

Importancia económica de las características fonéticas del idioma español y sus variedades regionales

Germán Coloma*

Universidad del CEMA

Resumen

El presente trabajo busca cuantificar económicamente la importancia de cinco características fonéticas (seseo, yeísmo, aspiración del fonema /s/, rehilamiento del fonema /ʒ/ y aspiración del fonema /x/), cuya presencia o ausencia sirve para distinguir entre diez variedades regionales del idioma español. Para ello se efectúa un análisis de la extensión de dichas características y se estiman las que corresponden al “hablante promedio” de español, que resultan ser idénticas a las del “español mexicano-centroamericano”. A través de estimaciones de ingreso per cápita, sin embargo, se llega a la conclusión de que la característica asociada con un mayor ingreso es la ausencia de seseo (típica del español peninsular norteño). Por último, mediante un análisis de regresión estadística, se verifica que dicha característica está asociada con un ingreso anual medio por habitante que es de U\$S 20.000 más que el de los hablantes seseantes. Dicho análisis nos muestra también que las otras dos características estadísticamente significativas como señales de un ingreso más elevado son la presencia de yeísmo y la ausencia de aspiración del fonema /x/.

Variedades regionales y características fonéticas

Aunque el español es un idioma relativamente compacto, en el sentido de que la mayoría de sus hablantes pueden comunicarse perfectamente aunque provengan de regiones muy distantes entre sí, existe una cantidad de características (fonéticas, morfológicas, sintácticas, léxicas, etc.) que permiten identificar diferentes dialectos o variedades regionales dentro de la lengua española. En este trabajo hemos identificado cinco de dichas características, todas ellas de carácter fonético, que son las siguientes:

1) Seseo: ausencia de distinción entre los fonemas /s/ y /θ/, y su reemplazo por un único fonema que en general se articula como una consonante alveolar fricativa sorda [s].

2) Yeísmo: ausencia de distinción entre los fonemas /y/ y /λ/, y su reemplazo por un único fonema que se articula utilizando alguno de los alófonos de /y/ (que en general es una consonante palatal aproximante sonora).

3) Aspiración de /s/: uso de una consonante fricativa aspirada glotal [h] como alófono del fonema /s/, especialmente cuando dicho fonema aparece en posición preconsonántica.

4) Aspiración de /x/: uso de [h] como articulación principal del fonema /x/, en vez de utilizar la consonante fricativa velar sorda [x].

5) Rehilamiento de /y/: articulación del fonema /y/ a través de una consonante asibilada posalveolar, que puede tener un sonido africado sonoro [dʒ], fricativo sonoro [ʒ] o fricativo sordo [ʃ].

1 * Agradezco los comentarios de Laura Colantoni, Mariana Conte Grand, Alejandro Corbacho, John Lipski, Andrés López, Francisco Moreno Fernández, Carolina Reynoso Savio y Jorge Streb a una versión anterior del presente trabajo. Los errores que subsisten son de mi exclusiva responsabilidad.

La distribución geográfica conjunta de estas características fonéticas permite identificar los dialectos del español según que cada una de dichas características se encuentre o no presente (ver cuadro 1), y delimita seis variedades en España (castellano tradicional, castellano moderno, peninsular mixto, valenciano, andaluz occidental y andaluz-canario) y diez en América Latina (mexicano occidental, mexicano oriental, caribeño, andino-centroamericano, andino ecuatorial, andino altoperuano, boliviano oriental, paraguayo, cordillerano y rioplatense). La caracterización en cuestión se llevó a cabo utilizando información de Moreno Fernández (2009) y Hualde (2005), y para algunos casos se emplearon además otros trabajos que analizan las características particulares de algunos países.² Nótese que algunos de los dialectos identificados son idénticos entre sí (en lo que se refiere a los cinco rasgos fonéticos individualizados). Esto ocurre con el español mexicano occidental y el español valenciano, y con el español caribeño y el español andaluz-canario.

