Date Accepted: August 31 2015

Date Published: August 31 2015

¿Qué twiteastes tú? Variation in second person singular preterit –s in Spanish tweets

Chelsea Escalante

University of California - Davis, cescalante@ucdavis.edu

Recommended Citation/Citación recomendada

Escalante, Chelsea (2015) "¿Qué twiteastes tú? Variation in second person singular preterit –s in Spanish tweets," *Entrehojas: Revista de Estudios Hispánicos*: Vol. 5 : Iss. 1 , Article 1.

¿Qué twiteastes tú? Variation in second person singular preterit –s in Spanish tweets

Abstract/Resumen

Spanish dialects throughout Spain and the Americas have shown variation in the second person singular form of the preterit tense; in certain cases, a non-standard –s is found at the end of the verb conjugation (*fuistes*, *comistes*, *dijistes*, etc). This has been mentioned descriptively by several researchers as a cross-dialectal as well as historical feature of many dialects, yet little empirical data is available on this topic related to what factors constrain the variable. In fact, only one study has looked at this phenomenon from a variationist perspective (see Barnes, 2012). This study borrows certain methodological aspects of Barnes' (2012) analysis of oral data but applies them to the analysis of the variable as it exists in the written sphere. Data is collected through the social media mogul Twitter and tabulated with the multiple regression logistic software, GoldVarb X (Sankoff, Tagliamonte, & Smith, 2005). The data suggests that verb frequency is the only factor group that significantly conditions one variant over another, with high frequency verbs highly conditioning the standard and low frequency verbs highly conditioning the non-standard.

Varios dialectos españoles que se encuentran en España y en la Américas tienen una tendencia de demostrar variación en la forma de segunda persona singular del pretérito; en ciertos casos, se encuentra una –s no-estándar al final de la conjugación del verbo (*fuistes*, *comistes*, *dijistes*, etc). Este fenómeno ha sido mencionado descriptivamente por varios investigadores como un rasgo inter-dialectal y también histórico, sin embargo, no hay muchos datos empíricos disponibles acerca de qué factores lingüísticos restringen la variable. De hecho, solamente un estudio ha explorado el fenómeno desde la perspectiva variacionista (véase Barnes, 2012). Este estudio toma prestado algunos aspectos metodológicos del análisis de data oral de Barnes (2012) pero los aplica al análisis de la variable dentro del ámbito escrito. Los datos se recogen a través la red social enorme Twitter y se tabulan con el programa estadístico GoldVarb X (Sankoff, Tagliamonte, & Smith, 2005). Se encuentra que la frecuencia del verbo es el único factor que significativamente restringe la variable; los verbos de alta frecuencia condicionan el uso de la forma estándar mientras que los de baja frecuencia favorecen la no-estándar.

Keywords/Palabras clave

variation, sociolinguistics, Twitter, second-person preterit -s, frequency

Creative Commons License

This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Labov (1963, 1966, 1969, 1972a, 1972b), often considered the founding father of modern day studies of language variation and change, was the first to show that language variation was not random, but instead characterized by “structured heterogeneity” (Weinreich, Labov, & Herzog, 1968, p. 99-101). That is, speakers’ choices among variable linguistic forms – be them conscious or unconscious decisions—are systematically constrained by multiple linguistic and social factors that reflect underlying grammatical systems. These underlying systems both indicate and partially constitute the speech communities to which the speakers belong (Bayley, 2013). The goal of this study is to provide a variationist analysis of one variable present in several dialects across the Spanish speaking world: nonstandard *-s* on second person singular (*tú*) preterite verb forms¹ (*comistes*, *fuistes*, *dijistes*, etc.). The current paper examines the variable—as shown in (1) and (2)—as it exists in Spanish tweets and analyzes the linguistic factors that constrain its use.

- (1) Standard:
 @xxx² Hey, ya **regresaste** al club. Wohooo! Jkasksk :D
 ‘@xxx Hey, you already came back to the club.’
- (2) Non-standard:
 ¿**regresastes** porque me quieres? Si me quisieras, nunca te hubieras ido.
 ‘You came back because you love me? If you loved me, you would have never gone.’

Several dialects throughout Latin America and Spain have shown variation in this form of the preterit tense, denoted by several researchers as a cross-dialectal phenomenon that is both historical as well as current. Cerda (1953) reports the use of the non-standard variant in Oaxaca, Veracruz, and Argentina; Lapesa (1965) cites its occurrence in Andalucía; Lance (1975) describes its common appearance in Texas Spanish; Vaquero de Ramirez (1998), Lipski (2002), and Frago García (2003) acknowledge its occurrence in various Latin American and Peninsular dialects; Mendez-Pidal (1962) cites evidence of the non-standard variant in early eighteenth century Spanish and hypothesizes that the

¹ Although there is more than one second person singular personal pronoun in Spanish, only the *tú* form presents the kind of variation discussed in this article. References to second person singular forms in this article refer only to the *tú* form and not *vos*.

