

UNA VERSIÓN ALTERNATIVA DE “EL VIENTO NORTE Y EL SOL” EN ESPAÑOL

GERMÁN COLOMA
UNIVERSIDAD DEL CEMA
BUENOS AIRES, ARGENTINA
gcoloma@cema.edu.ar

Resumen: En este trabajo se propone una versión española alternativa de la fábula “El viento norte y el sol”, utilizada por la Asociación Fonética Internacional para ilustrar la pronunciación de los distintos idiomas. La misma tiene una serie de ventajas respecto de la versión estándar que aparece en Martínez, Fernández y Carrera (2003) y en Monroy y Hernández (2015), ya que incluye dos fonemas más, tiene menos repeticiones de palabras, está más balanceada fonéticamente, y presenta una distribución de frecuencias de los fonemas que es más cercana a la distribución promedio de dichos fonemas en el lenguaje natural.

Palabras clave: Fonética española, frecuencia de fonemas, balance fonético.

Title: An Alternative Version of 'The North Wind and the Sun' in Spanish”.

Abstract: In this paper we propose an alternative Spanish version of the fable “The North

Wind and the Sun”, used by the International Phonetic Association to illustrate the pronunciation of different languages. This version has a number of advantages over the standard text that appears in Martínez, Fernández and Carrera (2003) and in Monroy and Hernández (2015), since it includes two additional phonemes, it has fewer word repetitions, it is more balanced phonetically, and it shows a phoneme frequency distribution that is closer to the average distribution in natural language.

Key words: Spanish phonetics, phoneme frequency, phonetic balance.

1. INTRODUCCIÓN

“El viento norte y el sol” es una fábula atribuida a Esopo, que viene siendo empleada desde hace más de cien años por la Asociación Fonética Internacional (IPA, por sus siglas en inglés) como un texto modelo para ilustrar la pronunciación de numerosas lenguas.¹ En español, la versión más comúnmente utilizada de esta fábula es la que apareció originalmente en IPA (1949), y luego nuevamente en Martínez, Fernández y Carrera (2003) y en Monroy y Hernández (2015).

La existencia de versiones de “El viento norte y el sol” en muchas lenguas distintas ha permitido llevar a cabo análisis comparativos sobre la fonética de diferentes idiomas, así como también otros que tienen que ver con comparaciones tipológicas o estadísticas.² Sin embargo, ha recibido también algunas críticas, sobre todo en lo que se refiere a su uso para ilustrar la pronunciación de lenguas no europeas. Dichas críticas señalan lo poco natural que en muchos casos puede resultar la historia para hablantes de idiomas originados en zonas cálidas, o en culturas que no tienen la tradición de personificar a las fuerzas de la naturaleza. Esto ha generado que algunos autores de ilustraciones del Alfabeto Fonético Internacional (AFI) hayan preferido emplear otros textos alternativos.³

Otra crítica que ha surgido acerca de “El viento norte y el sol”, y que encontró eco en la publicación de la propia IPA, es la expresada por Deterding (2006), quien resaltó el hecho de la gran repetición de palabras que presenta esta fábula, y de cómo eso puede ser un problema para el análisis fonético. También encontró, trabajando sobre el texto en inglés de “El viento norte y el sol”, que en el mismo no había ejemplos del fonema /z/, y que también faltaban algunos diptongos importantes y algunas combinaciones de consonantes que son comunes en el idioma inglés. En respuesta a ello, Deterding propuso la uti-

1 Véase IPA (1912), IPA (1949) e IPA (1999), o las numerosas ilustraciones del Alfabeto Fonético Internacional publicadas desde 1990 en el *Journal of the International Phonetic Association*.

2 Véase, por ejemplo, Coloma (2014), que es un estudio en el cual se utilizan estadísticas tomadas de versiones de “El viento norte y el sol” en diferentes idiomas para estudiar la posible existencia de correlación entre distintos aspectos de la complejidad lingüística.

3 Véase, por ejemplo, las ilustraciones de los idiomas taba (Bowden y Hajek, 1996), spokane (Carlson y Esling, 2000), ega (Connell, Ahoua y Gibbon, 2002), y bardi (Bowern, McDonough y Kelliher, 2012).

lización de un texto alternativo, que es una versión adaptada y más extensa de otra fábula de Esopo: “El pastor y el lobo”.

El objetivo del presente trabajo es evaluar la capacidad del texto en español de “El viento norte y el sol” de cumplir con una serie de requisitos que hacen a su aptitud para ilustrar la pronunciación de la lengua castellana en sus distintas variedades. Sin embargo, en vez de proponer la adopción de otro texto, lo que haremos será evaluar una versión alternativa de la fábula del viento norte y el sol, que tiene la misma extensión en cuanto a número de palabras pero que corrige sus principales problemas.⁴

En lo que sigue del trabajo reproduciremos primero el texto original y el alternativo, y haremos una comparación estadístico-descriptiva de los mismos (sección 2). Luego ilustraremos la capacidad de ambos textos para captar una serie de variantes en la pronunciación de los fonemas que son comunes dentro el ámbito hispanohablante (sección 3). A continuación (sección 4) estudiaremos el balance fonético de las dos versiones de la fábula, en tanto que en la sección 5 aproximaremos las distribuciones de frecuencias de los fonemas en los dos textos utilizando la denominada “distribución de Zipf”. La sección 6, por último, resumirá las principales conclusiones del trabajo.

2. LA FÁBULA DEL VIENTO NORTE Y EL SOL

La versión ortográfica de “El viento norte y el sol” que aparece en Martínez, Fernández y Carrera (2003) es la siguiente:

El viento norte y el sol porfiaban sobre cuál de ellos era el más fuerte, cuando acertó a pasar un viajero envuelto en ancha capa. Convinieron en que quien antes lograra obligar al viajero a quitarse la capa sería considerado más poderoso. El viento norte sopló con gran furia, pero cuanto más soplabla, más se arrebujaba en su capa el viajero; por fin el viento norte abandonó la empresa. Entonces brilló el sol con ardor, e inmediatamente se despojó de su capa el viajero; por lo que el viento norte hubo de reconocer la superioridad del sol.

