UNIVERSIDAD DEL CEMA
Buenos Aires
Argentina

Serie
DOCUMENTOS DE TRABAJO

Área: Lingüística y Estadística

A STATISTICAL COMPARISON BETWEEN TWO TEXTS TO ILLUSTRATE THE PHONETICS OF SPANISH

Germán Coloma

Septiembre 2017
Nro. 619

www.cema.edu.ar/publicaciones/doc_trabajo.html
UCEMA: Av. Córdoba 374, C1054AAP Buenos Aires, Argentina
ISSN 1668-4575 (impreso), ISSN 1668-4583 (en línea)
Editor: Jorge M. Streb; asistente editorial: Valeria Dowding <jae@cema.edu.ar>

A Statistical Comparison between Two Texts to Illustrate the Phonetics of Spanish

Germán Coloma*

Abstract

Following an idea proposed by Deterding (2006) for the English version of "The North Wind and the Sun", this paper compares the standard Spanish version of that fable with an alternative text which corresponds to another fable, "The Boy who Cried Wolf". The comparison is based on the phonetic features that appear in both texts, on their phonetic balance, and on the goodness of fit that they display when we compute their phoneme frequencies (and compare those frequencies with an average distribution for Spanish written texts). The conclusion is that "The Boy who Cried Wolf" seems to perform better than "The North Wind and the Sun" in all those dimensions.

Keywords: descriptive phonetics, Spanish, phonetic balance, Zipf distribution, Yule distribution.

1. Introduction

"The North Wind and the Sun" (NWS) is a fable attributed to Aesop, which has been used for more than a hundred years by the International Phonetic Association (IPA) as a "specimen" to illustrate the phonetics of many languages.[1] Spanish has been no exception to that rule, and a version of NWS can be found in Martínez, Fernández & Carrera (2003) and in Monroy & Hernández (2015), which is basically the same one that appears in IPA (1949).[2]

In several articles that illustrate the sounds of a series of non-European languages, some authors have argued against the use of NWS, mainly because they think that the plot of the story told in that text is unnatural for the speakers of those languages.[3] In

* CEMA University; Av. Córdoba 374, Buenos Aires, C1054AAP, Argentina. Telephone: 54-11-6314-3000. E-mail: gcoloma@cema.edu.ar. I thank Laura Colantoni, David Deterding, Luis Jesus and Adrian Simpson for their useful comments to a previous version of this paper. The opinions expressed in this publication are my own, and not necessarily the ones of CEMA University.

[1] See IPA (1912), IPA (1949) and IPA (1999), or the many "Illustrations of the IPA" published in the *Journal of the International Phonetic Association* since 1990.

[2] Other (slightly different) versions appear in Avelino (2017) and in Coloma (2017).

[3] See, for example, Bowern, McDonough & Kelliher (2012).

general, those authors have preferred to use alternative texts, which are supposed to be more suitable examples.[4]

In Deterding (2006), however, there is another objection against NWS, which refers to its use as an illustration of the phonetics of the English language. That objection has to do with the absence of some phonemes and allophones, and also with other problems related to rhythm and to the acoustic measurement of some vowels. As a consequence of those problems, Deterding proposed the use of an alternative text, which is an English version of another fable: "The Boy who Cried Wolf" (BCW).

The BCW text that Deterding analyzes is a substantially rewritten version of the original fable, and it is nearly twice as long as the English NWS text. In Spanish, however, there is a classical version of BCW, whose title is "*El zagal y las ovejas*". It was written by a relatively famous writer, Félix de Samaniego, who originally published it in 1781 as a part of a collection of fables.[5] That text has roughly the same extension than the NWS Spanish version.

In the following sections we will proceed to compare the relative advantages and disadvantages of NWS and BCW for the description of the phonetics of Spanish. We will first reproduce both texts and calculate a few descriptive statistics for them (section 2), and after that we will illustrate their main phonetic features and shortcomings (section 3). In section 4 we will study their phoneme frequency distributions, and finally in section 5 there will be some concluding remarks.

2. The North Wind versus the Wolf

The Spanish NWS text that appears in Martínez, Fernández & Carrera (2003), and in Monroy & Hernández (2015), is the following:

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

[4] See Bowden & Hajek (1996), Carlson & Esling (2000), Connell, Ahoua & Gibbon (2002) and Guerin & Aoyama (2009), among other illustrations of the IPA that do not use a NWS text.
[5] The text that we use here is the one that appears in Samaniego (2003:58).

sopló con gran furia, pero cuanto más soplaba, 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.

and its corresponding phonemic transcription would be this:

el 'biento 'noɾte i el 'sol poɾ'fiaban sobɾe 'kual de 'eʎos 'eɾa el 'mas 'fueɾte | kuando aθeɾ'to a pa'saɾ un bia'xeɾo em'buelto en 'antʃa 'kapa || kombi'nieɾon en ke kien 'antes lo'gɾaɾa obli'gaɾ al bia'xeɾo a ki'taɾse la 'kapa se'ɾia konside'ɾado 'mas pode'ɾoso || el 'biento 'noɾte so'plo kon 'gran 'fuɾia | peɾo 'kuanto 'mas so'plaba 'mas se arebu'xaba en su 'kapa el bia'xeɾo || poɾ 'fin el 'biento 'noɾte abando'no la em'pɾesa || en'tonθes bɾi'ʎo el 'sol kon aɾ'doɾ | e inme'diata'mente se despo'xo de su 'kapa el bia'xeɾo | poɾ lo ke el 'biento 'noɾte 'ubo de rekono'θeɾ la supeɾioɾi'dad del 'sol ||

The BCW text, which will be used as an alternative to NWS, is this:

Apacentando un joven su ganado,
gritó desde la cima de un collado:
«¡Favor!, que viene el lobo, labradores».
Estos, abandonando sus labores,
acuden prontamente,
y hallan que es una chanza solamente.
Vuelve a clamar, y temen la desgracia;
segunda vez los burla. ¡Linda gracia!
¿Pero qué sucedió la vez tercera?
Que vino en realidad la hambrienta fiera.
Entonces el zagal se desgañita,
y por más que patea, llora y grita,
no se mueve la gente escarmentada,
y el lobo le devora la manada.
¡Cuántas veces resulta de un engaño
contra el engañador el mayor daño! [6]

and it can be phonemically transcribed in the following way:

apaθen'tando un 'xoben su ga'nado | gɾi'to desde la 'θima de un ko'ʎado | fa'boɾ ke 'biene el 'lobo | labɾa'doɾes || 'estos | abando'nando sus la'boɾes |

[6] A relatively literal English translation of this text would be the following: “While looking after his sheep, a young man / shouted from the top of a hill: / ‘Help! The wolf is coming!’ / Some peasants, leaving their tasks, / arrive immediately, / and they find that it is only a prank. / He calls once more and they fear a tragedy. / They are fooled again. What a joke! / But what happened the third time? / The hungry beast actually appeared. / Then the boy bawls, / kicks, cries and shouts, / but the tired people do not move / and the wolf eats his flock. / How often is the worst harm from a lie / for the liar himself!”

a'kuden pɾonta'mente | i 'aʌan ke es una 'tʃanθa sola'mente || 'buelbe a kla'maɾ i 'temen la des'gɾaθia || se'gunda beθ los 'buɾla | 'linda 'gɾaθia || peɾo 'ke suθe'dio la 'beθ teɾ'θeɾa || ke 'bino en reali'dad la am'bɾienta 'fieɾa || en'tonθes el θa'gal se desga'ɲita | i poɾ 'mas ke pa'tea | 'ʌoɾa i 'gɾita | no se 'muebe la 'xente eskaɾmen'tada | i el 'lobo le de'boɾa la ma'nada || 'kuantas 'beθes re'sulta de un en'gaɲo | 'kontɾa el engaɲa'doɾ el maʝoɾ 'daɲo ||

