Journal of Child Language (2022), **49**, 1256–1267
doi:10.1017/S0305000921000507

CAMBRIDGE
UNIVERSITY PRESS

BRIEF RESEARCH REPORT

¡Casi te caístes!: Variation in second person singular preterit forms in Spanish Children

Elisabeth BAKER

University of New Mexico
Address for correspondence: Elisabeth Baker University of New Mexico Spanish & Portuguese MSC03 2100 1 University of New Mexico Albuquerque, NM 87131-0001 USA E-mail: bakere@unm.edu

(Received 8 May 2020; revised 20 February 2021; accepted 2 June 2021; first published online 23 July 2021)

Abstract

The current study investigates Spanish children's variation between the standard and non-standard forms for second person singular preterit –s (*caiste* ~ *caístes*). All second person singular preterit forms were extracted from the spontaneous speech of 78 children in Spain and analyzed for the effects of age, language contact setting, and lexical frequency. Results show that children in contact with Galician and Catalan produce more non-standard than children in non-contact areas like Madrid. Meanwhile, low-frequency verbs (e.g., *pillaste*) are more likely to occur with the non-standard –s than high-frequency verbs (e.g., *fuiste*). However, age is not a significant predictor of children's 2sg preterit production. These preliminary findings demonstrate that Spanish children do have the non-standard -s in their speech, and that their 2sg preterit forms are significantly conditioned by language contact and lexical frequency.

Keywords: morphosyntactic variation; acquisition of sociolinguistic variation; developmental sociolinguistics; acquisition of Spanish morphosyntax

Introduction

While there has been extensive research on adults' sociolinguistic variation (Labov, 2001), children's sociolinguistic variation is less well-understood; this is particularly the case for sociolinguistic variation involving morphosyntax. Thus far, the few studies investigating children's acquisition of morphosyntactic variation have yielded conflicting results. Some studies suggest that children's variable morphosyntactic patterns match those of their caregivers by age three (e.g., Smith & Durham, 2019), whereas others suggest that acquisition of morphosyntactic variation can take a long time to emerge during childhood (e.g., Shin, 2016). This developmental trajectory and the age at which children master a structure varies depending on the structure's complexity and its frequency in the input in addition to its perceived prestige in the community. For example, Smith and Durham (2019) note that caregivers are reluctant to use stigmatized variants with young children, and instead opt for the socially recognized standard form in Child Directed Speech (CDS). As such,

© The Author(s), 2021. Published by Cambridge University Press

https://doi.org/10.1017/S0305000921000507 Published online by Cambridge University Press

stigmatized morphosyntactic variants may develop later in children's output in comparison to other less salient forms.

This brief research report contributes to the growing field of developmental sociolinguistics by zeroing in on a sociolinguistic variable in Spanish that is well-studied and well-understood among adults, but not yet well-understood in child language. More specifically, the current study investigates Spanish children's variation between the second person singular (2sg) preterit standard forms (e.g., *caíste* 'you fell') and the non-standard forms, which include word-final -s (e.g., *caístes*). To do so, preterit forms were extracted from corpora of spontaneous speech produced by 78 children in Spain and analyzed for the effects of age, lexical frequency, and the possible influence from Catalan or Galician.

Background

In Spanish, the perfective past tense is often called the 'preterit'. The standard 2sg preterit inflection is *–ste*. As observable in Table 1, 2nd personal singular forms generally include the 2sg word-final –s morpheme, making this paradigm different from all other 2sg verbal paradigms and thus an irregular pattern. Moreover, Miller and Schmitt (2012) suggest that Spanish-speaking children associate the 2sg word-final –s morpheme to 2sg subjects by age 3;00.

The non-standard forms, which include word-final –s (e.g., *hablastes*, *comistes*, *vivistes*), are an example of analogical leveling; speakers regularize forms so that they follow the dominant pattern of the language (Bybee, 2015; Penny, 2000). While several studies show that Spanish-speaking adults vary between standard 2sg preterit forms, like *caíste*, and non-standard forms, like *caístes*, their use of the non-standard is constrained by social factors such as socioeconomic status and education level. Importantly, non-standard preterit –s is a stigmatized variant in many speech communities (Lipski, 2014; Tinoco, 2017). For example, Barnes' (2012) study, which draws on data from sociolinguistic interviews, television broadcasts, and public events, highlights the importance of speech register, with the non-standard showing a strong correlation with informal speech. This finding aligns with the general association found between informal speech and less-prestigious variants (Smith & Durham, 2019).