⁴ En ese aspecto, seguiremos el camino propuesto por Hiki y Okada (2011) para la versión japonesa de “El viento norte y el sol”, a la cual someten a una serie de ajustes para obtener lo que ellos denominan una versión “panfónica” del texto (es decir, una que contenga ejemplos de todos los fonemas y de los principales alófonos que se utilizan en japonés).

y su correspondiente transcripción fonémica es:

el 'biento 'no te i el 'sol por'fiaban sob e 'kual de 'eAos 'e a el 'mas 'fuerte | kuando aθer'to a pa'sar un bia'xero em'buelto en 'antja 'kapa || kombi'nie on en ke kien 'antes lo'grara obli'gar al bia'xero a ki'tarse la 'kapa se'ria konside'rado 'mas pode'roso || el 'biento 'norte so'plo kon 'gran 'furia | pero 'kuanto 'mas so'plaba 'mas se arebu'xaba en su 'kapa el bia'xero || por 'fin el 'biento 'norte abando'no la em'presa || en'tonθes bri'lo el 'sol kon ar'dor | e inne'diata'mente se despo'xo de su 'kapa el bia'xero | por lo ke el 'biento 'norte 'ubo de rekon'o'θer la superiori'dad del 'sol ||

Haciéndole algunos cambios al texto, que de cualquier modo no modifican la esencia del mismo, la fábula podría quedar escrita del siguiente modo:

El viento norte y el sol discutían sobre cuál de ellos era el más fuerte, cuando pasó un extraño viajero envuelto en unas ropas muy abrigadas. Conviniéron en que quien antes lograra obligarlo al transeúnte a quitarse el abrigo sería considerado más poderoso. El viento (...) sopló con gran furia, pero cuanto más soplaba, más se ceñía el hombre su ropa al cuerpo. Entonces se dio por vencido, y el sol empezó a brillar con mucha fuerza. Inmediatamente el viajero se despojó de su abrigo; y así ya quedó claro que el sol tenía superioridad respecto del viento.⁵

con lo cual su transcripción fonémica sería la siguiente:

el 'biento 'norte i el 'sol disku'tian sobre 'kual de 'eAos 'era el 'mas 'fuerte | kuando pa'so un eks'trapo bia'xero em'buelto en 'unas 'ropas mui abri'gadas || kombi'nieron en ke kien 'antes lo'grara obli'garlo al transe'unte a ki'tarse el a'brigo se'ria konside'rado 'mas pode'roso || el 'biento so'plo kon 'gran 'furia | pero 'kuanto 'mas so'plaba 'mas se θe'nia el 'ombre su 'ropa al 'kuerpo || en'tonθes se 'dio por ben'θido | i el 'sol empe'θo a bri'la kon 'mutja 'fuerθa || inne'diata'mente el bia'xero se despo'xo de su a'brigo | i as'i 'ajke'do 'klaro ke el 'sol te'nia superiori'dad res'pekto del 'biento ||

5 En esta versión hemos subrayado las palabras nuevas que han sido incorporadas, y que reemplazan a otras que están en la versión original. Los puntos suspensivos indican que la palabra “norte” fue suprimida en una de sus apariciones.

Las dos transcripciones llevadas a cabo utilizando los símbolos del AFI se escribieron empleando los siguientes 24 fonemas: /a/, /e/, /i/, /o/, /u/, /p/, /t/, /k/, /b/, /d/, /g/, /tʃ/, /j/, /f/, /s/, /x/, /m/, /n/, /ɲ/, /r/, /r/, /l/, /θ/ y /ʎ/. Si bien los últimos dos son inexistentes para la mayor parte de los hablantes de la lengua española,⁶ hemos preferido incluirlos para tener una versión “diasistémica” de los textos, que incluya al “seseo” y al “yeísmo” como casos particulares.

En el cuadro 1 pueden verse los principales estadísticos descriptivos para las dos versiones del texto. Ambas tienen 97 palabras, pero la cantidad de fonemas es algo mayor en la versión alternativa que en la original (439 contra 428). Esto se explica básicamente porque la versión alternativa tiene 9 tipos más de palabras que la original, y toda la diferencia está en las denominadas “palabras léxicas” (es decir, sustantivos, adjetivos, verbos, adverbios) y no en las “partículas” (artículos, preposiciones, conjunciones, etc.).

Como las primeras son en general más largas que las segundas, esto hace que la extensión promedio de las palabras del texto alternativo supere levemente a las del original (4,52 versus 4,41 fonemas por palabra).

1. Estadísticos descriptivos de “El viento norte y el sol”

| Concepto | Versión original | Versión alternativa |
|------------------------|------------------|---------------------|
| Palabras (ocurrencias) | 97 | 97 |
| Palabras léxicas | 54 | 55 |
| Partículas | 43 | 42 |
| Palabras (tipos) | 60 | 69 |
| Palabras léxicas | 40 | 49 |
| Partículas | 20 | 20 |
| Fonemas (ocurrencias) | 428 | 439 |
| Fonemas (tipos) | 22 | 24 |

6 En efecto, la distinción entre los fonemas /θ/y/s/ está confinada casi exclusivamente al ámbito peninsular español, en tanto que la distinción entre /j/y/ʎ/, que hasta hace unas décadas era también común en España, parece circunscribirse hoy en día en ese país a ámbitos rurales o muy tradicionales. Esta última distinción, sin embargo, se conserva aún con fuerza en algunos países sudamericanos (especialmente Bolivia y Paraguay). Sobre estos temas, véase Hualde y Colina (2014: 283-295).

Otro elemento que influye en los estadísticos descriptivos del cuadro 1 es que, al igual que lo que Deterding (2006) señala para la versión inglesa de “El viento norte y el sol”, en la versión original española hay una gran repetición de sustantivos: “viento”, “norte”, “viajero” y “capa” aparecen cuatro veces cada uno, y “sol” lo hace tres veces. En la versión alternativa que estamos proponiendo, varias de esas repeticiones han sido suprimidas, y solo quedan “viento” y “sol” (tres veces cada uno), y “abrigo” y “viajero” (dos veces).⁷

Otra característica muy significativa de la nueva versión propuesta es que incluye palabras que ejemplifican el uso de dos fonemas que faltan en la versión original, y que son /ɲ/ y /j/.⁸ El primero de ellos aparece ahora en dos palabras: “extraño” y “ceñía”; en tanto que para ilustrar la pronunciación del segundo hemos incluido la palabra “ya”.