The transcriptions that appear above were written using the following Spanish phonemes: /a/, /e/, /i/, /o/, /u/, /p/, /t/, /k/, /b/, /d/, /g/, /tʃ/, /ʝ/, /f/, /s/, /x/, /m/, /n/, /ɲ/, /ɾ/, /r/, /l/, /θ/ and /ʎ/. The last two of them, however, are inexistent for most Spanish-language speakers, who merge them with /s/ and /ʝ/, respectively.[7] We nevertheless decided to keep them in the transcriptions, in order to illustrate the possible differences in the pronunciation of those words for which some speakers use /θ/ or /ʎ/, while other speakers use /s/ or /ʝ/. In that sense, therefore, the transcriptions that use that list of 24 phonemes can be seen as a "diasystemic" version of the corresponding text, which includes the /s-θ/ and /ʝ-ʎ/ mergers as special cases.[8]

Table 1: Descriptive statistics for the NWS and BCW texts

Concept	**NWS**	**BCW**
Words (tokens)	97	95
Content words	54	53
Particles	43	42
Words (types)	60	72
Content words	40	51
Particles	20	21
Phonemes (tokens)	428	423
Phonemes (types)	22	24

On table 1 we can see the main descriptive statistics for both the NWS and BCW texts, concerning their number of words and phonemes. Note that BCW is slightly shorter than NWS: it has 95 words (instead of 97) and 423 phonemes (instead of 428). The NWS

[7] The /s-θ/ merger is also known as *seseo*, and the /ʝ-ʎ/ merger is also known as *yeísmo*. See Penny (2004:118-121).

[8] The /s-θ/ split is typical of Castilian Spanish, where it is standard. The /ʝ-ʎ/ split was also standard in that accent until a few decades ago, but it has largely receded in modern Spain (at least for urban speakers). It is still widely heard, however, in some South American countries, especially in Bolivia and Paraguay. See Hualde (2005:20-30).

text, however, has only 60 word types (against the 72 types found in the BCW text). That difference is basically explained by a greater repetition in "content words", which are only 40 in NWS (against 51 found in BCW). Indeed, the Spanish NWS text repeats the words *viento* ("wind"), *norte* ("north"), *viajero* ("traveler") and *capa* ("cloak") four times each, while *sol* ("sun") appears three times. The BCW text, on the contrary, repeats two content words: *lobo* ("wolf") and *vez* ("occurrence, time"), and each of them appears only twice. The other repeated words are particles (prepositions, articles, relative pronouns, etc.), and their repetition rate is also higher in NWS, since the token/type ratio for that group of words is equal to 2.15 in the NWS text, and equal to 2.00 in the BCW text.

Another characteristic that is worth noting is that, while BCW has examples for the 24 Spanish phonemes, in NWS there are two missing observations: /ɲ/ and /ʝ/. It can be argued that the last of them is not very important, since most Spanish speakers merge /ʝ/ with /ʎ/, and /ʎ/ appears twice in the NWS text (*ellos* /'eʎos/ "them", and *brilló* /bɾi'ʎo/ "shone"). This is not the case, however, for speakers who actually pronounce /ʎ/ and /ʝ/ differently, and exhibit variation in the realization of those phonemes.[9] On the other hand, although /ɲ/ is a low-frequency phoneme with a limited distribution (it rarely appears at the beginning of a word, and it never appears in syllabic coda), it may be subject to some interesting phonetic processes, such as its depalatalization and its merger with the combination /ni/.[10]

3. Illustration of phonetic features in NWS and BCW

The pronunciation of the Spanish phonemes is subject to variation due to some relatively general phonological rules, and also because of some dialect differences. In this section we will mention the most important sources of variation, and see the capability of both the NWS and BCW texts to illustrate them. In order to do that, we will group them based on the type of phonemes affected by each analyzed variation.

[9] For example, in Eastern Ecuador there are people who use [ʒ] for /ʎ/ and [ʝ] for /ʝ/, while in Northeastern Argentina there are people who use [ʝ] for /ʎ/ and [dʒ] for /ʝ/. In those cases, the lack of an example for /ʝ/ may induce the observer to think that a speaker merges /ʝ/ and /ʎ/, when in fact he or she pronounces those phonemes differently. See Sessarego (2013:57-68) and Colantoni (2006).
[10] See Moreno (2011).

3.1. Vowels

Both the NWS and the BCW texts have a relatively large number of vowels as a percentage of their total number of phoneme tokens (47% in NWS and 46% in BCW). The five vowel phonemes are substantially represented in the two texts, being /e/ the one with more occurrences in NWS (61 tokens) and /a/ the one with more occurrences in BCW (69 tokens). In both cases, the vowel with fewer occurrences is /u/ (13 tokens in NWS and 14 in BCW).

/e/, /o/, /i/ and /u/ have basically two types of allophones: the vowels themselves ([e], [o], [i] and [u]) and the glides [e̯], [o̯], [j] and [w].[11] Both types of sounds are represented in NWS and BCW. As, in Spanish, glides are used to form (biphonematic) diphthongs, their occurrence is related to the appearance of those diphthongs. In NWS, the total number of diphthongs that we found is equal to 34, and the total number of diphthong types is 11 ([ja], [je], [jo], [wa], [we], [ej], [ae̯], [ao̯], [e̯a], [o̯a] and [o̯e]). In BCW, the total number of diphthongs is equal to 19, but the total number of diphthong types is also 11 ([ja], [je], [jo], [wa], [we], [aj], [ew], [ow], [ae̯], [e̯a] and [o̯e]).[12]

In NWS there are also two instances of identical consecutive vowels (/ee/ in both cases) that can be reduced to a single realization of [e], while in BCW there are two instances of /ee/ and one instance of /aa/ (which can be reduced to [a]).

3.2. Voiced obstruents

In Spanish, voiced obstruent phonemes are contrasted by their place of articulation (labial, coronal or velar), but not by their manner of articulation (plosive or continuant). Therefore, [d] and [ð] are allophones of the same phoneme, and the same occurs with [b] and [β], and with [g] and [ɣ]. The distribution of those allophones varies with the position of the phoneme, and also with the surrounding phonemes. If we apply

[11] The last two symbols can also be written as [i̯] and [u̯], respectively. See Hualde (2005:54-55).

[12] These calculations are made counting the diphthongs that appear inside words (e.g., *furia* ['fuɾja] "fury") and also the diphthongs formed by synalepha, i.e., when pronouncing two consecutive words (e.g., *y hallan* ['jaʎan] "and they find"). They are nevertheless conservative, since it is assumed that no phoneme is elided when reading the text. If some usual elisions were allowed (e.g., *considerado* [konsiðe'ɾao̯] "considered"), new diphthongs would appear.

the standard rules described for Castilian Spanish,[13] we find that the expected realizations of /b/, /d/ and /g/ in NWS and BCW are the ones reported on table 2.