² User names have been deleted for privacy and substituted with @xxx

variation must be even older than that due to its common presence in Judeo-Spanish. Since Spanish inflectional morphemes have descended from the Latin endings –ASTI and –ISTI, which did not contain a final –s, it can be assumed that this is a non-etymological modification that coexists with the standard form (Barnes, 2012). Perhaps because of the fact that in all other tenses and moods, Spanish verbs in the second person singular form do end in –s (*comes* in present tense; *comías* in imperfect; *comerías* in conditional; *comas* in present subjunctive; *comerás* in future), the preterit non-standard –s has been thought to be the result of the linguistic phenomenon of regularization.

The present study looks at this phenomenon from the variationist perspective and attempts to discover the linguistic factors that condition or inhibit the production of the non-standard variant. Data is collected through the social media mogul *Twitter* and tabulated with the multiple regression logistic software, GoldVarb.

2. Previous studies

Despite its well-known occurrence in the Spanish-speaking world, and despite the plethora of data available on syllable-final –s weakening in Spanish, there has been very limited research of second person singular preterit –s. The above-mentioned sources describe the existence of the variable but very little quantitative data is available on the topic. This may be in part due to methodological questions; many times variationist data is collected through a sociolinguistic interview, however, since this variable exists only within the second-person, it is difficult to elicit if the researcher is the one asking the questions. A sociolinguistic interview participant might be wary about asking the researcher any question at all, much less using the informal *tú* second-person marking, thus making the emergence of either variant—the standard or non-standard—very rare.

To my knowledge, there has been only one study that looks at second-person singular –s usage from a variationist perspective (see Barnes, 2012). Using corpora from three different publically available sources, she analyzes 854 tokens of the variable found in transcriptions of oral data. She finds that two factors are significant in conditioning the production of the –s variant: verb frequency and following phonological context. High-frequency verbs were found to condition the standard variant while low-frequency verbs were found to condition the non-standard, a pattern that she views as a confirmation of Penny's (2002) claims of the analogical nature of the variation as well as support for the lexical diffusion model proposed by Bybee (1998, 2000, 2002), according to which “frequency of use affects the mental representation of lexical items and determines the direction

of phonological and morphological changes” (Barnes, 2012, p. 46).

In terms of following phonological context, following segment vowels were shown to condition the non-standard while following segment pauses conditioned the standard. Her analysis suggests that these patterns are in accordance with the preference for CV sequences across word boundaries in Spanish. In summarizing her findings, she asserts that these two significant factor groups are not isolated factors but rather are interconnected and interdependent; she characterizes the addition of *-s* on second person singular preterits as an analogically motivated process that is favored by certain phonological factors. She also notes that in the cases in which the addition of *-s* takes place, the analogical process overrides the general phonological tendency of weakening that affects word-final segments and that, according to the lexical diffusion model, is more likely to occur in high-frequency words first (p. 46). Barnes also suggests that further research is needed in order to determine if this phenomenon is a change in progress and what the direction of that change might be.

This current study aims to expand the scholarship begun by Barnes on this very common but underrepresented phenomenon in the literature. It provides the first variationist account for second person singular preterit *-s* in the written sphere as well as the first that takes into account bilingual (Spanish/English) data. Lastly, it provides another variationist analysis that uses the microblogging site Twitter as the primary data source, a trend in current sociolinguistic research. The specific research questions it aims to explore are: (1) What linguistic factors constrain the variable as it is found in Spanish tweets? (2) How do these constraints differ from those found to be significant in Barnes’ (2012) analysis of oral data?

3. Methodology

A total of 403 tokens were collected from social media posts publically available on Twitter. Due to its ability to function in a way that mimics snippets of a conversation, Twitter has been recognized as a particularly useful tool for students and instructors to partake in ongoing dialogue and for that reason can be considered advantageous to other social media platforms (Junco, Heiberger, & Loken, 2011). This service allows its users to post 140-character messages, with each author’s messages appearing in the newsfeeds of individuals who have chosen to ‘follow’ the author, though by default the messages are publicly available to anyone on the Internet. According to Bamman, Eisenstein, & Schnoebelen (2012), Twitter has relatively broad penetration across different ethnicities, genders, and income levels. Unlike Facebook, the majority of content on Twitter is explicitly public and unlike blogs, Twitter data is encoded in a single

format, facilitating large scale data collection. Large numbers of tweets may be collected using Twitter's streaming API, which delivers a stream that is randomly sampled from the complete set of public messages on the service (Bamman et. al, 2012, p. 11).

Twitter also allows users to search for certain words within tweets. For my data collection, I searched for the two variants in question of the 37 verbs listed in Table 1.