3. ILUSTRACIÓN DE VARIANTES FONÉTICAS

Uno de los usos que tiene el texto de “El viento norte y el sol”, en los distintos idiomas en los que se lo utiliza, es capturar las diferencias de pronunciación que puedan producirse entre las personas que lo leen. En ese sentido, la ausencia de /ɲ/ y /j/ en la versión original española puede ser un problema para captar algunas de esas diferencias, ya que este último fonema suele presentar pronunciaciones bastante disímiles entre hablantes de distintas regiones,⁹ en tanto que el primero también está sujeto a variaciones, relacionadas con su posible despatalización y reemplazo por la combinación /ni/.¹⁰

7 La palabra “norte”, por ejemplo, solo aparece ahora una vez, al principio del texto, y después se la omite por considerársela superflua. En cuanto a “viento”, también desaparece una vez, aprovechando la posibilidad que ofrece el español de utilizar sujetos tácitos en los enunciados. La palabra “viajero”, por su parte, alterna en la nueva versión con “hombre” y “transeúnte”, en tanto que “capa” ha sido reemplazada alternativamente por “abrigo”, “ropa” y “ropas”.

8 Nótese que esta última ausencia no es un problema si se quiere ilustrar la pronunciación de una variedad yeísta, ya que de todos modos en la versión original figuran las palabras “ellos” y “brilló”.

9 Estas variaciones van desde posibles pronunciaciones africadas ([tʃ], [dʒ]) hasta fenómenos de asibilación ([ʃ], [ʒ]) y vocalización ([j]). En este último caso, el fonema / / puede inclusive ser analizado como un alófono de /i/. Véase Hualde y Colina (2014: 162-169).

10 Véase, por ejemplo, Colantoni y Hualde (2013).

Además de estos fenómenos, existen en la lengua española otras circunstancias en las cuales puede aparecer variación en la pronunciación. En Coloma (2013), por ejemplo, hay una lista de ocho características fonéticas que sirven para delimitar zonas, y para definir “continuos dialectales” que ligan a dichas zonas cuando las mismas se encuentran contiguas geográficamente. La lista de dichas características incluye a los casos ya mencionados de seseo y yeísmo, y a seis rasgos fonéticos más, que son:

Aspiración de /s/: Uso de una consonante fricativa aspirada glotal [h] como alófono del fonema /s/, en coda silábica.

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

Asibilación de /j/: Articulación del fonema /j/ a través de una consonante asibilada postalveolar, que puede ser sonora [ʒ] o sorda [ʃ].

Asibilación de /r/: Articulación del fonema /r/ a través de una consonante fricativa asibilada [ʒ], en vez de utilizar el sonido vibrante [r].

Velarización de /n/: Uso de la consonante nasal velar [ŋ] como alófono del fonema /n/, no solo cuando dicho fonema precede a otra consonante velar sino también en posición final de palabra.

Desafricación de /tʃ/: Articulación del fonema /tʃ/ a través de la consonante fricativa postalveolar sorda [ʃ], en reemplazo de la consonante africada post-alveolar sorda [tʃ].

A estas características podemos agregarle muchas otras, que tienen una difusión mayor o menor en el ámbito de la lengua castellana. Entre ellas mencionaremos la espirantización restringida de las consonantes obstruyentes sonoras (que implica utilizar los alófonos oclusivos de /b/, /d/ y /g/ en algunos casos en los cuales la mayoría de los hablantes emplean los alófonos fricativos o aproximantes [β], [ð] y [ɣ]),¹¹ la elisión del fonema /d/ en ciertas posiciones (finales de palabra, e intervocálicas en sílabas no acentuadas) y la elisión del fonema /r/ en final de palabra.¹²

Con todo este conjunto de variaciones hemos preparado dos pares de transcripciones fonéticas de “El viento norte y el sol”, que corresponden a la versión original y a la versión alternativa presentada en la sección anterior. Para cada una

11 Véase Real Academia Española (2011: 144-162).

12 Sobre la distribución geográfica de estos fenómenos, véase Moreno Fernández (2009) y las referencias bibliográficas allí citadas.

de ellas hemos elaborado una transcripción “conservadora” (que no presenta ninguna de las características fonéticas mencionadas más arriba) y una transcripción “innovadora” (que las incluye a todas). Estas transcripciones no intentan representar ningún dialecto particular dentro del mundo hispánico, si bien la transcripción conservadora podría asociarse con el habla de ciertas zonas tradicionales del norte de España.¹³ Las transcripciones en cuestión son las siguientes:

Transcripción fonética conservadora (versión original)

el 'βjento 'norte jel 'sol por'fjaβan soβre 'kwal 'de'lo 'seraeɫ 'mas 'fwerte | kwandoα θer'toα pa'sarum bja'xeroem 'bweltoε 'nantʃa 'kapa || kombi'njero neɲ ke kje 'nantez lo'γarao βli'γa al βja'xeroα ki'tarse la 'kapa se'ria konsiðe'raðo 'mas poðe'roso || el 'βjento 'norte so'plo koɲ 'gram 'furja | pero 'kwanto 'maso 'plaβa 'maseα reβu'xaβaeɲ su 'kapaεɫ bja'xero || por 'finel 'βjento 'norɽeα βando'no laεm'presa || en'tonθez βri'ʎoεɫ 'sol ko nar'dor | ejnme'ðjata'mente se ðespo'xo ðe su 'kapaεɫ bja'xero | por lo kel 'βjento 'norte 'uβo ðe rekono'θer la superjori'ðaðel 'sol ||

Transcripción fonética innovadora (versión original)

el 'bjento 'no te jel 'sol por'fjaβaɲ soβre 'kwal 'dezo 'heraeɫ 'mah 'fwerte | kwandoα ser'toα pa'sawɲ bja'heroem 'bweltoε 'nanʃa 'kapa || kombi'njero neɲ ke kje 'nanteɲ lo'γrarao βli'γal bja'heroα ki'tarse la 'kapa se'ria konsiðe'rao 'mah poðe'roso || el 'bjento 'norte so'plo koɲ 'gran 'furja | pero 'kwanto 'maso 'plaβa 'maseα zeβu'haβaeɲ su 'kapaεɫ bja'hero || por 'finel 'bjento 'norɽeα βando'no laεm'presa || en'tonseɲ bri'zoεɫ 'sol ko nar'do | ejnme'ðjata'mente se ðehpo'hoε su 'kapaεɫ bja'he o | por lo kel 'bjento 'norte 'uβoε zekono'se la superjori'ðaeɫ 'sol ||

