porcentual de las vocales



La /o/ (9.28% en nuestro estudio y 9.71% promedio) aparece en el tercer lugar y la /u/ sería la vocal menos frecuente (nunca más allá del 3.05% alcanzado en nuestro recuento). Los resultados más dispares los presenta la vocal /i/, oscilando entre un 4.2% y un 8.6% (7.46% para el caso de nuestro corpus), no alcanzando a ser, de todos modos, más frecuente que la /o/ en ninguno de los casos (*Figura 2*).

Con respecto a las consonantes, un 31.25% del corpus representa la frecuencia de ocurrencia de las obstruyentes (32.6% en promedio) lo que equivale a un 58.11% del total de consonantes y un 22.53% a las resonantes (22.21% promedio) equivalente al 41.89% de todas las realizaciones consonantes, proporciones bastante semejantes en todos los estudios (*Figura 3*).

Figura 3. Gráfico comparativo de la distribución porcentual de las vocales, las consonantes obstruyentes y las consonantes resonantes



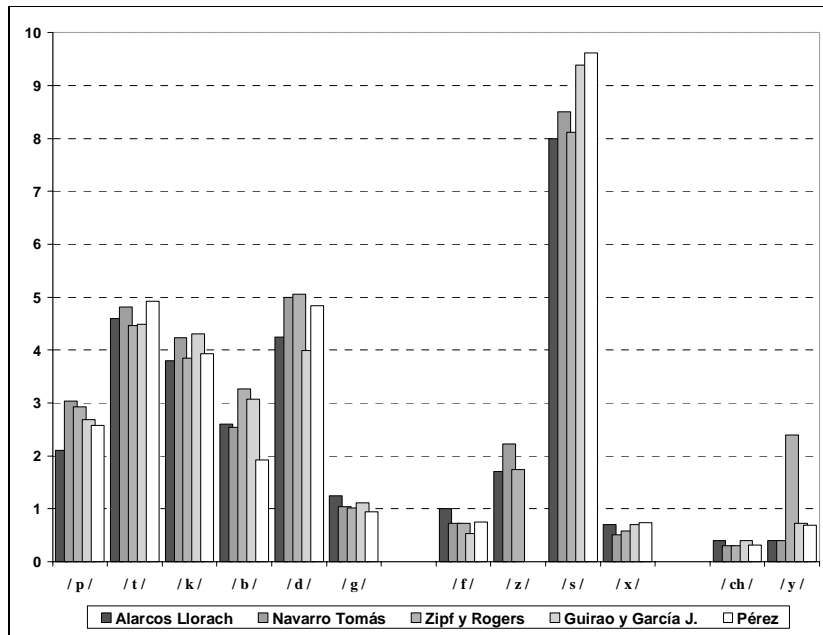
La más frecuente de las consonantes obstruyentes, e incluso de todas las consonantes, es la /s/ (sobre 8% en todos los estudios). Los dos estudios orientados al español de América muestran un índice sobre el 9.3% (9.61% más exactamente en nuestro caso), porque, como es bien sabido, éstos no consideran la consonante cuya realización es el fono interdental áfono (*Figura 4*).

De las consonantes obstruyentes, las menos frecuentes son las africadas, que manifiestan no más de un 0.4% de frecuencia en casi todos los casos, exceptuando /y/ que, por un lado, en el estudio de Zipf y Rogers aparece con un índice de un 2.4%, hecho cuya causa, según Alarcos Llorach, es "seguramente por haber tomado por fonema /y/ las variantes fonéticamente consonantes de /i/" (Alarcos Llorach, 1965:200); y que, por otro lado, aparece con un

índice cercano al 0.7% en los dos estudios orientados hacia el español de América (*Figura 4*).

En cuanto a la correlación /p-t-k/ vs. /b-d-g/ todos los estudios encontraron una proporción más o menos similar en los pares mínimos cuyas realizaciones se articulan en la zona anterior, siendo el más frecuente el de las consonantes cuyas realizaciones son los segmentos postdentales (alrededor de un 4.6% en promedio). La diferencia más significativa se manifiesta entre /k/ y /g/ (4.02% y 1.02% respectivamente), hecho que coloca a la consonante de realización velar áfona en el tercer lugar en términos de frecuencia dentro de la correlación, superando incluso a /d/ en el estudio de Guirao y García Jurado. En todos los estudios /g/ es la menos frecuente, no llegando a superar el 1.1% (*Figura 4*).

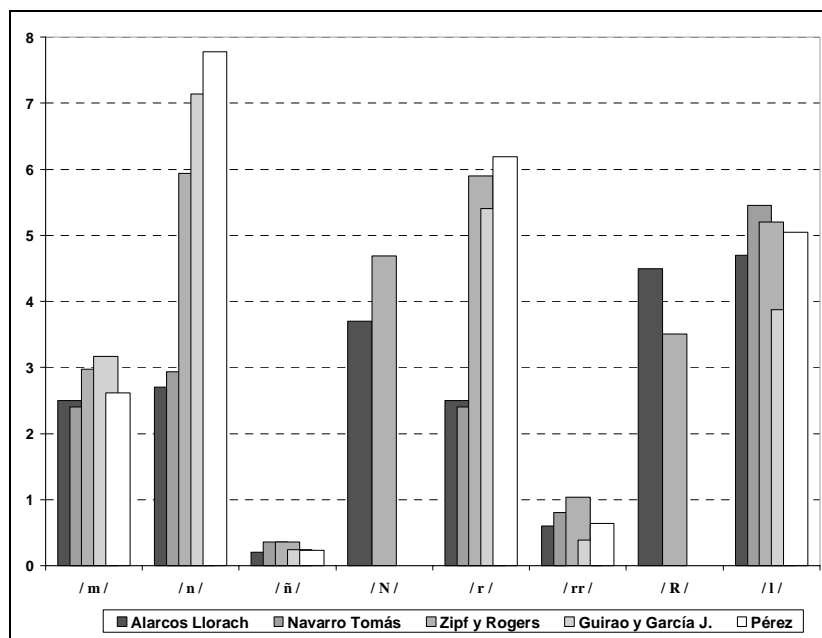
Figura 4. Gráfico comparativo de la distribución porcentual de las consonantes obstruyentes