Variation between standard and non-standard -s is also influenced by lexical frequency. Both Barnes (2012) and Escalante (2015) find that non-standard –s is more prevalent among low-frequency (23%, 48%, respectively) than high-frequency verbs (10%, 8%, respectively) in adult Spanish. Finally, previous studies of adults' 2sg preterit forms demonstrate varying rates of the non-standard –s across Spanish dialects with its acceptability relating to the speakers' variety (Aleza & Enguita, 2010; Lipski, 2014). While Barnes (2012) finds rates as low as 13% in an adult Spanish population, Lerner's (2016) adult Mexican population in the US produces non-standard –s at a rate of 33%. These varying rates will likely impact children's production of the non-standard -s. Indeed, previous research on developmental sociolinguistics suggests that children are more likely to show variation in their output when exposed to it in their input (Smith & Durham, 2019; Miller, 2013).

In addition to dialect variation, an open question is whether language contact affects speaker's variation between standard and non-standard –s. Previous research highlights the importance of crosslinguistic influence (CLI) in language contact settings noting that variable morphosyntactic structures are more susceptible to CLI (Silva-Corvalán,

https://doi.org/10.1017/S0305000921000507 Published online by Cambridge University Press

Table 1. 2sg verb conjugations in simple tenses

Verbal tense	Hablar	Comer	Vivir
Present	Hablas	Comes	Vives
Imperfect	Hablabas	Comías	Vivías
Future	Hablarás	Comerás	Vivirás
Conditional	Hablarías	Comerías	Vivirías
Preterit	Hablaste	Comiste	Viviste

2014). CLI can manifest in the form of accelerated or delayed acquisition of a construction or the transfer of an element from one language to another (Paradis & Genesee, 1996). Thus, while Barnes (2012) notes low percentages of the non-standard –s in Peninsular Spanish, Spanish speakers in Galicia and Catalonia may produce more non-standard 2sg preterit forms due to contact with Galician and Catalan, two languages in which the standard 2sg preterit forms include word final -s.

To date, there has been one study of Spanish-speaking children's variation between standard and non-standard -s. In her study of Mexican children from three to eight-years-old, Lerner (2016) finds that children use the non-standard variant at a rate (61%) that exceeds that of their caregivers. She also finds that both children and adults alike produce higher rates of the non-standard in informal than formal speech. Lerner observes an age effect whereby older children produce more non-standard –s than younger children. These results are consistent with previous research in developmental sociolinguistics showing children use increasingly more non-standard forms as they get older, a finding that has been explained as the result of caregivers' tendency to use fewer non-standard forms with younger children and more with older children and adults (Smith & Durham, 2019). The increased use of non-standard forms with age could also be the result of increased exposure to more verb forms, including low-frequency forms, as children age. While Lerner (2016) studies the factors influencing children's 2sg –s variation, she does not analyze the effect of lexical frequency, which is known to condition adults' variable 2sg preterit production in Spanish. Further, lexical frequency is known to impact child language: children overregularize irregular forms that are low-frequency, whereas high-frequency forms are less likely to undergo overregularization (Clahsen, Aveledo & Roca, 2002; Fernández-Dobao and Herschensohn, 2020; Johnson, 1995). How does lexical frequency impact children's use of the non-standard preterit –s in Spanish? To date, there is no study that has examined this question.

The current study aims to further our understanding of children's acquisition of sociolinguistic variation by analyzing Spanish children's 2sg preterit verb forms and the effects of age, contact setting, and lexical frequency. It is predicted that children, like adults, will produce both '*caiste*' and '*caistes*', and that their variation will be impacted by age, language contact, and lexical frequency.

Research questions and hypotheses

The following research questions and their corresponding hypotheses were developed based on previous research:

https://doi.org/10.1017/S0305000921000507 Published online by Cambridge University Press