Table 2: Expected pronunciation for voiced obstruent phonemes

Concept	**/b/**	**/d/**	**/g/**
NWS	19	14	3
Plosive [b, d, g]	3	5	1
Continuant [β, ð, ɣ]	16	9	2
BWS	18	23	10
Plosive [b, d, g]	2	4	3
Continuant [β, ð, ɣ]	16	19	7

In both the NWS and the BCW texts we have several instances of possible elision of /d/, which are common in some Spanish accents.[14] Those instances are more frequent in BCW than in NWS, since in NWS there are only three words for which /d/-elision could be reasonably expected (*de* [e] "of", *considerado* [konsiðe'ɾao̯] "considered", and *superioridad* [supeɾjoɾi'ða] "superiority"), while in BCW the number of likely /d/-elisions is higher (*ganado* [ga'nao̯] "livestock", *collado* [ko'ʎao̯] "hill", *realidad* [rẹ̯ali'ða] "reality", *escarmentada* [eskaɾmen'ta] "tired", *manada* [ma'na] "flock", and *engañador* [eŋgaɲa'oɾ] "liar").

The word *realidad*, which appears in BCW, is also a good example to check if the speaker actually pronounces the last /d/ as a standard [ð] or as a different sound (for example, [θ] or [t], which are usual allophones for that phoneme in that position in some regions of Spain).[15] Conversely, the word *superioridad*, which is in NWS, is not a good token to analyze this, because it appears in a context where the next word begins with another /d/ (and that is a situation where one expects to find assimilation of both sounds, which should be pronounced as a single [ð]).

Another variation that is reported for some accents is the use of [v] as an

[13] See, for example, González (2006). Those rules, however, are different for other accents (e.g., Colombian, Panamanian and Central American Spanish), which exhibit a more restricted use of the continuant allophones. See Piñeros (2002).
[14] See Samper (2011), Lipski (2011) and Monroy & Hernández (2015).
[15] See, for example, Molina (2008) and Gimeno & Gómez (2007).

allophone of /b/, especially for words written with the grapheme "v".[16] To test this possibility, both the NWS and the BCW texts provide a relatively good benchmark, since 10 instances out of 19 tokens of /b/ are written with "v" in NWS, and 11 instances out of 18 tokens of /b/ are written with "v" in BCW.

3.3. Voiceless fricatives

Variation in the pronunciation of the Spanish voiceless fricatives is basically related to the presence of the /s-θ/ merger or split, to the pronunciation of /s/, and to the pronunciation of /x/. All these phenomena are relatively well-illustrated in both the NWS and the BCW texts, although the number of occurrences of the phoneme /θ/ (and thus the number of chances to test if the speaker actually merges it with /s/) is much larger in BCW (12 cases) than in NWS (3 cases).

BCW therefore provides a better sample to check if someone who hesitates between the use of [θ] and [s] is more inclined towards the /s-θ/ merger or split.[17] Moreover, the three cases of /θ/ in NWS are in onset positions, while in BCW we have two cases of /θ/ in syllabic coda (and both of them appear before another consonant). Those cases are subject to additional variation related to possible processes of aspiration, elision and voicing, which are not common in onset positions.[18]

The number of occurrences of /s/, conversely, are roughly the same in the two texts (25 in NWS and 23 in BCW), and in both cases there is a considerable number of tokens in onset and coda positions.[19] In the NWS text, however, there are no examples of /s/ before a pause, while in the BCW text there are three cases like that. Both texts have examples of /s/ in coda before a vowel (1 in NWS, 2 in BCW), before a voiced consonant (2 in NWS, 5 in BCW) and before a voiceless consonant (4 in NWS, 3 in BCW), although in NWS two of such cases occur when the following phoneme is another /s/.

The phoneme /x/, finally, is more common in NWS than in BCW (6 tokens versus

[16] See Penny (2004:46-48).
[17] This might be interesting for speakers from Southern Spain (where the /s-θ/ merger co-exists with the /s-θ/ split), or for people who are bilingual in Spanish and Catalan (which is a language where [θ] has no phonemic status).
[18] See Hualde (2005:160-165).
[19] These last cases are also subject to processes that imply aspiration, elision and voicing.

2 tokens), but this is strongly influenced by the fact that the word *viajero* /bia'xeɾo/ is repeated four times in NWS. Both texts, however, have examples of /x/ in different positions (before /a/, /e/ and /o/ in NWS, and before /e/ and /o/ in BCW). This is good, because in some accents /x/ admits different pronunciations before different vowels.[20]

3.4. Nasal consonants

Spanish has three nasal consonant phonemes, which are /m/, /n/ and /ɲ/. The main source of variation within this group has to do with the neutralization of their phonemic opposition in syllabic coda, which generates allophones that adopt the point of articulation of the following consonant. This implies the use of [m] before /p/, /b/ and /f/, [ŋ] before /k/, /g/ and /x/, and [n] elsewhere.[21] In both NWS and BCW, there are relatively many cases where these phenomena can be illustrated, since there are 27 instances of nasal codas in each text.

Table 3: Expected pronunciation for nasals in syllabic coda

Position	Before	Predicted allophone	NWS	BCW
Interior	Consonant	[m]	3	1
Interior	Consonant	[n]	13	15
Interior	Consonant	[ŋ]	0	3
Final	Consonant	[m]	2	1
Final	Consonant	[n]	2	3
Final	Consonant	[ŋ]	2	3
Final	Vowel	[n]	5	1
Total			27	27

The figures that appear on table 3 show the distribution of the different nasal consonants in syllabic coda in NWS and BCW. In it we can see that in neither of these texts there are examples of nasal consonants before pauses, but that the two of them have

[20] In Chilean Spanish, for example, /x/ is typically pronounced as [ç] before /i/ and /e/, and [x] elsewhere. In many regions of Spain, conversely, it is pronounced as [χ] before /o/ and /u/, and [x] elsewhere. In Mexico, Argentina and other Latin American countries, the standard pronunciation for /x/ is [x] in all positions, while [h] is its typical pronunciation in places like Andalusia, Colombia, the Caribbean, and Central America. See Hualde (2005:154-155).

[21] This is a rather broad description of these allophones. A narrower one would imply the use of additional symbols such as [ɱ], [ɳ] and [ɴ]. See Martínez, Fernández & Carrera (2003).

examples of those consonants before other consonants, both in interior and in final positions. In NWS there are also five cases of final nasals before words that begin with a vowel, which is something that occurs only once in BCW. Both texts also have examples in which the predicted allophone for the nasal phonemes is [ŋ], although in the NWS text there are no such cases in interior positions.

The number of cases where table 3 indicates that the chosen allophone is [ŋ], however, is subject to a rule which predicts that pronunciation when a nasal phoneme appears before a velar consonant. In several Spanish dialects, however, velarization can occur in other contexts as well, especially when the nasal phonemes appear before a pause, or when they are in a final position and the following word begins with another consonant.[22] These cases, which are four in each text, can be used to test if a particular speaker belongs to one of those "velarizing dialects".

Another possible source of variation in the pronunciation of nasal phonemes has to do with the depalatalization of /ɲ/. This phoneme appears 4 times in BCW, but it does not appear in NWS. In BCW, moreover, it occurs in three different contexts: before /a/, before /o/, and before /i/. This can be useful because depalatalization of /ɲ/, and its corresponding substitution by the combination /ni/, may be less frequent before /i/ and more frequent before the other vowel phonemes.[23] It is therefore possible that the same speaker that uses [ɲ] for the word *desgañita* [dezɣa'ɲita] "bawls", pronounces the phoneme /ɲ/ as [nj] in *engaño* [eŋ'gaɲjo] "lie", *engañador* [eŋgaɲja'ðoɾ] "liar" and *daño* ['daɲjo] "harm".