Cuadro 1: Diferencias fonéticas entre las variedades regionales del español

Variedad / Rasgo fonético	Seseo	Yeísmo	Aspir /s/	Aspir /x/	Rehil /y/
Castellano tradicional	No	No	No	No	No
Castellano moderno	No	Sí	No	No	No
Peninsular mixto	No	Sí	Sí	No	No
Valenciano	Sí	Sí	No	No	No
Andaluz occidental	Sí	Sí	Sí	Sí	Sí
Andaluz-canario	Sí	Sí	Sí	Sí	No
Mexicano occidental	Sí	Sí	No	No	No
Mexicano oriental	Sí	Sí	No	Sí	Sí
Caribeño	Sí	Sí	Sí	Sí	No
Andino-centroamericano	Sí	Sí	No	Sí	No
Andino ecuatorial	Sí	No	No	Sí	No
Andino altoperuano	Sí	No	No	No	No
Boliviano oriental	Sí	No	Sí	No	No
Paraguayo	Sí	No	Sí	No	Sí
Cordillerano	Sí	Sí	Sí	No	No
Rioplatense	Sí	Sí	Sí	No	Sí

Importancia demográfica y económica de las características fonéticas

La tipificación de las distintas variedades regionales del idioma español, de acuerdo con las características fonéticas que poseen, posibilita efectuar un análisis cuantitativo de la importancia relativa de dichas características dentro del universo de hablantes de la lengua, tanto desde el punto de vista demográfico como económico. Para ello es necesario medir primero de manera aproximada el número de hablantes de las distintas variedades de español descriptas, así como el producto bruto generado por dichos hablantes. A tal efecto hemos tomado como base información del Banco Mundial (2009), que tiene la ventaja de que el producto bruto aparece medido utilizando unidades monetarias comparables, que son dólares estadounidenses ajustados por el concepto de “paridad del poder adquisitivo” (ver cuadro 2). A fin de cuantificar el ingreso medio por habitante, hemos calculado también el producto bruto interno *per*

2 Para el caso argentino se ha tomado información de Fontanella (2000), para el mexicano de Martín Butragueño (2010), para el peruano de Escobar (1978) y para el español peninsular de Lipski (2004).

cápita (PBIpc) correspondiente a cada país, que surge de dividir el producto bruto total por la población del país en cuestión.

De los veinte países que aparecen en el cuadro 2, hay doce en los cuales predomina una sola variedad regional de las mencionadas en el cuadro 1, que son: Guatemala, Honduras, Nicaragua, El Salvador y Costa Rica (español andino-centroamericano), Panamá, Cuba, República Dominicana y Puerto Rico (español caribeño), Chile (español cordillerano), Paraguay (español paraguayo) y Uruguay (español rioplatense). En los restantes ocho (España, México, Colombia, Venezuela, Ecuador, Perú, Bolivia y la Argentina), podemos identificar en cambio distintas áreas geográficas correspondientes a dialectos diferentes. Utilizando información de población y PBI por división administrativa de cada país, hemos efectuado entonces dicha división, y el resultado de la misma aparece resumido en el cuadro 3.

Cuadro 2: Habitantes y producto bruto por país hispanohablante (2008)

País	Población		Producto bruto		PBIpc (U\$\$/año)
	Millones	%	Mil mill U\$\$	%	
España	45,56	11,14%	1.442,94	26,00%	31.674
México	106,35	26,00%	1.549,49	27,92%	14.570
Guatemala	13,69	3,35%	65,15	1,17%	4.760
Nicaragua	5,67	1,39%	15,24	0,27%	2.689
El Salvador	6,13	1,50%	41,71	0,75%	6.799
Costa Rica	4,52	1,10%	50,76	0,91%	11.232
Honduras	7,32	1,79%	28,78	0,52%	3.932
Panamá	3,40	0,83%	42,48	0,77%	12.497
Cuba	11,21	2,74%	110,80	2,00%	9.888
República Dominicana	9,95	2,43%	80,87	1,46%	8.125
Puerto Rico	3,95	0,97%	86,50	1,56%	21.877
Venezuela	27,94	6,83%	358,07	6,45%	12.818
Colombia	45,01	11,01%	395,99	7,14%	8.797
Ecuador	13,48	3,30%	108,04	1,95%	8.015
Perú	28,84	7,05%	245,36	4,42%	8.508
Bolivia	9,69	2,37%	41,46	0,75%	4.277
Chile	16,80	4,11%	242,58	4,37%	14.436
Argentina	39,88	9,75%	570,85	10,29%	14.313
Uruguay	3,33	0,82%	42,49	0,77%	12.744
Paraguay	6,24	1,53%	29,35	0,53%	4.704
Total	408,96	100,00%	5.548,89	100,00%	13.568