1) Does age affect children's variation between standard and non-standard 2sg preterit forms?
   H1a) Based on Lerner's (2016) findings that older Spanish-speaking children produce more non-standard –s than younger children, as well as Smith and Durham's (2019) finding that caregivers use more non-standard forms in CDS as children age, it is predicted that older children will produce higher rates of non-standard –s.
   H1b) Based on findings that Spanish-speaking children associate the 2sg word-final –s to 2nd person subjects as early as 3;0 (Miller & Schmitt, 2012), it is predicted that younger children will overregularize the preterit forms and produce more non-standard –s.
2) Do Spanish children in language contact settings produce more non-standard –s than those in non-contact settings?
   H2a) If language contact influences 2sg preterit variation as has been observed in other variable morphosyntactic structures (Silva-Corvalán, 2014), children in Galicia and Catalonia are predicted to transfer the Galician and Catalan 2sg –s morpheme and thus produce higher rates of the non-standard forms in Spanish.
   H2b) Given the overall low rates of non-standard –s in Peninsular Spanish (Barnes, 2012), all Peninsular speakers are expected to produce low rates of non-standard –s.
3) Does lexical frequency impact children's production of non-standard –s?
   H3a) It is predicted that children will produce non-standard –s more often with low-frequency verbs. This outcome could reflect patterns found in the input since adults produce more non-standard preterit –s with low-frequency verbs (Barnes, 2012; Escalante, 2015) or could be the result of children's own experience with verb forms since children tend to overregularize more with low-frequency verbs (Clahsen et al., 2002; Johnson, 1995).
   H3b) A strong tendency to overregularize could override children's lexical learning of frequent forms, in which case children will produce non-standard –s on both high and low-frequency verbs.

Methodology

Participants and data extraction

Data for this study come from 12 Spanish-language corpora from the Child Language Data Exchange System (CHILDES) Talkbank database (MacWhinney, 2000). A total of 78 Peninsular Spanish speakers between the ages of 1;10 and 12;4 years were included (see Table 2).

In the 'other' category, four corpora were grouped together due to their low token count of 2sg preterit forms: Aguirre (2000; Madrid), Ornat (1994; Madrid), Linaza (1981; Madrid), and Vila (1990; Catalonia). The 12 corpora utilized included longitudinal and cross-sectional data consisting of naturalistic speech between children and caretakers/researchers, child play, and more formal production tasks such as narrating a story. All tokens of standard and non-standard 2sg preterit forms (N = 230) were extracted manually from the CHILDES transcripts. Tokens followed by an alveolar fricative must be excluded from studies of word-final -s, as this phonological context would make it difficult to distinguish between the 2sg preterit –s and the following consonant. However, there were no such contexts in the current data set.

https://doi.org/10.1017/S0305000921000507 Published online by Cambridge University Press

Table 2. Number of children and tokens of 2sg preterit forms per CHILDES corpus

Corpus	Location	N children	Age range	N tokens
BecaCESNo	Madrid	24	3;6 –12;4	39
Koine	Galicia	24	2;2–4;4	52
DiezItza	Asturias	12	3;0–3;10	29
FernAgudo	Navarre	8	2;11–4;5	15
FerFuLice	Castile-Leon	2	4;4–5;7	5
LlinasOjea	Asturias	1	1;10–3;2	65
Marrero	Canary Islands	2	3;3–4;7	10
OreaPine	Madrid	1	2;0–2;5	6
Other	Other	4	2;4–4;9	9
Total		78	1;10 –12;4	230

Note: BecaCESNo = CHILDES Databate BecaCESNO Corpus (no date), Koine = Fernández Pérez (2015), DiezItza = Diez-Itza (1995), FernAguado = CHILDES Database FernAguado Corpus (no date), FerFuLice = Liceras et al. (2008), LlinasOjea = CHILDES Database LlinasOjea Corpus (no date), Marrero = Cappelli et al. (1994), OreaPine = Aguado-Orea & Pine (2015).

Coding

All 2sg preterit tokens were coded for whether the form was the standard form without –s (1) or the non-standard form with –s (2).

(1) *Me metiste aquí dentro.* (Alejandro, 4;00; Corpus: FernAgudo)
'You put me here inside.'
(2) *¡Me pillastes!* (Rosa, 7;11; Corpus: BecaCESNo)
'You found me!'

Furthermore, in order to address the research questions, all tokens were coded for (i) the participant's age, (ii) whether the children were in a language contact setting, and (iii) lexical frequency.