3.5. Other consonants

The remaining Spanish consonant phonemes are the voiceless plosives /p/, /t/ and /k/, the laterals /l/ and /ʎ/, the tap /ɾ/, the trill /r/, the affricate /tʃ/, and the voiced fricative /ʝ/. The voiceless plosives are typically unaspirated in Spanish, and they are not subject to much allophonic variation. In some accents, however, /k/ may be pronounced as [c]

[22] This is typical of Galicia, Extremadura, Andalusia, the Canary Islands, Central America, the Caribbean, and the Pacific Coast of Colombia, Ecuador and Peru. See Samper (2011) and Lipski (2011).
[23] See Colantoni & Kochetov (2010).

before /i/ and /e/. In NWS there are four cases in which that pronunciation could be found (*quien* [cjen] "who", *quitarse* [ci'taɾse] "take off", and two instances of *que* [ce] "that"), while in BCW there are five instances of *que* but no examples of /k/ before /i/.

Variations within the pronunciation of /l/ and /ɾ/ are more important, because both phonemes are sometimes confused in certain Spanish accents, especially when they appear in syllabic coda in the interior of a word. NWS has nine cases like that (*porfiaban*, *fuerte*, *acertó*, *envuelto*, *quitarse*, and 4 instances of *norte*), while BCW has three of such cases (*vuelve*, *burla* and *tercera*). /ɾ/ and /l/ are also subject to possible elision, especially when they occur at the end of a word.[24]

In South America, the phoneme /r/ also has an important source of variation related to its possible pronunciation as a fricative sound (which could be something like [ɹ] or, more commonly, [ʐ̞]).[25] That phoneme appears twice in NWS (*arrebujaba* /arebu'xaba/ "folded around", and *reconocer* /rekono'θeɾ/ "to confess") and twice in BCW (*realidad* /reali'dad/ "reality" and *resulta* /re'sulta/ "turns out").

The affricate phoneme /tʃ/, conversely, appears only once in NWS (*ancha* /'antʃa/ "wide") and only once in BCW (*chanza* /'tʃanθa/ "prank"). Its main variation has to do with its possible deaffrication (which implies pronouncing it as [ʃ]) or voicing (which implies pronouncing it as [dʲ]).[26] The phoneme /ʲ/, finally, appears once in BCW (*mayor* /ma'ʲoɾ/ "largest") and, as we mentioned before, does not appear in NWS (unless the reader merges it with /ʎ/, which appears twice). It is a phoneme that exhibits considerable variation in Spanish, which goes from its possible assibilation (which implies pronouncing it as [ʒ] or [ʃ]) to its affrication (which implies pronouncing it as [ʄ] or [dʒ]) and its vocalization (which implies using the glide [j]).[27]

3.6. A comparison of NWS and BCW for Castilian and Andalusian accents

[24] For an account of this in Latin America and Spain, see Lipski (2011) and Samper (2011). See also Monroy & Hernández (2015), for a detailed description of this phenomenon in Murcian Spanish.
[25] This is typical of the Spanish spoken in Bolivia and Paraguay, and in some parts of Colombia, Ecuador, Peru, Chile and Argentina. See, for example, Escobar (2011) and Colantoni (2006).
[26] This last variation is associated with the Canary Islands (see Penny, 2004:129-131). /tʃ/-deaffrication, conversely, has been reported in very different places such as Andalusia, Chile, the Caribbean, and Northern Mexico. See Lipski (2011) and Villena (2008).
[27] See, for example, Kochetov & Colantoni (2011) or Hualde (2005:165-172).

In Coloma (2012), there is a list of ten phonetic features whose presence or absence is useful to characterize 28 dialect areas within the Spanish-speaking world. Those features are: /s-θ/ merger, /ʝ-ʎ/ merger, /s/-aspiration, /x/-aspiration, /ʝ/-assibilation, /r/-assibilation, /n/-velarization, /tʃ/-deaffrication, /x/-uvularization and /tʃ/-voicing. Nine of the defined dialect areas belong to Spain, while the remaining nineteen are located in different parts of Latin America.

The two dialect areas that are more extreme, in the presence or absence of the reported features, are the ones that correspond to the so-called "Traditional Castilian" accent (TC), which lacks all those features except /x/-uvularization, and the so-called "Western Andalusian" accent (WA), that possesses all of them except /r/-assibilation, /x/-uvularization and /tʃ/-voicing. Taking into account their relatively dissociated distribution of phonetic features, in this section we will use these two accents to exemplify the differences that can be found in the phonetic transcriptions of NWS and BCW.

Table 4: Differences between TC and WA transcriptions

Difference	**NWS**	**BCW**
/s-θ/ merger	3	12
/ʝ-ʎ/ merger	2	3
/s/-aspiration	4	13
/x/-aspiration	6	2
/ʝ/-assibilation	2	4
/n/-velarization	4	4
/tʃ/-deaffrication	1	1
/x/-uvularization	1	1
/d/-elision	4	6
/s/-elision	2	4
/ɾ/-elision	4	1
Total	33	51

Table 4 shows the number of differences between a TC and a WA phonetic transcription for both the NWS and the BCW texts.[28] Those differences are counted as

[28] All four transcriptions are reproduced in appendix 1. None of them comes from an actual recording, but the TC transcription for NWS is very similar to the ones that appear in Martínez, Fernández & Carrera (2003) and in Canepari (2005:254).

the number of phenomena that appear in the WA transcription but not in the TC transcription, or vice versa. This includes eight characteristics mentioned by Coloma (2012), plus three additional features related to possible elision of sounds.[29] As a result of this, we end up with transcriptions that exhibit 32 differences for the NWS text, and 51 differences for the BCW text.

4. Phoneme frequency distributions

Another possible comparison between the Spanish versions of NWS and BCW could be made considering the frequency distributions of the phonemes that appear in those texts. Those distributions, which come from counting the number of occurrences for each phoneme, can be contrasted with the ones reported in the literature for natural language. In order to perform such contrasts, we first study the "phonetic balance" of our two texts. After that, we try to estimate their corresponding distribution functions, assuming certain theoretical shapes and relating the frequencies of the different phonemes with their corresponding positions in the ranking of occurrences.

4.1. Phonetic balance

Following Sinclair (2005), we can state that a certain corpus is balanced if "the proportions of the different kinds of text that it contains correspond with informed and intuitive judgments". For a set of phonemes in a particular text, a general rule to assess this is to analyze if all possible phonemes appear in the text, if it uses a frequency which is close to natural language, if it contains examples from all relevant phonotactic rules, if it includes the smallest possible number of words, and if its words are in current use.[30]

The Spanish NWS text does not fulfill one of the conditions mentioned in the previous paragraph, since, as seen in section 2, it lacks two phonemes. Concerning their length, both NWS and BCW seem to be good examples, since their extension is relatively short for texts whose aim is to represent the different phonemes of a language. Most

[29] For an account of these phenomena, see Hernández & Villena (2009).
[30] See Jesus, Valente & Hall (2015).

phonotactic rules, moreover, are covered in both texts, although there are a few missing cases (e.g., NWS cannot detect /s/-aspiration or elision after a pause, nor some relatively rare cases of the /ʝ-ʎ/ split). Finally, most of the words that appear in NWS and BCW are relatively common, although three of them (*arrebujaba* "folded around" in NWS, and *collado* "hill" and *zagal* "boy" in BCW) may sound rather archaic in modern Spanish.