En base a las cifras del cuadro 3 y a la relación entre variedades regionales y rasgos fonéticos resumida en el cuadro 1, resulta posible relacionar dichos rasgos fonéticos con los datos de población y producto bruto. Eso es lo que aparece en el cuadro 4, en el cual hemos incluido también columnas referidas al producto bruto *per cápita* de los hablantes de español que poseen y que no poseen cada una de las características fonéticas consideradas.

Cuadro 3: Habitantes y producto bruto por variedad regional (2008)

Variedad regional	Población		Producto bruto		PBIpc (U\$\$/año)
	Millones	%	Mil mill U\$\$	%	
Castellano tradicional	2,39	0,59%	84,48	1,52%	35.281
Castellano moderno	24,69	6,04%	872,71	15,73%	35.349
Peninsular mixto	5,16	1,26%	132,24	2,38%	25.646
Valenciano	3,03	0,74%	91,55	1,65%	30.184
Andaluz occidental	5,16	1,26%	132,19	2,38%	25.635
Andaluz-canario	5,13	1,25%	129,77	2,34%	25.311
Mexicano occidental	85,12	20,81%	1.348,83	24,31%	15.845
Mexicano oriental	16,80	4,11%	175,25	3,16%	10.429
Caribeño	61,38	15,01%	684,55	12,34%	11.153
Andino- centroamericano	108,25	26,47%	861,74	15,53%	7.960
Andino ecuatorial	8,62	2,11%	75,21	1,36%	8.726
Andino altoperuano	13,46	3,29%	59,71	1,08%	4.436
Boliviano oriental	3,03	0,74%	13,32	0,24%	4.399
Paraguayo	9,93	2,43%	49,79	0,90%	5.012
Cordillerano	21,57	5,27%	293,96	5,30%	13.627
Rioplatense	35,23	8,62%	543,60	9,80%	15.429
Total	408,96	100,00%	5.548,89	100,00%	13.568

De las cifras del cuadro 4 puede inferirse que algunos rasgos fonéticos son muy mayoritarios entre los hablantes de la lengua castellana (como por ejemplo el seseo y el yeísmo, que son compartidos por más del 90% de los hispanohablantes), en tanto que otros son relativamente minoritarios (por ejemplo, la aspiración de /s/ y el rehilamiento de /y/). La aspiración de /x/ también es mayoritaria si se la evalúa en términos del porcentaje de la población que la utiliza (50,21%), pero no si se la mide en términos del producto bruto generado por dicha población (37,10%).

Cuadro 4: Habitantes y producto bruto por característica fonética (2008)

Característica	Población		Producto bruto		PBI <i>per cápita</i>	
	Millones	%	Mil mill U\$\$	%	Sí	No
Seseo	376,72	92,12%	4.459,46	80,37%	11.838	33.792
Yeísmo	371,53	90,85%	5.266,38	94,91%	14.175	7.547
Aspiración de /s/	146,58	35,84%	1.979,42	35,67%	13.504	13.604
Aspiración de /x/	205,34	50,21%	2.058,70	37,10%	10.026	17.140
Rehilamiento de /y/	67,13	16,41%	900,83	16,23%	13.420	13.597

Las cifras que aparecen en el cuadro 4 son también útiles para caracterizar un “hablante promedio del idioma español”. Dicho hablante es alguien que posee al mismo tiempo seseo, yeísmo y aspiración de /x/, pero que no aspira el fonema /s/, ni rehila el fonema /y/. Dichas características coinciden exactamente con las que presenta un dialecto determinado del idioma castellano, que es el español andino-centroamericano. Sin embargo, si en vez de medir los porcentajes en términos de población lo hacemos en términos de producto bruto, el hablante promedio pasa a ser alguien

que es seseante y yeísta pero que no presenta aspiración de /x/. Dicho hablante coincide con el de español mexicano occidental (o con el de español valenciano).