Con respecto a las consonantes resonantes, en las nasales se dan los dos casos extremos, la resonante que mayor frecuencia de aparición manifiesta es /n/ con un 5.3% y la de menor

aparición es /ñ/ con un 0.28%, siendo a la vez, en promedio, el fonema menos frecuente de todos (Figura 5).

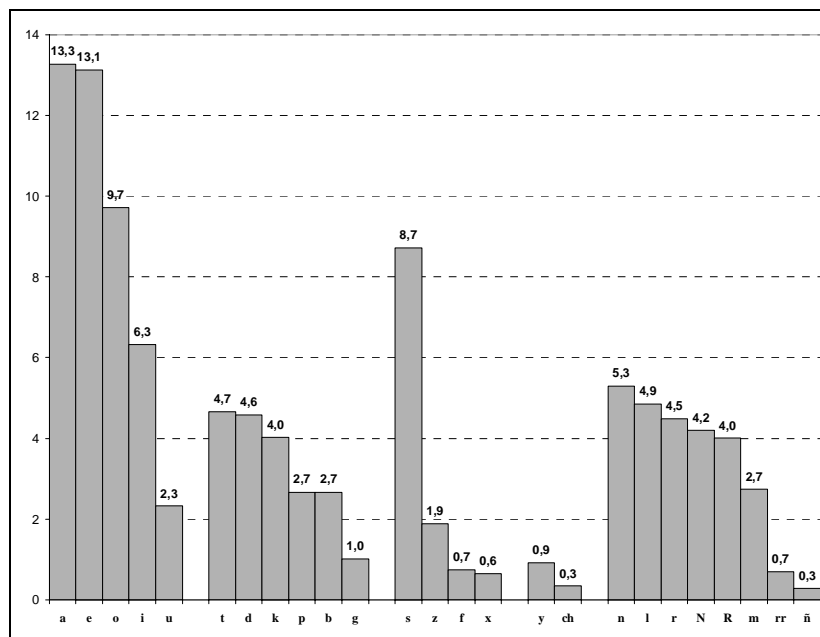
Figura 5. Gráfico comparativo de la distribución porcentual de las consonantes resonantes



Cabe destacar que tanto Alarcos Llorach como Zipf y Rogers consideran, como puede verse en la *Figura 5*, los archifonemas /N/ y /R/, los que presentan un porcentaje de aparición que no puede ignorarse (4.2% y

4.01% respectivamente) y que se refleja en la significativa diferencia en los porcentajes de /n/ y /r/ en dichos estudios con respecto a sus similares.

Figura 6. Gráfico comparativo de la distribución porcentual promedio de la frecuencia de los fonemas del español



Cabe destacar también que Alarcos Llorach considera además los archifonemas /B/, /D/ y /G/; pero en estos casos los porcentajes son tan bajos que no implican mayores diferencias en ningún sentido.

Finalmente, en la *Figura 6*, se puede observar la distribución frecuencial promedio de los elementos de todo el sistema, considerando además las diferencias de criterio en los diferentes estudios.

En síntesis, podemos decir que vocales y consonantes se distribuyen en el sistema en proporciones casi similares, manifestando un porcentaje levemente más alto las consonantes, pero reducido si consideramos la cantidad de elementos que compone cada conjunto.

Desde el punto de vista sistémico, podemos decir que no se aprecia ninguna relación entre frecuencia de aparición y correlaciones o clases.

Desde el punto de vista acústico, podemos decir que se aprecia una mayor tendencia hacia los sonidos agudos y a los que presentan una mayor sonorancia (energía sonora no atenuada).

Desde el punto de vista articulatorio, podemos apreciar que existe una mayor

frecuencia de fonemas de realizaciones medias y centrales, más que de realizaciones extremas, pero siempre con una leve tendencia hacia la zona anterior. Así también podemos decir que hay una tendencia hacia la salida fluida de la corriente de aire, más que a los sonidos que se generan con interrupciones.

Referencias bibliográficas

ALARCOS LLORACH, E. 1965. *Fonología española*. Madrid, Gredos

GUIRAO, M & M. GARCÍA JURADO. 1993. *Estudio estadístico del español*. Buenos Aires, Consejo Nacional de Investigaciones Científicas y Técnicas

NAVARRO TOMÁS, T. 1946. "Escala de frecuencia de los fonemas españoles" en *Estudios de fonología española*. Syracuse, pp. 15-30, citado en Alarcos Llorach 1965.

_____. 1957. *Manual de pronunciación española*. New York, Hafner Publishing Company.

QUILIS, A & J. FERNÁNDEZ. 1969. *Curso de fonética y fonología españolas*. Madrid, Consejo Superior de Investigaciones Científicas.

ZIPF, G. K. & F. M. ROGERS. 1939. "Phonemes and Variophones in four present-day Romance Languages and Classical Latin from the viewpoint of dynamic Philology", en *Archives Néerlandaises de Phonétique Expérimentale* 15, pp. 111-147, citado en Alarcos Llorach 1965.