Table 1. List of verbs

aprender	criar	enterar	hacer	matar	poner	salir	ver
casar	decir	entrar	ir	meter	preguntar	sentir	vivir
comentar	dejar	equivocar	llamar	mirar	preparar	tener	
conocer	empezar	escoger	llegar	montar	quedar	tomar	
contar	encontrar	estar	llevar	notar	regresar	venir	

Barnes (2012) found that these verbs showed variation in the three oral corpora that she analyzed: the oral archive of the *Corpus de Referencia del Español Actual* (CREA), the *Habla Popular* Mexican Spanish corpus (Lope Blanch, 1976) and the *Corpus del Español* (Davies). For reasons of comparability, only the same verbs were included in the present study. Tokens were collected from Twitter during one 24-hour period (October 29, 2013), at which time each of the variants of the verbs in Table 1 were gleaned from the Twitter feed, resulting in a total of 403 tokens.

Because personal information including sex, age, socioeconomic status, occupation, and location are not available for all users, social factors were not taken into account in this study. Barnes (2012) suggests that these social factors are in fact important, stating that “previous dialectological descriptions of different varieties of Spanish have pointed to a possible relation between certain social factors, particularly those of social class and education” (p. 40). Even though these factors have a potential impact on the variable, due to lack of social information available from Twitter, only linguistic factors were taken into account. Tokens were coded for six different linguistic factors: presence of pronominal subject, frequency of preterit form, code-switching within tweet, directionality of tweet, type of conjugation, and following phonological environment.

Following the coding process, tokens were submitted to the multiple logistical regression software often used in sociolinguistic research, GoldVarb X (Sankoff, Tagliamonte, & Smith, 2005), to determine the probabilistic weights of the factors under study in conditioning the use of one variant over another.

3.1 Presence of pronominal subject.

Following Barnes' (2012) methodology, tokens were coded based on the presence or absence of the pronominal subject. Spanish is a “null-subject” or “pro-drop” language; speakers may express the subject of a finite verb as a lexical, pronominal, or phonetically null noun phrase (Cameron, 1996, p. 62). In order to explore how the variable may be constrained by the presence of a pronominal subject, tokens were coded for the inclusion (3) or absence (4) of *tú*³.

- (3) Overt pronoun:
Tu viste el slogan de Virgin Mobile @xxx? “A higher calling” #gold
‘Did you see the slogan of Virgin Mobile @xxx? “A higher calling” #gold’
- (4) Null pronoun:
WEY YA VISTE EL TRAILER DE DAYS OF FUTURE PAST?
 @xxx
‘Dude did you see the trailer for Days of Future Past? @xxx’

Due to fact that several Caribbean dialects of Spanish aspirate/delete syllable-final /s/ and have been reported to have higher overt second-person subject pronoun use, it was conceived by Sabater (1978) that this process was due to functional compensation, or the production of a form needed to compensate for information that has been deleted by another process. Because second person verbs rely on the final *-s* morpheme to distinguish it from third-person singular forms (and in many cases, first-person singular as well), the addition of the overt pronoun was thought to be a way that speakers could compensate for the information lost when the morpheme *-s* was aspirated/deleted. Sabater explains his interpretation of the data found in Dominican Spanish:

Especially noteworthy is the presence of the pronoun *tú*, a phenomenon which I interpret as an obvious readjustment to the disappearance of the verbal morpheme of second person: the *-s* of *amas*. The Dominican, confronted by this loss, tends to adopt the same resource as Old French

³ The second-person pronominal subject *tú* is orthographically represented in standard dialects with an accent mark (*tilde*) in order to distinguish it from the non-accented possessive pronoun *tu* ‘your’. However, since non-standard orthography is common in social media posts, any *tu* positioned in front of a verb—with or without accent mark—was coded as a subject pronoun and not considered a possessive

and to use therefore the prefixed pronoun as a constant marker of verbal person (Quoted in Cameron, 1996: 66, his translation).

The Functional Compensation Hypothesis (Hochber, 1986a,b) in relation to overt pronoun realization has been supported by some data (Poplack, 1980) but rejected by other data (Cameron, 1993, 1996). In Cameron's rejection of the hypothesis, he found that *-s* aspiration does not constrain the presence of a pronominal subject (1993) and that the higher percentage of Puerto Rican overt pronoun use was due to compensation only in case of [-specific] *tú*, but not in the case of [+specific] *tú* (1996). We would find support for the functional hypothesis in the current data if we find that there is less *-s* marking in cases of overt pronoun use. Contrastingly, we would find grounds to reject the hypothesis if we find more *-s* marking in cases of overt pronoun use.