De la información referida al PBI *per cápita* de las personas que poseen y que no poseen cada uno de los rasgos fonéticos individualizados, por su parte, llama la atención que para algunas características existen diferencias muy grandes. Si tomamos las cifras del cuadro 4 y vemos qué características están asociadas con mayores niveles de PBI *per cápita* nos encontramos con el resultado esperable de que las mismas coinciden con las del hablante de español castellano moderno, ya que los hablantes de mayor ingreso son los que poseen yeísmo pero no seseo, ni aspiración de /s/ o de /x/, ni rehilamiento de /y/.

Valoración económica mediante un análisis de regresión estadística

La información de los cuadros 2, 3 y 4 resulta útil para tener una idea aproximada de la importancia económica de los distintos países, dialectos y rasgos fonéticos que caracterizan a dichos dialectos, pero no nos dicen qué rasgos son más importantes como signos de un ingreso medio más elevado, ni cómo dichos rasgos interactúan entre sí. Una forma analítica de abordar ese problema es utilizar una metodología elaborada originalmente para el campo de la economía, y que se conoce con el nombre de “valuación por precios hedónicos”.

La valuación por precios hedónicos consiste en correr una regresión por mínimos cuadrados en la cual la variable dependiente es una magnitud monetaria (por ejemplo, el precio de algún bien, o el ingreso de un grupo de personas) y las variables independientes representan características asociadas con el bien o grupo de personas al cual se refiere la variable dependiente. El objetivo de este método es asignarle un valor implícito a dichas características asociadas, a través de una lógica de descomposición del valor total a explicar en una serie de “valores parciales” imputables a las características a evaluar.³

El análisis de regresión por mínimos cuadrados es relativamente común en fonética para encontrar correlaciones entre variables de tipo acústico que tienen que ver con la caracterización de los sonidos. También es muy común en economía para explicar el comportamiento de variables como puede ser el PBI *per cápita*. No resulta en cambio habitual para correlacionar variables económicas con variables lingüísticas, si bien existen algunos trabajos que han intentado avanzar en dicha línea, sobre todo en lo que se refiere a la relación entre capacidades lingüísticas y nivel de ingreso.⁴ En la literatura sobre sociolingüística existe además toda una corriente que ha empleado distintos tipos de regresión estadística en los cuales la variable dependiente es una variable fonética y las variables independientes tienen que ver con características tales como la edad de las personas, su sexo, su clase social o su nivel de ingreso.⁵

En relación con estos temas, sin embargo, resulta útil diferenciar entre modelos que tratan de utilizar las variables lingüísticas como explicativas de las variables económicas y modelos en los cuales lo que se intenta es correlacionar y medir el valor de determinada característica de la lengua como signo de una variable socioeconómica. Esto último es lo que intentamos hacer en esta sección de nuestro trabajo, en la cual corremos una regresión que tiene la siguiente forma:

$$PBIPC = \alpha_0 + \alpha_1 * SESEO + \alpha_2 * YEISMO + \alpha_3 * SPIRS + \alpha_4 * SPIRX + \alpha_5 * REHIL;$$

donde PBIPC es la variable dependiente; SESEO, YEISMO, SPIRS, SPIRX y REHIL son las variables independientes; y α_0 , α_1 , α_2 , α_3 , α_4 y α_5 son los coeficientes a estimar.

3 Para una reseña de la literatura económica sobre precios hedónicos que incluye una explicación detallada de las metodologías a utilizar, véase Nesheim (2006).