Transcripción fonética conservadora (versión alternativa)

el 'βjento 'norte jel 'sol disku'tian soβre 'kwal 'de'lo 'seraeɫ 'mas 'fwerte | kwando pa'sow neks'traɲo βja'xeroem 'bweltoε 'nunaz 'ropaz mu jaβri'γaðas || kombi'njero neɲ ke kje 'nantez lo'γra ao βli'garloal transe'unteα ki'tarse la'βri'γo se'ria konsiðe' aðo 'mas poðe' oso || el 'βjento so'plo koɲ 'gram 'furja | pe o 'kwanto 'maso 'plaβa 'mase θe'ɲiaε 'lombre su 'ropal 'kwerpo || en'tonθe se 'ðjo por βen'θiðo | jel 'so lempe'θoα

¹³ La transcripción conservadora de la versión original es de hecho bastante similar a la que aparece en Martínez, Fernández y Carrera (2003).

βri'la kon 'mutfa 'fwerθa || inme'ðjata'mentel βja'xero se ðespo'xo ðe swa'βriʎo | ja'si ja ke'ðo 'klaro kel 'sol te'nia superjori'ðað res'pekto ðel 'βjento ||

Transcripción fonética innovadora (versión alternativa)

el 'bjento 'norte jel 'sol dihku'tian soβre 'kwal 'dezo 'heraɣl 'mah 'fwerte | kwando pa'sow neh'tranjo βja'herqem 'bweltɔe 'nunah 'zopah mu jabri'ʎah || kom-bi'njero neɲ ke kje 'nanteh lo'ʎrafaɔ βli'ʎaloal transe'untea ki'tarse la'βriʎo se'ria konsiðe'raɔ 'mah poðe'roso || el 'bjento so'plo koɲ 'graɲ 'furja | pero 'kwanto 'maso 'plaβa 'mase se'ɲiaɕ 'lombre su 'zopal 'kwerpo || en'tonse se 'ðjo por ben'sio | jel 'so lempe'sɔa βri'za koɲ 'muɟa 'fwersa || inme'ðjata'mentel bja'hero se ðehpo'hɔe swa'βriʎo | ja'si za ke'ðo 'klaro kel 'sol te'nia superjori'ða zeh'pektɔel 'bjento ||

Si ahora comparamos la transcripción conservadora y la innovadora para las dos versiones de nuestra fábula, veremos que en la versión original las dos transcripciones tienen 42 diferencias entre sí, en tanto que en la versión alternativa tienen 45 diferencias. Las mismas son las que aparecen expuestas en el cuadro 2, en el cual han sido clasificadas según la característica que origina cada diferencia.

2. Diferencias entre transcripciones conservadoras e innovadoras

| Diferencia | Versión original | Versión alternativa |
|-----------------------------|------------------|---------------------|
| Seseo | 3 | 5 |
| Yeísmo | 2 | 2 |
| Aspiración de /s/ | 6 | 11 |
| Aspiración de /x/ | 6 | 3 |
| Asibilación de /j/ | 2 | 3 |
| Velarización de /n/ | 4 | 3 |
| Desafricación de /t / | 1 | 1 |
| Asibilación de /r/ | 2 | 2 |
| Elisión de /d/ | 4 | 7 |
| Elisión de /r/ | 4 | 2 |
| Despalatalización de /ɲ/ | 0 | 1 |
| Espirantización restringida | 8 | 5 |
| Total | 42 | 45 |

Tal como puede observarse, la versión alternativa tiene más ejemplos de variación originados en la presencia de seseo, aspiración de /s/, asibilación de /j/ y elisión de /d/, y menos ejemplos en lo que se refiere a aspiración de /x/, velarización de /n/, elisión de /r/ y espirantización restringida. Tiene además la ventaja de tener un caso que puede presentar despalatalización del fonema /ɲ/ (el correspondiente a la palabra “extraño”), cosa que no existe en la versión original porque en ella no hay ejemplos de pronunciación de dicho fonema.

4. BALANCE FONÉTICO

Otro modo de evaluar la capacidad de un texto para representar los sonidos de un idioma es estudiar lo que se conoce como su “balance fonético”, que es un concepto que tiene que ver con las proporciones de los diferentes sonidos. Puede afirmarse que un texto está fonéticamente balanceado si todos los fonemas del idioma están reflejados en él, si la frecuencia con la que aparecen es cercana a la del lenguaje natural, si incluye el número más reducido de palabras que puede cumplir con los requisitos antes mencionados, y si dichas palabras son todas de uso corriente.¹⁴

Claramente, la versión original española de “El viento norte y el sol” no cumple con uno de los requisitos mencionados en el párrafo anterior, ya que le faltan ejemplos de ocurrencia de dos fonemas. En cuanto a su extensión, la misma parece en cambio ser adecuada, ya que es relativamente breve para un texto que busca reflejar todos los fonemas del español de manera balanceada. En lo que se refiere al léxico de la versión original de la fábula, puede decirse que el mismo resulta un poco anticuado para un hablante del siglo XXI, pero de cualquier modo todas las palabras que contiene deberían ser comprendidas por dicho hablante. A efectos de evitar algunas expresiones no tan usuales, sin embargo, en la versión alternativa eliminamos “porfiaban”, “ancha capa”, “arrebujaba” y “ardor”, y las reemplazamos por “discutían”, “ropas muy abrigadas” “ceñía” y “mucho fuerza”.¹⁵

14 Sobre este punto, véase Jesus, Valente y Hall (2015).

15 De paso, al llevar a cabo estos reemplazos, aprovechamos para incluir un ejemplo de /s/ en coda silábica antes de pausa (que no existe en la versión original de “El viento norte y el sol”), así como realizaciones de los fonemas /θ/, /ɲ/ y /tʃ/.