3.2 Frequency of preterit form.

Again following the methodology of Barnes (2012), tokens were coded by verb frequency. In order to categorize verbs into those of high, medium, and low frequency, Barnes determined the calculated frequency of the preterit forms under study provided by CREA's "List of Frequencies." In this online list, each lexical item that appears in the corpus is assigned a calculated frequency based on the absolute appearance of that item. For the preterit forms of the verbs under question, Barnes found that the frequencies ranged in value from 0.5-13.41. She then split them into three groups; those with values from 0.5-1 were considered low frequency, those from 1-5 considered medium, and those with values above 5 considered high (For further information, consult Barnes, 2012, p. 41). High, medium, and low frequency preterit verbs from the tokens collected for this study are illustrated in examples 5, 6, and 7 with the standard variant illustrated in (a) and the non-standard in (b) of each example:

(5a) High:

Me **dijiste**: "Haz lo que quieras" y aquí estoy, esperándote.

'You told me: "Do what you want" and here I am, waiting for you'

(5b) High:

@xxx that's why you love talking to me lol and ya **dijistes** eh

'@xxx that's why you love talking to me lol and you said it eh'

(6a) Medium:

@xxx OH MY GOSH me **dejaste** loca

'@xxx OH MY GOSH you left me crazy'

- (6b) Medium:
 Hoy un año, un año desde que te perdimos, **dejastes** muchas cosas por hacer, pero sin embargo, a tus 31?
'Today a year, a year since we lost you, you left a lot of things pending, however, at age 31?'
- (7a) Low:
 Elastic girl? Te **casaste** con elastic girl?
'Elastic girl? You married elastic girl?'
- (7b) Low:
 @xxx bien hayy amigo ya te **casastes** y ahora estas pensando en el futuro
'@xxx good job buddy you got married and now you are thinking about the future'

3.3 Code-switching within tweet.

Code-switching, the phenomenon present in bilingual communities by which two or more languages come into contact and alternate at the level of clauses and sentences (Montes-Alcalá, 2000, p. 218), has been the topic of an increasing amount of research since the 1970s. Since that time, linguists have devoted considerable effort to challenge popular notions that equate code-switching with imperfect acquisition, and instead have suggested that this process not considered as a deficiency or anomaly, but rather a system that is constrained by specific grammatical and pragmatic conditions (Lipski, 2014, p. 24). However, few studies have investigated to what extent other variables, such as the one in question in this study, are constrained by code-switching.

In order to examine if the final -s variant appears to a higher degree among bilingual code-switchers, tokens were coded as either being exclusively in Spanish (8) or containing elements of English-Spanish code-switching, including intersentential switching (9a), intrasentential switching (9b), or intra-word switching (9d).

- (8) Exclusively Spanish:
 Feliz cumple Iara...ame las tortas que llevastes
'Happy birthday Iara...I loved the cakes you brought'
- (9a) Intersentential code-switching:

@xxx Shaki great quedastes I feel so proud to follow you for so many years.

'@xxx Shaki you turned out great I feel so proud to follow you for so many years'

(9b) Intrasentencial code-switching:

@xxx he dont want you no more ?? lo dejastes so he crushing on someone else ??

'@xxx he dont want you no more ?? you left him so he crushing on someone else ??'

(9d) Intra-word code-switching:

Rihanna bebé cuando twerkiastes me matastes

'Rihanna baby when you twerked you killed me'

In (9d), the tweeter uses only Spanish, but with one English word (“twerk”) adopted into Spanish morphology.

3.4 Directionality of tweet

The factor group of directionality considers whether a tweet is directed at one person in particular, or at an unspecified, public audience. This is related to Barnes' (2012) factor group of subject specificity, in which she categorized tokens as being ‘particular’ or ‘general’. ‘Particular’ NPs refer to a group of entities whose members are interchangeable, while particularizing NPs are used for unique referents. Many times, it is easy to determine if the verb is particular or general, but in certain cases, it can be unclear (10).

(10) Te acuerdas cuando a principios de semestre **dijiste**: "Este semestre si le voy a echar ganas"?

'Do you remember when at the beginning of the semester you said: "This semester I am going to give it my all"?''

In this case, it is difficult to judge if the person only has one referent in mind, or if the members are interchangeable. Is the speaker asking this question to one specific person, or to various, interchangeable classmates?

One advantage of Twitter is the ability to use the @ symbol to direct a comment at a specific person. Users can then carry on a conversation with a specific person using their public messages by “tagging” that person in their tweet. Due to this specific role of the @ symbol in the Twitter context, it can be used to measure directionality. In the current study, the tweet was considered to

be directed to a specific twitter follower if it tagged a person (11), whereas it was considered to be directed to the general twitter audience (12) in the absence of the tag.

(11) Specific:

@xxx @xxx ok, ya dijiste. :) Gotta go for real now. :) Good Night!
 '@xxx @xxx ok, you said it. :) Gotta go for real now. :) Good Night!'

(12) General:

No se pierde lo que no tuviste, no se mantiene lo que no es tuyo y no puedes aferrarte a algo que no se quiere quedar.
 'You don't lose what you didn't have, you don't maintain what is not yours and you can't hang on to something that doesn't want to stay.'