4 Para una buena reseña en español de esta literatura sobre “economía de la lengua”, véase Jiménez (2006).

5 Sobre este punto, véase Labov (2006).

El total de observaciones utilizadas en nuestra regresión es de 36, y corresponden a los 12 países en los cuales hemos identificado una única variedad de español y a 24 observaciones más de los ocho países en los cuales hemos identificado más de una variedad (6 observaciones para España, 3 para Colombia, la Argentina, México y Bolivia, y 2 para Venezuela, Ecuador y Perú). La regresión se llevó a cabo utilizando un método de mínimos cuadrados en el cual el peso de las observaciones se ponderó por el número de habitantes asociado a cada observación, y se le aplicó una corrección por heterocedasticidad basada en la metodología de White (1980).⁶

Los resultados del análisis de regresión son los que aparecen en el cuadro 5. En él hemos expuesto los valores obtenidos para los coeficientes estimados, así como también el correspondiente desvío estándar, el estadístico-t, y el valor-p de dichos coeficientes, que nos muestran una estimación de la probabilidad de que los mismos sean en rigor iguales a cero. Otro resultado de la regresión que vale la pena reportar es su coeficiente de determinación (R^2), que en este caso tomó un valor igual a 0,9832 (el cual puede considerarse como indicativo de un ajuste estadístico muy bueno).

Cuadro 5: Resultados del análisis de regresión del PBI *per cápita*

Variable	Coficiente	Desv std	Estad-t	Valor-p
Constante	26592,45	2355,63	11,28888	0,0000
Seseo	-19342,13	639,04	-30,26755	0,0000
Yeísmo	8399,08	2291,43	3,66543	0,0009
Aspiración de /s/	724,02	1124,76	0,64371	0,5247
Aspiración de /x/	-5758,55	610,87	-9,42682	0,0000
Rehilamiento de /y/	-234,25	1024,24	-0,22871	0,8206

Los resultados del análisis llevado a cabo muestran que, como signo del ingreso por habitante de quien la posee, la característica fonética que tiene un efecto más importante es el seseo, ya que su existencia reduce el valor esperado del PBI *per cápita* anual en una suma de U\$S 19.342, y es estadísticamente significativa al 1%. También son estadísticamente significativas al 1% el yeísmo (que se asocia con un aumento esperado del PBI *per cápita* de U\$S 8.399 anuales), y la aspiración de /x/ (que reduce el valor esperado del ingreso por habitante en U\$S 5.759 por año). Controlando por la presencia de seseo, yeísmo y aspiración de /x/, en cambio, la aspiración de /s/ y el rehilamiento de /y/ tienen un efecto que no resulta estadísticamente significativo para ningún nivel razonable de probabilidad.

Las cifras del cuadro 5 reafirman la idea de que las características fonéticas asociadas con un mayor nivel de ingreso coinciden en general con las del castellano moderno. Según esta visión, lo más relevante como signo de un ingreso más elevado es distinguir entre los sonidos de los fonemas /s/ y / /, no aspirar el fonema /x/ y no distinguir entre los fonemas / / y /y/, y no serían en cambio de importancia significativa la aspiración de /s/ ni el rehilamiento de /y/. Nótese, sin embargo, que esto no quiere decir que los rasgos fonéticos analizados puedan ser interpretados como causantes de incrementos o disminuciones en el ingreso *per cápita*, sino que los mismos serían más bien signos o “marcadores sociolingüísticos” de un nivel de ingreso más alto o más bajo de los hablantes con los cuales pueden ser asociados.

6 Para una descripción y justificación de la metodología utilizada, véase Coloma (2010).

Conclusiones

De todo lo expuesto en el presente trabajo, resulta posible extraer conclusiones sobre el valor de las características fonéticas que identifican las distintas variedades regionales de la lengua castellana. Una primera conclusión es que, dentro del universo de hablantes del idioma español, existe una dispersión muy considerable entre el nivel de ingreso de quienes presentan ciertos rasgos fonéticos respecto del nivel de ingreso de quienes presentan otros rasgos. Eso puede apreciarse observando los ingresos *per cápita* promedio asociados con la presencia o ausencia de determinadas características, pero un modo más preciso de aislar los valores de cada rasgo dialectal es emplear un procedimiento de regresión por mínimos cuadrados que utilice como variable dependiente al ingreso *per cápita* y calcule los correspondientes precios hedónicos de los distintos rasgos.