Table 5: Phoneme frequency distributions

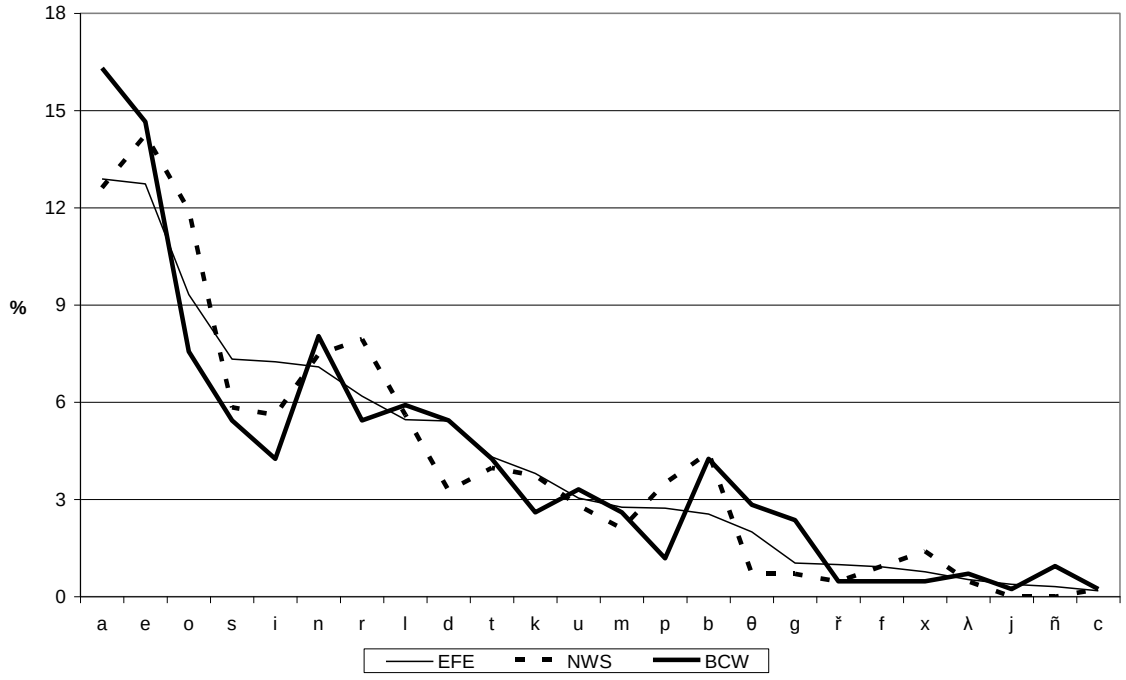
Phoneme	EFE		NWS		BCW	
	%	Ranking	%	Ranking	%	Ranking
a	12.89	1	12.62	2	16.31	1
e	12.74	2	14.25	1	14.66	2
o	9.32	3	11.92	3	7.57	4
s	7.33	4	5.84	6	5.44	7
i	7.25	5	5.61	7.5	4.26	10
n	7.09	6	7.48	5	8.04	3
ɾ	6.19	7	7.94	4	5.44	7
l	5.46	8	5.61	7.5	5.91	5
d	5.42	9	3.27	13	5.44	7
t	4.31	10	3.97	10	4.26	10
k	3.80	11	3.74	11	2.60	14.5
u	3.04	12	2.80	14	3.31	12
m	2.76	13	2.10	15	2.60	14.5
p	2.73	14	3.50	12	1.18	17
b	2.55	15	4.44	9	4.26	10
θ	2.00	16	0.70	18.5	2.84	13
g	1.04	17	0.70	18.5	2.36	16
r	0.99	18	0.47	20.5	0.47	21
f	0.92	19	0.93	17	0.47	21
x	0.77	20	1.40	16	0.47	21
ʎ	0.53	21	0.47	20.5	0.71	19
ʝ	0.38	22	0.00	23.5	0.24	23.5
ɲ	0.31	23	0.00	23.5	0.95	18
tʃ	0.18	24	0.23	22	0.24	23.5

In order to check if the phoneme frequency distributions are close to the one found in natural language, it is necessary to approximate the actual frequency of Spanish phonemes. To do that, we use one of the alternatives that appear in Moreno et al. (2008). That distribution comes from a large number of tokens (480,000 words and 2,511,856

phonemes), and it is based on a written corpus from the EFE news agency.[31]

On table 5, we have the phoneme frequency distributions for the NWS and BCW texts, together with the one that comes from the EFE corpus. We also report the rankings of phonemes derived from those distributions, and, when two phonemes have the same frequency in a certain distribution, we compute an "average ranking" for them.[32] The three frequency distributions are represented on figure 1, in which the order of the phonemes is the one that corresponds to the EFE distribution ranking.

Figure 1: Phoneme frequency distributions



One relatively direct measure of the similarity between two variables (e.g., two frequency distributions) is their standard (Pearson) correlation coefficient. In this case, if we calculate this measure for the EFE, NWS and BCW distributions, we see that their

[31] Other available alternatives are either shorter, or older, or are based on varieties of Spanish for which /ʎ/ and /θ/ are merged with /ʝ/ and /s/. Moreno et al. (2008) also reports an alternative frequency distribution based on an oral corpus of 1,244,411 phoneme tokens, but we preferred to use the EFE corpus because it was larger and it was based on written texts.

[32] This is necessary to calculate rank correlations between the distributions, and also to run regression equations that explain the shape of the frequency distributions as functions of the corresponding rankings.

correlation is very high ("r = 0.9634" for EFE vs. NWS, and "r = 0.9470" for EFE vs. BCW). The same occurs if we compute their rank (Spearman) correlation coefficients, which are correlation coefficients between the ranking variables. These are "r = 0.9443" for EFE vs. NWS, and "r = 0.9298" for EFE vs. BCW.

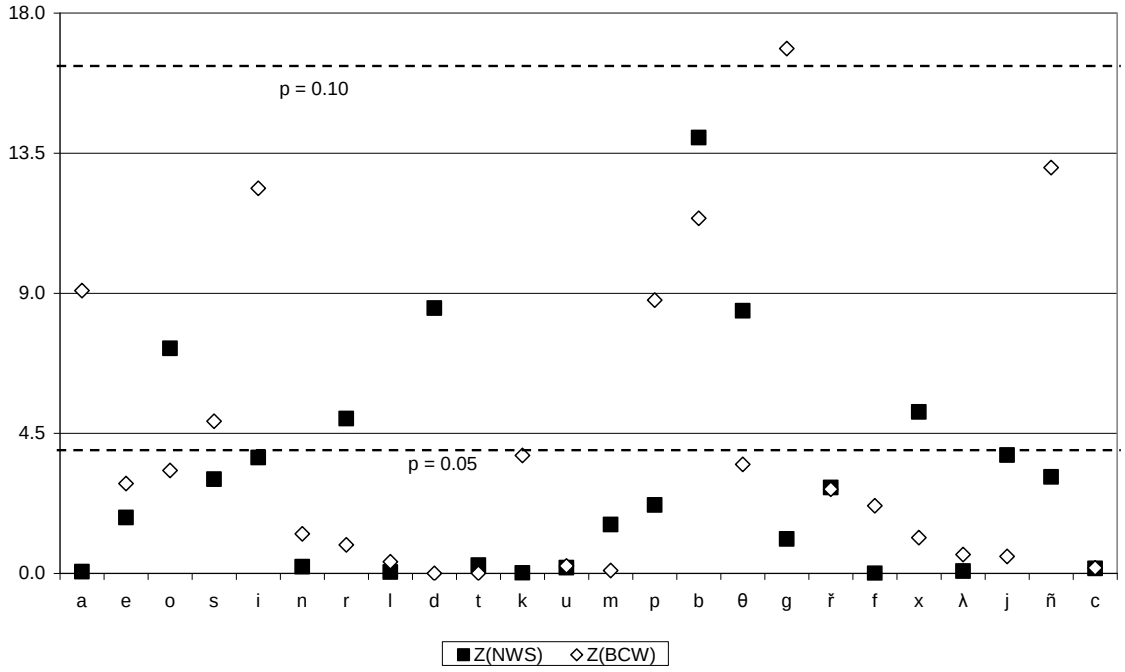
These very high correlation coefficients can be seen as an indication that our two texts are phonetically balanced, but this kind of average measures could be hiding some problems which might have an impact on particular phonemes. To find those problems, Jesus, Valente & Hall (2015) have used, in their study of the Portuguese version of the NWS text, a method created by Bland & Altman (1986) for assessing agreement between two samples. This method consists of calculating the following Z-statistics:

$$Z(NWS)_i = \frac{\left(f(NWS)_i - f(EFE)_i\right)^2}{f(EFE)_i} \quad ; \qquad Z(BCW)_i = \frac{\left(f(BCW)_i - f(EFE)_i\right)^2}{f(EFE)_i} \quad ;$$

where $f(EFE)_i$, $f(NWS)_i$ and $f(BCW)_i$ are the corresponding frequencies for an individual phoneme in the EFE, NWS and BCW distributions. These Z-statistics have a chi-squared distribution with one degree of freedom, and are statistically different from zero with a 10% probability level if their value is greater than 15.8, and statistically different from zero with a 5% probability level if their value is greater than 3.9.

On figure 2 we have depicted the Z-statistics for each phoneme in both the NWS and the BCW distributions. Only one of them is statistically different from zero at a 10% probability level, and twelve additional cases are significant at a 5% level. This can be seen on the figure, because there is one point above the line that represents "*p = 0.10*" and twelve additional points between that line and the one that represents "*p = 0.05*". The first of those points corresponds to /ɡ/, which is overrepresented in BCW, while the others correspond to /ɓ/ (overrepresented in both distributions), /ɑ/ (overrepresented in BCW), /o/ (overrepresented in NWS), /i/ (underrepresented in BCW), /ɾ/ (overrepresented in NWS), /ɗ/ (underrepresented in NWS), /k/ (underrepresented in BCW), /p/ (underrepresented in BCW), /θ/ (underrepresented in NWS), /x/ (overrepresented in NWS), and /ɲ/ (overrepresented in BCW).

Figure 2: Z-Statistics for NWS and BCW



4.2. Goodness of fit

Another way to compare the NWS and BCW phoneme frequency distributions is to estimate parameters for those distributions and to test if they are significantly different from the ones that correspond to the actual distribution of the Spanish phonemes. If they are not, one can say that those distributions have a similar shape than the actual distribution (which we are here approximating through the EFE frequency distribution). We can also calculate the goodness of the different parametric distributions to fit the data that comes from the different texts, through the use of some statistical measures.

In order to do all that, it is necessary to run regression analyses, using some functional form and some variables which are supposed to determine the corresponding phoneme frequencies. In the quantitative linguistics' literature, the most common function used for this is the Zipf distribution function,[33] which assumes that phonemes follow a distribution like this:

[33] See Baayen (2001:13-19)

$$f = a \cdot r^{b} \quad \Rightarrow \quad \log(f) = \log(a) + b \cdot \log(r) \quad ;$$

where *f* is the phoneme frequency, *r* is the ranking of the corresponding phoneme, and *a* and *b* are parameters.

The Zipf distribution, however, can be seen as a particular case of a more general function called the Yule distribution, whose formula is the following:

$$f = a \cdot r^{b} \cdot c^{r} \quad \Rightarrow \quad \log(f) = \log(a) + b \cdot \log(r) + \log(c) \cdot r \quad ;$$

where *c* is an additional parameter. This more general distribution has been tested by Tambovtsev & Martindale (2007) for a sample of 95 languages, and has been found to fit the data better than the Zipf distribution.

Table 6: Regression results

Concept	Zipf		Yule	
	Coefficient	p-value	Coefficient	p-value
EFE frequency equation				
Parameter a	42.7134	0.0000	12.4465	0.0000
Parameter b	-1.2573	0.0000	0.5094	0.0010
Parameter c			0.7993	0.0000
R-square	0.7314		0.9701	
NWS frequency equation				
Parameter a	57.8964	0.0000	13.0855	0.0000
Parameter b	-1.4565	0.0000	0.6754	0.0011
Parameter c			0.7631	0.0000
R-square	0.7101		0.9613	
BCW frequency equation				
Parameter a	45.3057	0.0000	14.4025	0.0000
Parameter b	-1.2968	0.0000	0.3459	0.0689
Parameter c			0.8118	0.0000
R-square	0.7547		0.9562	

On table 6 we can see the main results for three regression equations (corresponding to the EFE, NWS and BCW phoneme frequencies) that were run under both the Zipf and Yule specifications. The equations were linearized using natural logarithms, and the log of the observed frequency has been explained as a function of the log of the phoneme ranking (and the phoneme ranking itself). The Yule distribution has a

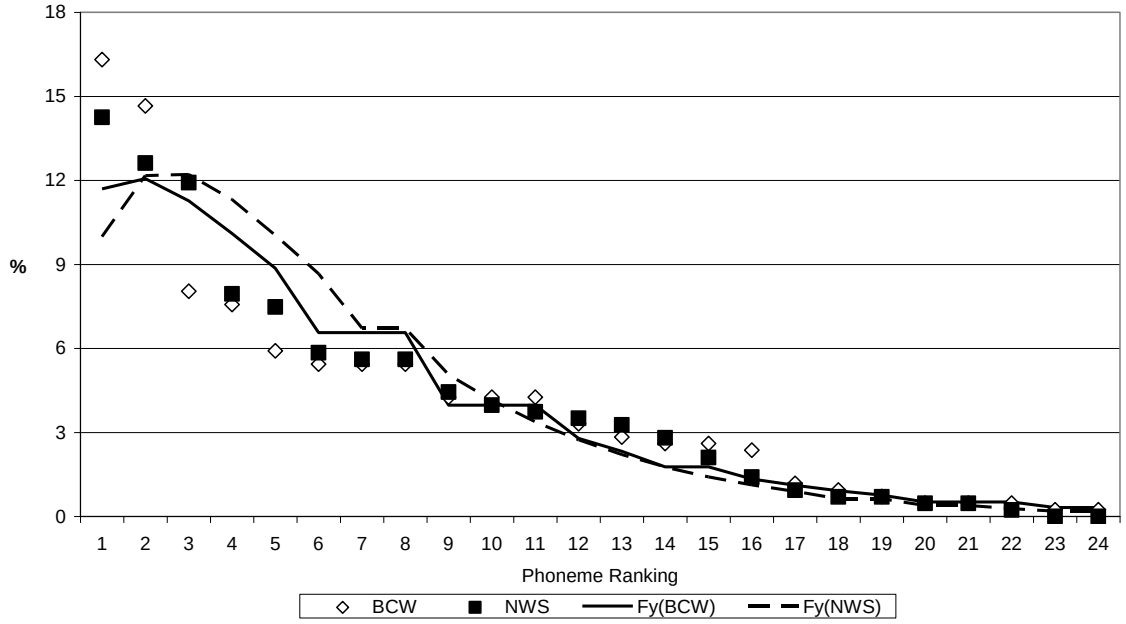
better fit than the Zipf distribution for the three analyzed equations, as can be seen by looking at the corresponding "R^2 coefficients" (which are always substantially higher when regressions are run using the Yule specification). These coefficients also show that the BCW equation has a better fit than the NWS equation in the Zipf specification, and a slightly worse one in the Yule specification.