3. Distribuciones de frecuencia de los fonemas

| Fonema | Corpus EFE | | VNS original | | VNS alternativa | |
|--------|------------|---------|--------------|---------|-----------------|---------|
| | % | Ranking | % | Ranking | % | Ranking |
| a | 12.89 | 1 | 12.62 | 2 | 10.93 | 3 |
| e | 12.74 | 2 | 14.25 | 1 | 13.90 | 1 |
| o | 9.32 | 3 | 11.92 | 3 | 11.16 | 2 |
| s | 7.33 | 4 | 5.84 | 6 | 7.52 | 4 |
| i | 7.25 | 5 | 5.61 | 7.5 | 7.29 | 5 |
| n | 7.09 | 6 | 7.48 | 5 | 6.15 | 7 |
| r | 6.19 | 7 | 7.94 | 4 | 7.06 | 6 |
| l | 5.46 | 8 | 5.61 | 7.5 | 5.24 | 8 |
| d | 5.42 | 9 | 3.27 | 13 | 3.64 | 12.5 |
| t | 4.31 | 10 | 3.97 | 10 | 4.10 | 9 |
| k | 3.80 | 11 | 3.74 | 11 | 3.87 | 10.5 |
| u | 3.04 | 12 | 2.80 | 14 | 3.87 | 10.5 |
| m | 2.76 | 13 | 2.10 | 15 | 2.73 | 15 |
| p | 2.73 | 14 | 3.50 | 12 | 2.96 | 14 |
| b | 2.55 | 15 | 4.44 | 9 | 3.64 | 12.5 |
| θ | 2.00 | 16 | 0.70 | 18.5 | 1.14 | 17 |
| g | 1.04 | 17 | 0.70 | 18.5 | 1.37 | 16 |
| r | 0.99 | 18 | 0.47 | 20.5 | 0.68 | 19 |
| f | 0.92 | 19 | 0.93 | 17 | 0.68 | 19 |
| x | 0.77 | 20 | 1.40 | 16 | 0.68 | 19 |
| ʎ | 0.53 | 21 | 0.47 | 20.5 | 0.46 | 21.5 |
| ɟ | 0.38 | 22 | 0.00 | 23.5 | 0.23 | 23.5 |
| ɲ | 0.31 | 23 | 0.00 | 23.5 | 0.46 | 21.5 |
| tʃ | 0.18 | 24 | 0.23 | 22 | 0.23 | 23.5 |

A efectos de evaluar si la distribución de frecuencias de los distintos fonemas en una versión y en otra de “El viento norte y el sol” se encuentra cerca de la que se supone que existe en la lengua española usual, resulta necesario aproximar esa última frecuencia utilizando algún estudio estadístico. De los que se encuentran disponibles, optamos por uno de los que aparecen en Moreno Sandoval y otros (2006), que proviene de un análisis basado en un gran número de casos (480.000 palabras y 2.511.856 fonemas) extraído de un corpus escrito de la agencia de noticias EFE.¹⁶

En el cuadro 3 aparecen los resultados de calcular las frecuencias relativas de los 24 fonemas españoles en las versiones original y alternativa de “El viento norte y el sol” (VNS), junto con el que surge del corpus de la agencia EFE. En cada caso se reporta también el respectivo *ranking* de fonemas, y todo ello aparece a su vez representado en el gráfico 1, en el cual el orden de los fonemas corresponde al del corpus EFE.

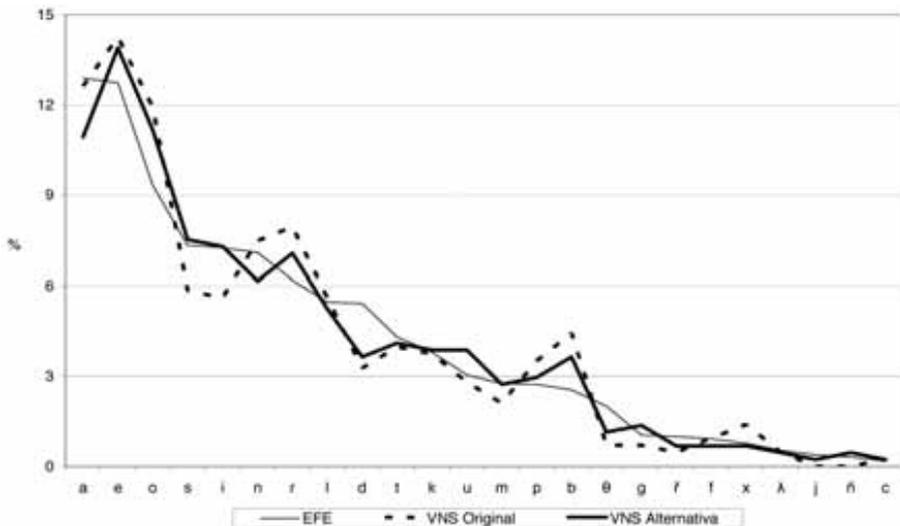


Gráfico 1: Distribuciones de frecuencia de los fonemas.

16 Las otras alternativas disponibles (por ejemplo, Alarcos, 1965:159-162; Quilis y Esgueva, 1980; Guirao y García Jurado, 1993; Pérez, 2003; González y Mejía, 2011) son más antiguas, o están basadas en menos observaciones, o corresponden a variedades del español que presentan seseo y yeísmo (y no registran por lo tanto frecuencias para los fonemas /q/ y /ʎ/). Moreno Sandoval y otros (2006) presentan también una distribución alternativa basada en un corpus oral de 1.244.411 fonemas, pero preferimos usar el corpus de la agencia EFE porque es más extenso y está basado en textos escritos.

Una medida relativamente directa de la semejanza entre dos variables (por ejemplo, dos distribuciones de frecuencia) es su coeficiente de correlación “r” (llamado también “coeficiente de Pearson”). En este caso, si calculamos dicho coeficiente para las distribuciones correspondientes al corpus EFE y a las dos versiones de VNS, vemos que la correlación obtenida es muy alta (“r = 0,9634” para EFE versus VNS original, y “r = 0,9755” para EFE versus VNS alternativa). Lo mismo ocurre si computamos el denominado “coeficiente de correlación de rangos” o “coeficiente de Spearman”, que no es otra cosa que el coeficiente de correlación entre los respectivos *rankings* de frecuencias. Dicho indicador nos da un valor “r = 0,9443” para la correlación entre el corpus EFE y la versión original de VNS, y “r = 0,9813” para la correlación entre dicho corpus y la versión alternativa.¹⁷