It is important to note that this factor group is related to that of Barnes (2012) and Cameron (1996) in terms of categorizing +/- specificity, but it does not measure the exact same linguistic information as those studies do. In many cases, a tweet without the @ symbol was seen to be obviously directed at only one person in particular even though the @ symbol was not used. This was prevalent in sentimental tweets, in cases in which the user was tweeting a negative message about another person, and in tweets with direct quotes.

3.5 Type of conjugation

Once again, following Barnes' (2012) methodology, all tokens were coded for their preterit conjugation: *–aste* (13) or *–iste* (14) with the goal of exploring how inflectional morphemes could potentially constrain the variable.

(13) *–aste* conjugation:

Dios Por que Me **Dejaste** Ver Esooo
 Buuuuuuuuuuuuuuuuuuuuuuuuuuuuu..!
 'God why did you let me see that Boooooooooooooooooooooo!'

(14) *–iste* conjugation:

We own the night. Si no la **viste**, mirala.
 'We own the night. If you didn't see it, watch it.'

3.6 Following phonological environment

Because *–s* aspiration and deletion have been found to be constrained by the following phonological context (Terrell, 1979), all the tokens were coded for

whether they were followed by a consonant (15), a vowel (16), or a pause (17), in order to examine whether the following segment also constrains the addition of a sound in word-final position. Pauses were identified when the following segment was a punctuation mark, a hashtag, the @ symbol, or the end of a tweet.

- (16) –s followed by consonant:

@xxx he dont want you no more lo **dejastes** so he crushing on someone else

'@xxx he don't want you no more. You left him so he's crushing on someone else'

- (17) –s followed by vowel:

Fui lo mejor que te pudo haber pasado y me **dejaste** ir.

'I was the best thing could have happened to you and you let me go.'

- (18) –s followed by pause:

#Heartbreaker #AllThatMatters #HoldTight #Recovery JUSTIN

BIEBER! si querias explotar mis ovarios!, Pues ya lo **hiciste** *-* ♥

#Heartbreaker #AllThatMatters #HoldTight #Recovery JUSTIN

BIEBER! 'If you wanted to make my ovaries explode, you just did it'

Weakening processes of /s/ in Spanish typically appear in implosive position; when /s/ is found in word-final position, it is more likely to be weakened before consonants than before vowels due to the resyllabification process and the universal tendency for Spanish to prefer CV syllables whenever possible (see Harris, 1983; Hualde, 1991a,b). However, since resyllabification is a typical process of rapid speech, it is unknown if the same patterns will be found within the written sphere.

4. Results and discussion

In the first part of this section, I will provide the results by factor group. I will then discuss the only factor group found by GoldVarb to be statistically significant: verb frequency.

Table 2 displays all of the results, including the raw number of tokens and percentages per factor. The factor groups and their individual factors are presented in the left hand column. In the subsequent columns are the raw number of standard variant tokens (N_{standard}), the percent of standard variant tokens within the factor group ($\%_{\text{standard}}$), the raw number of non-standard variant tokens ($N_{\text{non-standard}}$), the percentage of non-standard variant tokens within the factor group, the

total raw number of tokens per factor (T) and the total percentage (%), respectively.

Table 2. Raw number of tokens and their percentages by factor

Factor group/ factors	N_{standard}	%_{standard}	N_{non-standard}	%_{non-standard}	T	%
Presence of Pronominal Subject						
Tú	18	69.2	8	30.8	26	6.4
(Ø)	292	77.2	86	22.8	378	93.6
Verb Frequency						
High	174	91.6	16	8.4	190	47
Medium	80	75.5	26	24.5	106	26.2
Low	56	51.9	52	48.1	108	26.7
Code-switching						
Spanish only	212	77.7	61	22.3	273	67.6
Spanish/English	98	74.8	33	25.2	131	32.4
Directionality of Tweet						
Specific	128	75.3	42	24.7	170	42.1
General	182	77.8	52	22.2	234	57.9
Type of Conjugation						
-aste	75	56.8	57	43.2	132	32.7
-iste	235	86.4	37	13.6	272	67.3
Following Phonological Environment						
Consonant	138	77.1	41	22.9	179	44.3
Vowel	96	77.4	28	22.6	124	30.7
Pause	76	75.2	25	24.8	101	25.0

4.1 Presence of pronominal subject

As seen in Table 3, there was an overwhelmingly high percentage (93.6%) of tokens with the null pronoun as compared to the explicit one (6.4%). When comparing the occurrences of only the non-standard variant, we see that the percentage of *-s* used with the explicit pronoun (30.8%) is higher than the null (22.8%). These results do not support the Functional Compensation Hypothesis, which would predict that the explicit pronoun would be less frequent in cases with the non-standard *-s* due to increased redundancy in the second-person marking.

However, the presence of a pronominal subject was not found to be a significant factor in conditioning the non-standard variant.