Aplicando dicho método, hemos visto que las tres características más significativas como signos de un estatus socioeconómico alto en el mundo hispanohablante son la ausencia de seseo, la presencia de yeísmo y la ausencia de aspiración de /x/. Cabe mencionar, sin embargo, que en este estudio hemos empleado únicamente cifras agregadas (por país, o por división administrativa dentro de cada país). Esto es una limitación para la metodología utilizada que, por supuesto, podría usarse también para analizar bases de datos en las cuales las observaciones se refirieran al ingreso y a los rasgos fonéticos dialectales de personas individuales. Los resultados obtenidos, sin embargo, podrían servir para plantear hipótesis que fueran luego de utilidad para explorar otras dimensiones relacionadas con la sociolingüística de los rasgos fonéticos dialectales del español, y ser de ese modo el paso inicial en futuras investigaciones sobre esos temas.

Referencias bibliográficas

- Banco Mundial. *Population and Gross Domestic Product 2008*. 2009. Washington DC, World Bank.
- Coloma, Germán. 2010. "An Econometric Method to Detect the Social Significance of Linguistic Variables", mimeo. Buenos Aires, Universidad del CEMA.
- Escobar, Alberto. 1978. *Variaciones sociolingüísticas del castellano en el Perú*. Lima, Instituto de Estudios Peruanos.
- Fontanella, Beatriz. 2000. *El español de la Argentina y sus variedades regionales*. Buenos Aires, Edicial.
- Hualde, José. 2005. *The Sounds of Spanish*. Nueva Cork, Cambridge University Press.
- Jiménez, Juan Carlos. 2006. "La economía de la lengua: una visión de conjunto", DT 01/06. Madrid, Instituto Complutense de Estudios Internacionales.
- Labov, William. 2006. "A Sociolinguistic Perspective on Sociophonetic Research", *Journal of Phonetics*, vol 34, pp. 500-515.
- Lipski, John. 2004. *Lecture Notes on the Spanish of Spain*. University Park, Pennsylvania State University.
- Martín Butragueño, Pedro. 2010. "La división dialectal del español mexicano", mimeo. México, El Colegio de México.
- Moreno Fernández, Francisco. 2009. *La lengua española en su geografía*. Madrid, Arco.
- Nesheim, Lars. 2006. "Hedonic Price Functions", CEMMAP WP 18/06. Londres, University College.
- White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", *Econometrica*, vol. 48, pp. 817-838.

GERMÁN COLOMA ES DOCTOR EN ECONOMÍA POR LA UNIVERSIDAD DE CALIFORNIA, LOS ANGELES (EE.UU), Y ACTUALMENTE SE DESEMPEÑA COMO PROFESOR TITULAR DE ECONOMÍA EN LA UNIVERSIDAD DEL CEMA (BUENOS AIRES, ARGENTINA), EN LA CUAL ES TAMBIÉN DIRECTOR DE LA MAESTRÍA EN EVALUACIÓN DE PROYECTOS Y CO-EDITOR DE LA REVISTA CIENTÍFICA JOURNAL OF APPLIED ECONOMICS. TIENE TRES LIBROS PUBLICADOS: ANÁLISIS ECONÓMICO DEL DERECHO, ECONOMÍA DE LA ORGANIZACIÓN INDUSTRIAL Y DEFENSA DE LA COMPETENCIA Y VEINTIDÓS ARTÍCULOS EN REVISTAS INTERNACIONALES CON REFERATO (BERKELEY ELECTRONIC JOURNAL OF THEORETICAL ECONOMICS, EMPIRICAL ECONOMICS, INTERNATIONAL JOURNAL OF BUSINESS AND ECONOMICS, INTERNATIONAL REVIEW OF LAW AND ECONOMICS, JOURNAL OF ECONOMIC DYNAMICS AND CONTROL, REVIEW OF INDUSTRIAL ORGANIZATION, ETC).