We should nevertheless point out that, in order to perform our regression analyses, the NWS frequencies had to be adjusted using a simplified version of the so-called "Good-Turing estimates".[34] This adjustment was necessary because two observations (the ones that correspond to /ɲ/ and /ʝ/) are equal to zero in the NWS frequency series, and it was therefore impossible to calculate natural logarithms for those observations without using a technique that imputes an estimated positive value. The technique that we used consisted of estimating a probability for the missing phonemes (/ɲ/ and /ʝ/) that is equal to the probability of the observed phoneme with the lowest frequency (which in this case is /tʃ/). This probability was evenly divided between /ɲ/ and /ʝ/, and then the frequencies for all the observations were adjusted so that the sum of all frequencies added up to 100%.

On figure 3, we have depicted the results of our regression estimations in a diagram that shows the actual and predicted frequencies for the different phonemes (ordered by their rankings) in the NWS and BCW texts. In both cases, the predicted frequencies are graphed using the results of the Yule distribution regression equations (Fy), which are the ones that have the best fit in both cases. Note that the prediction for the NWS distribution is rather awkward, since it has a pronounced positive slope for the first two observations, as if the third phoneme in the ranking had a higher probability of occurrence than the first two ones.

[34] For an explanation of this concept, see Baayen (2001:57-63) or Sampson (2001:94-108).

Figure 3: Yule distributions for NWS and BCW



A last possible comparison between NWS and BCW is a statistical test that estimates the joint probability that the parameters for these distributions are actually the same ones that were computed for the EFE distribution. That test was performed using a chi-square statistic of the null hypothesis under which we alternatively supposed that, for the Zipf distributions, it held that:

a(EFE) = a(NWS), b(EFE) = b(NWS) ; a(EFE) = a(BCW), b(EFE) = b(BCW) ;

while for the Yule distributions it held that:

a(EFE) = a(NWS), b(EFE) = b(NWS), c(EFE) = c(NWS) ;
a(EFE) = a(BCW), b(EFE) = b(BCW) c(EFE) = c(BCW) .

After running those tests, we found that the probability that the null hypothesis is true for the Zipf specification of the NWS distribution is equal to 0.5693, while the probability that the null hypothesis is true for the Zipf specification of the BCW distribution is equal to 0.9707. When using the Yule specifications, those probabilities ended up being equal to 0.0104 for the NWS frequency distribution, and equal to 0.8576

for the BCW frequency distribution.[35] As we can see, both pairs of tests show a very clear preference for the phoneme frequency distribution that comes from the BCW text over the one that comes from the NWS text, in terms of their closeness to the theoretical frequency distribution that is behind the EFE corpus.

5. Concluding remarks

The main findings from the analyses performed, concerning the relative advantages and disadvantages of NWS and BCW to illustrate the phonetics of Spanish, can be summarized as follows:

a) Both texts are relatively short, especially if we compare them with other texts that could be phonetically balanced.[36] Their phoneme frequency distributions also display very high correlation coefficients when they are contrasted with the EFE frequency distribution (which is based on a written corpus from an important Spanish news agency, and is calculated using a very large number of tokens).

b) BCW has examples for all 24 Spanish phonemes, while NWS lacks two of them. NWS also has a higher word repetition rate, and lacks examples for a few important phonetic contrasts (e.g., /s/ before a pause, /d/ before another phoneme).

c) When used to exemplify two relatively extreme Spanish accents (Traditional Castilian and Western Andalusian), the phonetic transcriptions for NWS exhibit 33 differences, while the ones for BCW exhibit 51 differences (i.e., 55% more).

d) If we apply regression analysis, and approximate the different phoneme frequencies using Zipf and Yule distributions, the parameters found for BCW are relatively close to the ones estimated for the EFE frequency distributions. This does not occur with the coefficients estimated in the NWS regressions, whose probability of being equal to the EFE distribution parameters is much smaller.

---

[35] These numbers, like the ones that come from the regression analyses, were calculated using the program EViews 3.1.

[36] This is due to the fact that the probability value for the less frequent phoneme in Spanish is equal to 0.18% (if we use the EFE distribution shown on table 6) and therefore, on average, we need 555 phoneme tokens to have all the Spanish phonemes in a balanced sample. The NWS and BCW texts have 428 and 423 phoneme tokens, respectively, so it is not likely that texts that are shorter than them are phonetically balanced and, at the same time, have tokens for all the Spanish phonemes.

As a result of all this, we can state that the proposed BCW text seems to be considerably better than the standard NWS text to illustrate the phonetics of the Spanish language. This conclusion is similar to the one obtained in Deterding (2006) for the phonetics of the English language.

Appendix 1: Phonetic transcriptions

The North Wind and the Sun (Traditional Castilian)

el 'βjento 'noɾte jel 'sol poɾ'fjaβan soβɾe 'kwal 'deʎo 'seɾaˌel 'mas 'fweɾte | kwandoˌa θeɾ'toˌa pa'saɾum bja'xeɾoˌem 'bweltoˌe 'nantʃa 'kapa || kombi'njeɾo neŋ ke kje 'nantez lo'ɣɾaɾaoˌ βli'ɣaɾal βja'xeɾoˌa ki'taɾse la 'kapa se'ɾia konsiðe'ɾaðo 'mas poðe'ɾoso || el 'βjento 'noɾte so'plo koŋ 'gram 'fuɾja | peɾo 'kwanto 'maso 'plaβa 'maseˌa reβu'xaβaˌen su 'kapaˌel βja'xeɾo || poɾ 'finel 'βjento 'noɾteˌa βando'no laˌem'pɾesa || en'tonθez βɾi'ʎoˌel 'sol ko naɾ'doɾ | ejnme'ðjata'mente se ðespo'χo ðe su 'kapaˌel βja'xeɾo | poɾ lo kel 'βjento 'noɾte 'uβo ðe rekono'θeɾ la supeɾjoɾi'ðaðel 'sol ||

The North Wind and the Sun (Western Andalusian)

el 'βjento 'noɾte jel 'sol poɾ'fjaβaŋ soβɾe 'kwal 'deʒo 'heɾaˌel 'ma 'fweɾte | kwandoˌa seɾ'toˌa pa'sawŋ bja'heɾoˌem 'bweltoˌe 'nanʃa 'kapa || kombi'njeɾo neŋ ke kje 'nanteh lo'ɣɾaɾaoˌ βli'ɣal βja'heɾoˌa ki'taɾse la 'kapa se'ɾia konsiðe'ɾaoˌ 'mah poðe'ɾoso || el 'βjento 'noɾte so'plo koŋ 'graŋ 'fuɾja | peɾo 'kwanto 'maso 'plaβa 'maseˌa reβu'haβaˌeŋ su 'kapaˌel βja'heɾo || poɾ 'finel 'βjento 'noɾteˌa βando'no laˌem'pɾesa || en'tonse βɾi'ʒoˌel 'sol ko naɾ'do | ejnme'ðjata'mente se ðehpo'hoˌe su 'kapaˌel βja'heɾo | poɾ lo kel 'βjento 'noɾte 'uβoˌe rekono'se la supeɾjoɾi'ðaˌel 'sol ||