4.2 Verb frequency

Verbs of high frequency were recorded in almost half of the tokens (47%), and verbs of medium and low frequency in about one quarter of the tokens each (26.2 and 26.7, respectively). With respect to use of the non-standard variant, we see a very clear pattern: verbs of high frequency reflected low numbers of the non-standard (only 8.4%), while verbs of low frequency reflected high numbers (48.1%). Medium frequency verbs fell in the middle (24.5%). In other words, as frequency drops, use of the non-standard variant increases, to the point of being almost 50-50 in low frequency verbs. This is an extremely interesting finding, especially due to the fact that the rate of use of the *-s* variant in low frequency verbs in this study was found to be almost twice the rate of that of Barnes' (2012) study. This factor group was found to be statistically significant factor group in conditioning the use of one variant over another and will be discussed in the following section.

4.3 Code-switching

Tweets that included code-switching were less common overall than those that were written in Spanish only (32.4% contained code-switching versus 67.6% contained only Spanish). If we consider only the non-standard tokens, we see that there is a very similar percentage of tweets written in Spanish only (22.3%) and with code-switching (25.2%). Considering that data was collected broadly from both monolingual and bilingual communities, the relatively high number of tweets containing code-switching demonstrates that it is an important linguistic tool used by bilingual speakers. According to Toribio (2003), despite the low prestige associated with code-switching, covert norms value the duality conveyed by the linguistic alternations. In other words, code-switching is foregrounded in the speech of some Latinos because it serves the important function of signaling social identity (p. 115). However, in terms of the relation between code-switching and use of preterit final *-s*, the relationship is less clear. The data suggests that using the non-standard variant in code-switching is slightly more common than using it without code-switching, but not enough to be considered a statistically significant factor. This suggests that code-switching does not condition the use of the non-standard variant and that final *-s* is not a highly salient feature of the bilingual speech reported in this study.

4.5 Directionality of tweet

Overall, there was a higher number of tweets directed at a general audience (57.9%) as compared to those directed at a specific person (42.1%). Interestingly, another person is tagged in nearly half of the tweets in the data, meaning that users are very involved in intercommunication and do not only use Twitter to post comments that are directed toward the greater public audience.

Taking into account only those tokens that contain the non-standard variant, we see very similar rates of the token being embedded in a general tweet (22.2%) and a specific tweet (24.7%). Using the non-standard variant in a tweet that is directed at a specific person is slightly more common than using it in a general tweet, but as in the case of code-switching, this difference is not enough to make specific directionality a statistically significant factor.

4.6 Following phonological context

Our data demonstrates that the segment most commonly found to follow the variable is a consonant (44.4%), followed by a vowel (30.7%), and then by a pause (25%). Interestingly, there was very similar stratification in standard and non-standard variant use between each of the three factors; the standard variant was found in roughly three times as common as the non-standard, independent of whether the following segment was a consonant, vowel, or pause. Looking at just the non-standard variant, we see that when the following segment was a consonant, the non-standard appeared in 22.9% of the tokens, when the following segment was a vowel, it appeared in 22.6% of the tokens, and when it was a pause, the non-standard appeared in 24.8% of the tokens. We see that when the following segment is a pause, it is more likely to find the non-standard variant than when the following segment is a consonant or vowel; however, this difference is minute and the factor group is not found to be a significant predictor of the variable.

The data related to this factor group is in contrast to that of Barnes (2012) oral data, who found that vowels were the most likely of any following segment to condition the use of the non-standard variant. Her results support the idea that Spanish prefers CV syllables whenever possible. The data presented in this study, however, perhaps since it is written language as opposed to oral, does not show the same tendency. The distribution of the standard and non-standard variants is similar regardless of the following segment.

4.7 Probabilistic weight of factor group frequency

We now turn our attention to the only group that emerged as statistically significant in the GoldVarb analysis: that of verb frequency. As stated previously,

the factor group frequency showed marked stratification in terms of preferring the standard variant in verbs of high frequency and the non-standard in those of low frequency. To obtain a more in-depth analysis of the factors contributing to the use of one variant over another, a Varbrul (variable rule) analysis can be used. Varbrul analyses allow researchers to be able to see the probabilistic weight in conditioning the use of one variant over another. Weights fall between 0 and 1, with weights of 0 signifying categorical use of variant x and weights of 1 signifying categorical use of variant y. Weights between 0 and 0.5 denote that the factor conditions the use of variant x and inhibits that of y; weights between 0.5 and 1 represent that the factor conditions the use of variant y and inhibits that of x. Table 3 contains the Varbrul weights associated with the factor group verb frequency.

Table 3. Conditioning of the non-standard variant of the factor group verb frequency

Verb frequency	Weight
High	0.279
Medium	0.577
Low	0.796

Log likelihood = -188.735

As seen in Table 3, high frequency verbs highly inhibit the use of the *-s* variant, while verbs of medium frequency slightly condition it and verbs of low frequency highly condition it. Although the data in Barnes' (2012) study show similar patterns, the stratification evident in the present study is even stronger than that of Barnes. Table 4 show her results of the same factor group.