Estos coeficientes de correlación tan elevados pueden leerse como una indicación de que tanto la versión original como la alternativa están fonéticamente balanceadas, en el sentido de que la frecuencia con la que aparecen en ellas los distintos fonemas es muy cercana a la del lenguaje natural. Este tipo de coeficientes promedio, sin embargo, pueden a veces ocultar problemas que tienen impacto sobre algunos fonemas en particular. Para descubrir dichos problemas, Jesus, Valente y Hall (2015) han utilizado, en su estudio sobre la versión portuguesa de “El viento norte y el sol”, un método creado por Bland y Altman (1986) para evaluar la semejanza entre dos muestras. Dicho método se basa en calcular los siguientes “estadísticos Z”:

$$Z(VNSO)_i = \frac{(f(VNSO)_i - f(EFE)_i)^2}{f(EFE)_i} ; \quad Z(VNSA)_i = \frac{(f(VNSA)_i - f(EFE)_i)^2}{f(EFE)_i} ;$$

donde $f(EFE)_i$, $f(VNSO)_i$ y $f(VNSA)_i$ son las frecuencias para los distintos fonemas individuales que corresponden a las distribuciones de los mismos en el corpus EFE, y en las versiones original y alternativa de “El viento norte y el sol”. Los estadísticos Z así calculados tienen una distribución ji cuadrado con un grado de libertad, y son significativamente distintos de cero desde un punto de vista estadístico (con una probabilidad de 5%) si su valor es mayor que 0.39.

En el gráfico 2 aparecen representados los estadísticos Z de cada uno de los 24 fonemas españoles, tanto para la versión original como para la alternativa, así como también una

17 Para una explicación del cálculo de los coeficientes de correlación, véase Hernández Campoy y Almeida (2005: 237-240).

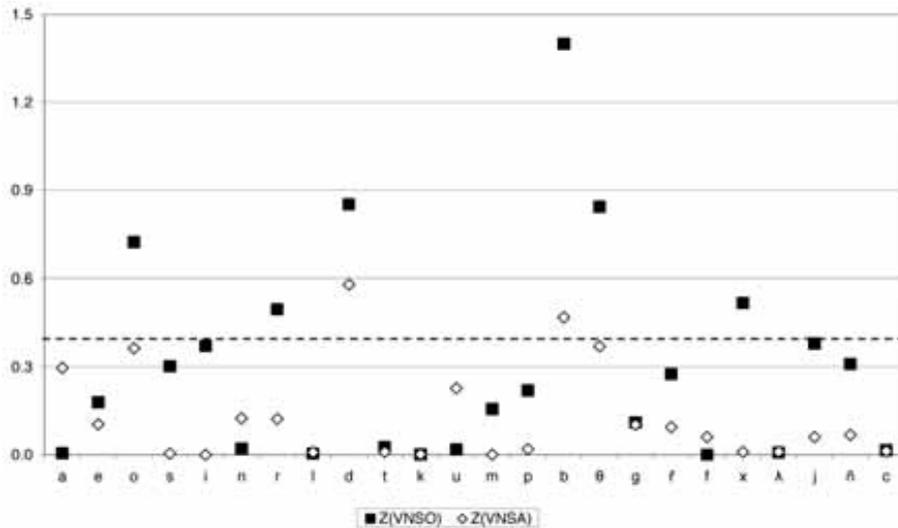


Gráfico 2: Estadísticos Z para las versiones original y alternativa.

línea de puntos que corresponde al valor por encima del cual tales estadísticos son significativamente distintos de cero. En dicho gráfico puede observarse que, para la versión original de VNS, existen seis fonemas cuyos estadísticos Z son significativamente distintos de cero, lo cual indica que las frecuencias de dichos fonemas están sobrerrepresentadas o subrepresentadas en ese texto. Tales fonemas son /o/, /r/, /b/ y /x/ (sobrerrepresentados), y /d/ y /θ/ (subrepresentados). En el caso de la versión alternativa, en cambio, los únicos dos fonemas para los cuales Z(VNSA) toma un valor que se ubica por encima de la línea del 5% de significación estadística son /b/ y /d/, y aún así el grado de sobre o subrepresentación que dichos fonemas tienen resulta ser menor que en la versión original.

5. ANÁLISIS DE REGRESIÓN ESTADÍSTICA

Otro modo de comparar las distribuciones de frecuencia de las versiones original y alternativa de “El viento norte y el sol” consiste en estimar paraméricamente dichas distribuciones, y luego testear si las mismas son significativamente distintas de la distribución real de frecuencias en el lenguaje natural. Si no lo son, puede entonces afirmarse que tales distribuciones tienen una forma similar a la verdadera distribución de frecuencias del español (que aquí estamos aproximando a través de la que se verifica en el corpus de la agencia EFE).

Para llevar a cabo esas comparaciones, resulta necesario correr una serie de regresiones, utilizando una forma funcional que resulte apropiada y alguna variable que se suponga que determina las frecuencias bajo estudio. En la literatura sobre lingüística cuantitativa, la función más común para este tipo de análisis es la función de distribución de Zipf,¹⁸ que supone que los fonemas siguen una distribución como esta:

$$f = a \times r^b \Rightarrow \log(f) = \log(a) + b \times \log(r) \quad ;$$

donde f es la frecuencia de cada fonema, r es el *ranking* de dicho fonema, y a y b son parámetros.

En el cuadro 4 aparecen los resultados de correr sendas regresiones para las tresseries de frecuencias de los fonemas (las correspondientes al corpus EFE, a VNS original y a VNS alternativa) como funciones de sus respectivos *rankings*. Para llevarlas a cabo, primero transformamos a f y a r en logaritmos naturales, y después corrimos las regresiones transformadas utilizando el método de mínimos cuadrados ordinarios. Una vez hallados los resultados, calculamos el antilogaritmo del coeficiente obtenido para la constante de cada regresión, el cual no es otra cosa que el coeficiente estimado para el parámetro a correspondiente a dicha regresión.

Debe aclararse además que, para poder llevar a cabo la regresión correspondiente a la ecuación de VNS original, fue necesario ajustar la variable f para que las observaciones de los fonemas / η / y / j / tuvieran un valor positivo, ya que el logaritmo natural de cero es un número inexistente. Se supuso entonces que la probabilidad conjunta estimada para los fonemas no observados era igual a la probabilidad del fonema de menor frecuencia (que en este caso es / t /), y dicha probabilidad fue dividida en partes iguales entre / η / y / j . Luego de eso, las frecuencias de todas las observaciones fueron ajustadas para que la suma total de las mismas fuera igual al 100%.¹⁹

En el gráfico 3 hemos representado los resultados del cuadro 4 a través de dos líneas que le asignan a cada fonema la frecuencia estimada por la regresión

18 Para una explicación de la forma de esta función, véase Rodríguez Luna (2012: 129-141).

19 La técnica utilizada para realizar estos ajustes es una versión simplificada de la denominada “estimación de Good y Turing”. Para una explicación de este concepto, véase Sampson (2001: 94-108).