The Boy who Cried Wolf (Traditional Castilian)

apaθen'tandowŋ 'χoβen su ɣa'naðo | gɾi'to ðezðe la 'θima ðewŋ ko'ʎaðo | fa'βoɾ ke 'βjenel 'loβo | laβɾa'ðoɾes || 'estos | aβando'nando suz la'βoɾes | a'kuðem pɾonta'mente | 'jaʎaŋ ke suna 'tʃanθa sola'mente || 'bwelβeˌa kla'maɾi 'temen la ðez'ɣɾaθja || se'ɣunda βeð loz 'βuɾla | 'linda 'ɣɾaθja || peɾo 'ke suθe'ðjo la 'βeθ teɾ'θeɾa || ke 'βinoˌen reˌali'ðað lam'bɾjenta 'fjeɾa || en'tonθe sel θa'ɣal se ðezɣa'ɲita | i poɾ 'mas ke pa'tea | 'ʎoɾaj 'ɣɾita | no se 'mweβe la 'xentes kaɾmen'taða | jel 'loβo le ðe'βoɾa la ma'naða || 'kwantaz 'βeθez re'sulta ðew neŋ'gaɲo | 'kontɾaˌe leŋgaɲa'ðoɾel maˌɲoɾ 'ðaɲo ||

The Boy who Cried Wolf (Western Andalusian)

apasen'tandowŋ 'hoβeŋ su ɣa'naoˌ | gɾi'to ðehðe la 'sima ðewŋ ko'ʒaoˌ | fa'βo keˌ 'βjenel 'loβo | laβɾa'ðoɾe || 'ehto | aβando'nando suh la'βoɾe | a'kuðeŋ pɾonta'mente | 'jaʒaŋ ke huna 'ʃansa sola'mente || 'bwelβeˌa kla'maɾi 'temeŋ

la ðeh'ɣɾasja || se'ɣunda βeh loh 'βuɾla | 'linda 'ɣɾasja || peɾo 'ke suse'ðjo la 'βeh teɾ'seɾa || ke 'βinɸeŋ rɸali'ða lam'bɾjenta 'fjeɾa || en'tonse hel sa'ɣal se ðehɣa'ɲita | i poɾ 'mah ke pa'tea | 'ʒoɾaj 'ɣɾita | no se 'mweβe la 'henteh kaɾmen'ta | jel 'loβo le ðe'βoɾa la ma'na || 'kwantah 'βese re'sulta ðew neŋ'gaɲo | 'kontɾaɸ leŋgaɲa'oɾel maʒoɾ 'ðaɲo ||

References

Avelino, Heriberto (2017). Illustrations of the IPA: Mexico City Spanish. *Journal of the International Phonetic Association*, https://doi.org/10.1017/S0025100316000232.

Baayen, Harald (2001). *Word Frequency Distributions*. Dordrecht, Kluwer.

Bland, John & Douglas Altman (1986). Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement. *Lancet* 1: 307–310.

Bowden, John & John Hajek (1996). Illustrations of the IPA: Taba. *Journal of the International Phonetic Association* 26(1): 55-57.

Bowern, Claire, Joyce McDonough & Katherine Kelliher (2012). Illustrations of the IPA: Bardi. *Journal of the International Phonetic Association* 42(3): 333-351.

Canepari, Luciano (2005). *A Handbook of Pronunciation*. Munich: Lincom Europa.

Carlson, Barry & John Esling (2000). Illustrations of the IPA: Spokane. *Journal of the International Phonetic Association* 30(1): 97-102.

Colantoni, Laura (2006). Micro and Macro Sound Variation and Change in Argentine Spanish. *Proceedings of the 9th Hispanic Linguistics Symposium*, 91-102. Somerville, Cascadilla.

Colantoni, Laura & Alexei Kochetov (2010). Palatal Nasals or Nasal Palatalization? *Linguistic Symposium on Romance Languages* (LSRL 40). Seattle: University of Washington.

Coloma, Germán (2012). The Importance of Ten Phonetic Characteristics to Define Dialect Areas in Spanish. *Dialectologia* 9: 1-26.

Coloma, Germán (2017). Illustrations of the IPA: Argentine Spanish. *Journal of the International Phonetic Association*, https://doi.org/10.1017/S0025100317000275.

Connell, Bruce, Firmin Ahoua & 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.

Escobar, Anna (2011). Spanish in Contact with Quechua. In Díaz-Campos, Manuel, ed: *Handbook of Hispanic Sociolinguistics*, 323-352. Oxford, Wiley-Blackwell.

Gimeno, Francisco & José Gómez (2007). Spanish and Catalan in the Community of Valencia. *International Journal of the Sociology of Language* 184: 95-107.

González, Carolina (2006). The Phonetics and Phonology of Spirantization in North-Central Peninsular Spanish. *ASJU International Journal of Basque Linguistics and Philology* 40: 409-436.

Guerin, Valerie & Katsura Aoyama (2009). Illustrations of the IPA: Mavea. *Journal of the International Phonetic Association* 39(2): 249-262.

Hernández, Juan & Juan Villena (2009). Standardness and Nonstandardness in Spain: Dialect Attrition and Revitalization of Regional Dialects of Spanish. *International Journal of the Sociology of Language* 196: 181-214.

Hualde, José (2005). *The Sounds of Spanish*. New York: Cambridge University Press.

IPA (1912). *Principles of the International Phonetic Association*. Paris: International Phonetic Association.

IPA (1949). *Principles of the International Phonetic Association*. London: University College.

IPA (1999). *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.

Jesus, Luis, Ana Valente & 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.

Kochetov, Alexei & Laura Colantoni (2011). Coronal Place Contrasts in Argentine and Cuban Spanish: An Electropalatographic Study. *Journal of the International Phonetic Association* 41(3): 313-342.

Lipski, John (2011). Socio-Phonological Variation in Latin American Spanish. In Díaz-Campos, *op. cit.,* 72-97.

Martínez, Eugenio, Ana Fernández & Josefina Carrera (2003). Illustrations of the IPA: Castilian Spanish. *Journal of the International Phonetic Association* 33(2): 255-260.

Molina, Isabel (2008). The Sociolinguistics of Castilian Dialects. *International Journal of the Sociology of Language* 193: 57-78.

Monroy, Rafeel & Juan Hernández. 2015. Illustrations of the IPA: Murcian Spanish. *Journal of the International Phonetic Association* 45(2), 229–240.

Moreno, Antonio, Doroteo Toledano, Raúl de la Torre, Marta Garrote & José Guirao (2008). Developing a Phonemic and Syllabic Frequency Inventory for Spontaneous Spoken Castilian Spanish and their Comparison to Text-Based Inventories. *Proceedings of the LREC 2008*, 1097-1100. Marrakech, ELRA.

Moreno, Francisco (2011). Internal Factors Conditioning Variation in Spanish Phonology. In Díaz-Campos, *op. cit.,* 54-71.

Penny, Ralph (2004). *Variation and Change in Spanish*. Cambridge: Cambridge University Press.

Piñeros, Carlos (2002). Markedness and Laziness in Spanish Obstruents. *Lingua* 112: 379-413.

Samaniego, Félix (2003). *Fábulas en verso castellano para uso del Real Seminario Vascongado*. Alicante: Biblioteca Virtual Cervantes.

Samper, José (2011). Sociophonological Variation and Change in Spain. In Díaz-Campos, *op. cit.,* 98-120.

Sampson, Geoffrey (2001). *Empirical Linguistics*. London, Continuum.

Sessarego, Sandro (2013). *Chota Valley Spanish*. Frankfurt, Iberoamericana/Verbuert.

Sinclair, John (2005). Corpus and Text: Basic Principles. In Wynne, Martin, ed: *Developing Linguistic Corpora*, 1-16. Oxford, Oxbow Books.

Tambovtsev, Yuri & Colin Martindale (2007). Phoneme Frequencies Follow a Yule Distribution. *SKASE Journal of Theoretical Linguistics* 4(2):1-11.

Villena, Juan (2008). Sociolinguistic Patterns of Andalusian Spanish. *International Journal of the Sociology of Language* 193: 139-160.