Table 4. Conditioning of the non-standard variant of the factor group verb frequency in Barnes (2012).

Verb frequency	Weight
High	0.43
Medium	0.51
Low	0.66

The same pattern exists, with low frequency verbs favoring the *-s* variant most, followed by medium frequency, and then low frequency; however, the weights are not as polarized as they are in my results.

What is the reason for such stark polarization between high and low frequency verbs when it comes to second person singular preterit *-s*? As in Barnes' (2012) study, the data here also supports the lexical diffusion model proposed by Bybee (1998, 2000, 2002), which states that low frequency verbs may be subject to regularization and changes influenced by the general patterns of the language.

Bybee (2007, p. 29) examines six verbs in the past tense with a lax vowel (*creep*, *weep*, *leap*, *sleep*, *leave*, and *keep*) and finds that frequency is an indicator of paradigm leveling; in other words, the three low frequency verbs of the group are often re-shaped with a regular *-ed* ending (*creeped*, *leaped*, and *wept*) instead of the standard *crept*, *lept*, and *wept*) whereas the high frequency verbs (*sleep*, *leave*, and *keep*) are more likely to resist that re-shaping and are produced as *slept*, *left*, and *kept*, not **slepted*, **leaved*, and **kepted*.

This hypothesis seems to be supported by the data; however, it is still unclear as to why the Twitter data seems to be more polarized than the oral data of Barnes (2012). One possible explanation is that perhaps the mental representation is weaker among written exemplars as compared to among those spoken. Written exemplars, therefore, may possibly be subject to even further regularization patterns than those found in spoken language due to their low frequency and possible weak mental representation. To my knowledge, there has not been any research that explores the effect of the lexical diffusion model on language context (oral versus written language), thus further research that compares the effect of frequency in different contexts is encouraged in order to explore the possible relationship between mental representation, frequency, and context.

5. Conclusion

The current paper explores the role of six linguistic factors in the production of second person preterit *-s* in Spanish (*fuiste/fuistes*). Since there has only been one empirical study to my knowledge that looks at this linguistic phenomenon from a variationist perspective (see Barnes, 2012), the current study was intentionally designed similarly so that results could be compared. The main differences between the two is that the current study examines tokens collected from Twitter posts and includes bilingual data while that of Barnes examines oral data found in three different publically available corpora and only includes monolingual data.

The results of the current study show that the only factor found to be statistically significant in conditioning the use of one variant over the other is the factor group verb frequency. It was found that high frequency verbs tend to strongly condition the use of the standard variant while low frequency verbs tend

to strongly condition the use of the non-standard with medium verbs only very slightly conditioning the use of the non-standard. It is suggested that this is a result of the lexical diffusion model proposed by Bybee (1998, 2000, 2002) which states that lower frequency forms are more likely to undergo change due to their “weaker representation in the mental lexicon of the speaker” (Barnes, 2012, p. 44). These forms have a “higher propensity to be analyzed and...re-shaped on the basis of the most frequent pattern present in the Spanish verb paradigm” (p. 44).

Other factors investigated here were found to not be statistically significant in conditioning the use of one variant over the other; however, in order to discard them entirely, more research is needed. This is especially important with the factor group of code-switching. Several researchers have noted descriptively that second-person singular preterit *-s* is a tendency in Mexican-American Spanish (Hidalgo, 1990 and references within). It would seem, then, that the non-standard variant would be more commonly found in US Spanish than in other monolingual varieties. Since US Spanish often contains code-switching, it would seem that code-switching might condition the use of the non-standard. The results presented here did not find that code-switching was a significant factor, but it would be interesting to conduct further research on US Spanish to see if geographic location influenced second person singular preterit *-s*.

References

- Bamman, D., Eisenstein, J., & Schnoebelen, T. (2012). Gender in Twitter: Styles, Stances, and Social Networks. Forthcoming. <http://arxiv.org/abs/1210.4567>.
- Barnes, S. (2012). ¿Qué dijistes?: A Variationist Reanalysis of Non-standard *-s* on Second Person Singular Preterit Verb Forms in Spanish. In K. Geeslin and M. Díaz-Campos (Eds.), *Selected Proceedings of the 14th Hispanic Linguistics Symposium*. (pp. 38-47), Sommerville, MA: Cascadilla Proceedings Project.
- Bayley, R. (2013). Variationist Sociolinguistics. In R. Bayley, R. Cameron, & C. Lucas (Eds.), *The Oxford handbook of sociolinguistics* (pp. 11-30). Oxford University Press.
- Bybee, J. (1998). Usage-based Phonology. In M. Darnell, E. Moravcsik, F. Newmeyer, M. Noonan & K. Wheatley (Eds.), *Functionalism and Formalism in Linguistics* (pp. 211-242). Amsterdam; Philadelphia, PA: John Benjamin's.
- Bybee, J. (2000). The phonology of the lexicon: evidence from lexical diffusion. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp.