4. Resultados de las regresiones de la función de distribución de Zipf

| Concepto | Coefficiente | Desvío Std. | Estad t | Valor p |
|--------------------------|--------------|-------------|---------|---------|
| Ecuación corpus EFE | | | | |
| Parámetro a | 42.7134 | 4.4780 | 9.5384 | 0.0000 |
| Parámetro b | -1.2573 | 0.1624 | -7.7399 | 0.0000 |
| R cuadrado | 0.7314 | | | |
| Ecuación VNS original | | | | |
| Parámetro a | 57.8964 | 6.8582 | 8.4419 | 0.0000 |
| Parámetro b | -1.4565 | 0.1984 | -7.3415 | 0.0000 |
| R cuadrado | 0.7101 | | | |
| Ecuación VNS alternativa | | | | |
| Parámetro a | 45.3629 | 4.7791 | 9.4919 | 0.0000 |
| Parámetro b | -1.2980 | 0.1658 | -7.8268 | 0.0000 |
| R cuadrado | 0.7358 | | | |

para su correspondiente *ranking*. Los fonemas fueron a su vez ordenados de acuerdo con su *ranking* en cada una de las dos series (VNS original y VNS alternativa), y sus frecuencias reales también están graficadas en la figura a través de los puntos marcados como *VNSO* y *VNSA*. Vemos que en ambos casos la distribución de Zipf ($F_z(VNSO)$ y $F_z(VNSA)$) predice que la frecuencia de cada fonema irá decreciendo, siguiendo una curva cuya forma es “convexa al origen”.

La curva predicha por $F_z(VNSA)$, sin embargo, logra un ajuste de los datos que es mejor que el que consigue la curva predicha por $F_z(VNSO)$. Esto puede verse en el cuadro 4 a través de los coeficientes “R cuadrado” reportados para las dos regresiones, que son respectivamente iguales a 0,7358 y a 0,7101.²⁰ El coeficiente R cuadrado de la regresión que corresponde a las frecuencias de la versión alternativa de “El viento norte y el sol” es inclusive mayor que el que corresponde a las frecuencias de los distintos fonemas en el corpus de la agencia EFE, el cual es igual a 0,7314.

20 Para una explicación de este concepto, véase Hernández Campoy y Almeida (2005: 240-247).

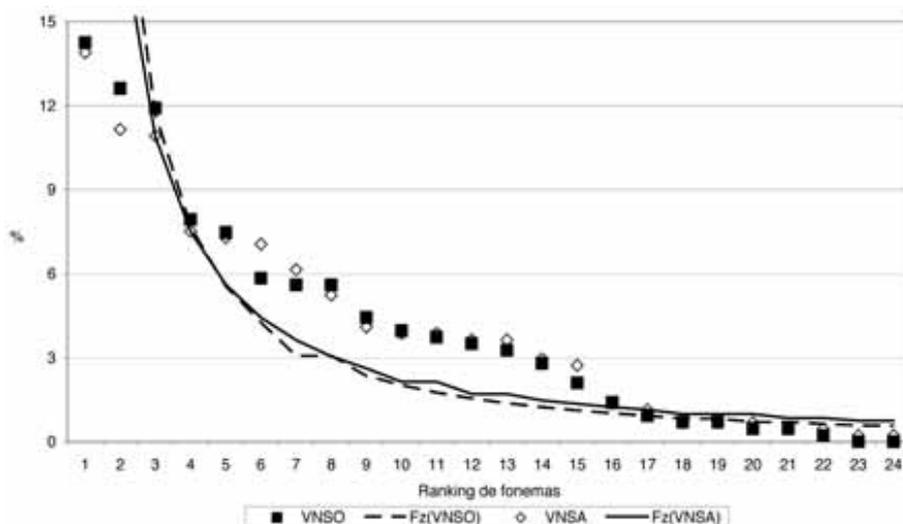


Gráfico 3: Distribuciones de Zipf para VNS original y alternativa.

Una última comparación posible entre las versiones original y alternativa de VNS consiste en un test estadístico que calcule la probabilidad conjunta de que los parámetros de las distribuciones de Zipf estimados en las regresiones coincidan con los de la distribución de las frecuencias del corpus EFE. Para llevar a cabo ese test, debe utilizarse un estadístico ji cuadrado de una hipótesis nula que suponga alternativamente que se cumple que:

$a(EFE) = a(VNSO)$, $b(EFE) = b(VNSO)$; $a(EFE) = a(VNSA)$, $b(EFE) = b(VNSA)$.
 donde $a(EFE)$, $a(VNSO)$, $a(VNSA)$, $b(EFE)$, $b(VNSO)$ y $b(VNSA)$ son los valores estimados para los parámetros a y b en las regresiones correspondientes a las frecuencias del corpus EFE, y a las frecuencias de las versiones original y alternativa de "El viento norte y el sol".

El resultado de esos tests es que la probabilidad de que la hipótesis nula sea verdadera para la versión original de VNS es igual a 0,5693, en tanto que la probabilidad de que sea verdadera para la versión alternativa es igual a 0,9701.²¹ Tal

²¹ Estos números, al igual que los que provienen de los análisis de regresión, fueron calculados utilizando el programa informático EViews 3.1.

como puede observarse, dichos resultados muestran una preferencia muy clara por la distribución de frecuencias proveniente de la versión alternativa de “El viento norte y el sol”, cuyos parámetros son prácticamente los mismos que los que surgen de correr una regresión con las frecuencias del corpus EFE. No es ese el caso, en cambio, de las frecuencias que exhiben los distintos fonemas en la versión original de VNS, cuya regresión genera una distribución de Zipf que es sustancialmente distinta de la del corpus EFE.

6. COMENTARIOS FINALES

La principal conclusión que puede obtenerse de los distintos análisis llevados a cabo en el presente trabajo es que la versión original de “El viento norte y el sol” en español, empleada por la IPA para ilustrar la pronunciación de nuestro idioma, adolece de cuatro fallas principales:

- a) No tiene ejemplos de uso de los fonemas /ɲ/ y /j/.
- b) Tiene una tasa de repetición muy elevada de algunos sustantivos.
- c) Tiene una distribución de frecuencias en la cual algunos fonemas están sobre o subrepresentados en relación con la frecuencia que tienen en el lenguaje natural.
- d) Esto último genera que, al representar dichas frecuencias utilizando una función de distribución de Zipf, surjan parámetros que son significativamente distintos de los que exhibe la misma distribución en el lenguaje natural.

La versión alternativa que estamos proponiendo, que tiene el mismo número de palabras que la original y que difiere de ella en solamente 35 vocablos, soluciona todos esos problemas, ya que incluye dos ejemplos de /ɲ/ y uno de /j/, tiene una tasa de repetición mucho menor, solo tiene un fonema sobrerrepresentado y uno subrepresentado (y en proporciones más bajas que las del texto original), y los parámetros que exhibe su función de distribución de Zipf no son significativamente distintos de los que aparecen en la función correspondiente al lenguaje natural (calculada en base a las frecuencias que surgen de tomar como base un corpus de la agencia de noticias EFE).

Por todo esto puede concluirse entonces que sería bueno sustituir la versión original de “El viento norte y el sol” por esta otra versión alternativa, tanto en los estudios sobre pronunciación en los que dicho texto vaya a ser utilizado

como en las próximas ilustraciones del AFI sobre variedades del idioma español que puedan aparecer en el *Journal of the International Phonetic Association*.

REFERENCIAS BIBLIOGRÁFICAS

- Alarcos, Emilio (1965). *Fonología española*. Madrid: Gredos.
- Bland, John y Douglas Altman (1986). "Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement", *Lancet* 1: 307-310.
- Bowden, John y John Hajek (1996). "Illustrations of the IPA: Taba", *Journal of the International Phonetic Association* 26(1): 55-57.
- Bowern, Claire, Joyce McDonough y Katherine Kelliher (2012). "Illustrations of the IPA: Bardi", *Journal of the International Phonetic Association* 42(3): 333-351.
- Carlson, Barry y John Esling (2000). "Illustrations of the IPA: Spokane", *Journal of the International Phonetic Association* 30(1): 97-102.
- Colantoni, Laura (2008). "Variación micro y macro fonética en español", *Estudios de Fonética Experimental* 17: 65-104.
- Colantoni, Laura y José Hualde (2013). "Variación fonológica en el español de la Argentina", en Colantoni, L. y C. Rodríguez Louro (eds.): *Perspectivas teóricas y experimentales sobre el español de la Argentina*, 21-35. Madrid: Iberoamericana Vervuert.
- Coloma, Germán (2013). "Los continuos dialectales de la lengua española", *Zeitschrift für romanische Philologie* 129(2): 366-382.
- Coloma, Germán (2014). "La existencia de correlación negativa entre distintos aspectos de la complejidad de los idiomas", Documento de Trabajo Nro. 536. Buenos Aires: Universidad del CEMA.
- Connell, Bruce, Firmin Ahoua y Dafydd Gibbon (2002). "Illustrations of the IPA: Ega", *Journal of the International Phonetic Association* 32(1): 99-104.
- Deterding, David (2006). "The North Wind versus a Wolf: Short Texts for the Description and Measurement of English Pronunciation", *Journal of the International Phonetic Association* 36(2): 187-196.
- González, María y Jorge Mejía (2011). "Frecuencia fonemática del español de Colombia", *Forma y función* 24(2): 69-102.
- Guirao, Miguelina y María García Jurado (1993). *Estudio estadístico del español*. Buenos Aires: Conicet.

- Hernández Campoy, Juan y Manuel Almeida (2005). *Metodología de la investigación sociolingüística*. Málaga: Comares.
- Hiki, Shizuo y Hideo Okada (2011). "A Panphonic Version of the Recording Text of 'The North Wind and the Sun' for the Illustration of the IPA of Japanese (Tokyo Dialect) Consonants", mimeo. Tokyo: Waseda University.
- Hualde, José y Sonia Colina (2014). *Los sonidos del español*. Nueva York: Cambridge University Press.
- IPA (1912). *Principles of the International Phonetic Association*. París: Asociación Fonética Internacional.
- IPA (1949). *Principles of the International Phonetic Association*. Londres: University College.
- IPA (1999). *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.
- Jesus, Luis, Ana Valente y Andreia Hall (2015). "Is the Portuguese Version of the Passage 'The North Wind and the Sun' Phonetically Balanced?", *Journal of the International Phonetic Association* 45(1): 1-11.
- Martínez, Eugenio, Ana Fernández y Josefina Carrera (2003). "Illustrations of the IPA: Castilian Spanish", *Journal of the International Phonetic Association* 33(2): 255-260.
- Monroy, Rafeel y Juan Hernández Campoy (2015). "Illustrations of the IPA: Murcian Spanish", *Journal of the International Phonetic Association* 45(2): 229-240.
- Moreno Sandoval, Antonio, Doroteo Torre, Natalia Curto y Raúl de la Torre (2006). "Inventario de frecuencias fonémicas y silábicas del castellano espontáneo y escrito", *Actas de las IV Jornadas en Tecnologías del Habla*, 77-81. Zaragoza: Universidad de Zaragoza.
- Moreno Fernández, Francisco (2009). *La lengua española en su geografía*. Madrid: Arco.
- Pérez, Hernán (2003). "Frecuencia de fonemas", *E-rthabla* 1(4).
- Quilis, Antonio y Manuel Esgueva (1980). "Frecuencia de fonemas en el español hablado", *Lingüística Española Actual* 2(1): 1-25.
- Real Academia Española (2011). *Nueva gramática de la lengua española: fonética y fonología*. Madrid: Espasa-Calpe.

- Rodríguez Luna, Manuela (2012). *Ampliación y perfeccionamiento de los métodos cuantitativos y leyes clásicas en recuperación de la información*. Valencia: Universidad Politécnica de Valencia.
- Sampson, Geoffrey (2001). *Empirical Linguistics*. Londres: Continuum.

Fecha de recepción: 1 de julio de 2015
Fecha de aceptación: 14 de julio de 2015