- 65-85). Stanford; California: Center for the Study of Language and Information.
- Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14, 261-190.
- Cameron, R. (1993). Ambiguous agreement, functional compensation, and non-specific tú in the Spanish of San Juan, Puerto Rico, and Madrid, Spain. *Language Variation and Change*, 5, 305-334.
- Cameron, R. (1996). A Community-Based Test of a Linguistic Hypothesis. *Language in Society*, 25(1), 61-111.
- Cerda, G, Cabaza, B, & Farias, J. (1953). *Vocabulario español de Tejas*. Austin: University of Texas Press.
- Davies, M. Corpus del español. Available at <http://corpusdelespanol.org/>
- Frago García, J. A. (2003). *El español de América*. Cádiz: Servicio de Publicaciones de la Universidad.
- Harris, J. (1983). *Syllable structure and stress in Spanish*. Cambridge, Massachusetts: The MIT Press.
- Hidalgo, M. (1990). On the Question of 'Standard' Versus 'Dialect:' Implications for Teaching Hispanic College Students. In J. Bergen (Ed.), *Spanish in the United States: Sociolinguistic Issues* (pp. 110-126). Washington, D.C.: Georgetown University Press.
- Hochberg, J. (1986). Functional compensation for /s/ deletion in Puerto Rican Spanish. *Language*, 62, 609-629.
- Hualde, J. (1991a). On Spanish syllabification. In H. Campos & F. Martínez-Gil (Eds.), *Current studies in Spanish Linguistics* (pp. 475-493). Washington, D.C. Georgetown University Press.
- Hualde, J. (1991b). Aspiration and resyllabification in Chinato Spanish. *Probus*, 3(1), 55-76.
- Junco, R., Heiberger, G., & Loken, E. (2011). The Effect of Twitter on College Student Engagement and Grades. *Journal of Computer Assisted Learning*, 27, 119-132.
- Labov, W. (1963). The social motivation of a sound change. *Word*, 19, 273-309.
- Labov, W. (1966). *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- Labov, W. (1969). Contraction, deletion, and inherent variability of the English copula. *Language*, 45, 715-62.
- Labov, W. (1972a). *Language in the inner city: Studies in the Black English vernacular*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (1972b). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

- Lance, D. (1975). Dialectal and Non-standard forms in Texas Spanish. In E. Hernández-Chavez, A. Cohen & A. Beltramo (Eds.), *El Lenguaje de los Chicanos* (pp. 37-51). Arlington, CA: Center for Applied Linguistics.
- Lapesa, R. (1965). *Historia de la lengua Española, sixth edition*. Madrid: Escelicer, S.A.
- Lipski, J. (2002). *El español de América*. Fuenlabrada, Madrid: Cátedra.
- Lipski, J. M. (2014). Spanish-English code-switching among low-fluency bilinguals: Towards an expanded typology. *Journal of Sociolinguistic Studies*, 8(1), 23-55.
- Lope Blanch, J. (1976). El habla popular de México: materiales para su estudio. México: Universidad Nacional Autónoma de México (UNAM).
- Mendez-Pidal, R. (1962). *Española. Diccionario manual e ilustrado de la lengua Española, eleventh edition*. Madrid: Espasa-Calpe, S.A.
- Montes-Alcalá, C. (2000). Attitudes Towards Oral and Written Code-switching in Spanish- English Bilingual Youths. In A. Roca (Ed.), *Research on Spanish in the United States* (pp. 218-227). Somerville: Cascadilla Press.
- Poplack, S. (1980). Deletion and disambiguation in Puerto Rican Spanish. *Language*, 56(2), 371- 385. RAE. CREA: Corpus de Referencia del Español Actual. Available at <<http://corpus.rae.es/creanet.html>>
- Sankoff, D., Tagliamonte, S. & Smith, E. (2005). GoldVarb X: a variable rule application for Macintosh and Windows. Department of Linguistics, University of Toronto.
<<http://individual.utoronto.ca/tagliamonte/goldvarb.html>>
- Terrell, T. (1979). Final /s/ in Cuban Spanish. *Hispania*, 62, 599-612.
- Toribio, A. J. (2003). Spanish-English code-switching among US Latinos. *International Journal of the Sociology of Language*, 158, 89-119.
- Vaquero de Ramírez, M. (1998). *El español de América II: Morfosintaxis y léxico*. Madrid: Arco Libros.
- Weinreich, U., Labov, W. & Herzog, M. (1968). Empirical foundations for a theory of language change. In W. Lehmann & Y. Malkiel (Eds.), *Directions for Historical Linguistics* (pp. 95-195). Austin: U. of Texas Press.