Age

Children ranged from 1;10 to 12;4 years. To address possible age effects on 2sg preterit production, children were split into three groups:

- 1;10-3;11 (N = 46)
- 4;0-6;7 (N = 24)
- 7;1-12;4 (N = 12)

Contact setting

To address language contact effects, children were categorized in two groups: 1) contact and 2) non-contact. Children in the contact group (N = 25) came from Galicia and Catalonia, while the non-contact children (N = 53) came from Madrid, Navarre, Castile-Leon, The Canary Islands, and Asturias. Children were grouped in this manner because both Galician and Catalan, which are official languages alongside Spanish in their respective provinces, mark –s on 2sg preterit forms.

https://doi.org/10.1017/S0305000921000507 Published online by Cambridge University Press

Table 3. Frequency categories and examples with frequency counts from Davies' corpus

Category	Frequency in Davies' corpus	Examples ranging from least to most frequent
Low	< 5,000	Mancaste (0), Viniste (1,561), Pusiste (2,524)
Mid	5,001–10,000	Diste (5,887), Tuviste (8,620), Dijiste (9,441)
High	>10,000	Hiciste (18,363), Fuiste (19,870), Viste (46,505)

Lexical frequency

Lexical frequency was coded for each 2sg preterit verb form produced by the children by obtaining frequency counts for each token's standard preterit form (*-ste*) in the Davies' (2018) Corpus del Español NOW. The most frequent form was *viste* 'you saw' with a total of 46,505 examples in Davies' corpus. The second most frequent form was *fuiste* 'you went'/'you were' with 19,870 tokens. There was one token extracted from the CHILDES corpora that did not appear at all in the Davies' corpus: *mancaste* 'you maimed'. The Davies' corpus is updated with new data monthly; as such, the frequency counts increase quite steadily. Therefore, the coding for frequency was completed within one month to gain an accurate comparison of frequency categories. Following Barnes (2012), frequency was split into three categories: low, mid, and high (Table 3).

Low-frequency verbs accounted for 55% (N = 126) of the overall tokens, with mid and high together accounting for the remaining 45% (12%, 33%, respectively). The following examples show children's production of a low-frequency verb (5), a mid-frequency verb (6), and a high-frequency verb (7).

(5) *¿Y con qué la pintastes (.) con boli?* (Raquel, f, 3;06; Corpus: BecaCESNo)
'What did you paint it with (.) with a pen?'
(6) *No me lo distes.* (Alejandro, m, 4;00; Corpus: FernAguado)
'You didn't give it to me.'
(7) *No viste el otro día esa película de…*(Sandra, f, 6;05; Corpus: BecaCESNo)
'You didn't see that movie the other day about…'

Results

Overall distribution

The overall rate of non-standard was 17% (N = 38) out of a total of 230 2sg preterit forms. Thus, while the non-standard forms were indeed present, the sampled children produced the standard form more frequently. Table 4 demonstrates the children's distribution according to their age group and contact setting and their rates of non-standard –s. Meanwhile, Table 5 presents the distribution of tokens across frequency categories

Table 4 demonstrates varying non-standard rates across age groups, with the oldest children in non-contact settings producing an increased rate of non-standard –s (31%). Further, children in the two younger age groups residing in contact areas have comparable non-standard rates (28%, 27%,). At the same time, Table 4 displays a high degree of variance especially within categories with low 2sg preterit token counts (e.g., Contact: 4;0-6;7; Non-contact: 7;1-12;4). Lastly, Table 5 shows a frequency effect whereby low-frequency verbs are produced as the non-standard variant most often, whereas high-frequency verbs almost always occur in their standard form.

https://doi.org/10.1017/S0305000921000507 Published online by Cambridge University Press

Table 4. Distribution of children classified by contact setting and age

Contact setting	Age group	N children	N tokens	%non-standard	SE
Contact	1;10-3;11	22	45	28	9.0
	4;0-6;7	5	11	27	13.3
Noncontact	1;10-3;11	24	117	6	6.2
	4;0-6;7	19	41	24	9.0
	7;1-12;4	12	16	31	14.2

Note. While the total number of participants in the current study is 78, there are N = 82 children accounted for in Table 5 as four children produced tokens at different periods spanning the two youngest age groups.

Table 5. Distribution of tokens across frequency categories

Frequency category	N tokens	%non-standard	SE
Low	126	24	3.8
Mid	27	11	6.1
High	77	6	2.8

To investigate whether the children's variable 2sg preterit production is significantly influenced by (i) age, (ii) contact setting, and (iii) lexical frequency, a mixed-effects binary logistic regression was run in R (R Core Team, 2019) using the glmer function in the lme4 package (Bates, Mächler, Bolker & Walker, 2015). The regression analysis included data from all 78 children with non-standard vs. standard 2sg preterit forms as the dependent variable with the model set to predict non-standard. The final mixed-effects model (Table 6), which included participant as a random factor to ensure generalizability beyond individual, included the following categorical independent variables: age group, contact setting, and lexical frequency. Multicollinearity was investigated using the vif function in the car package in R (Fox & Weisberg, 2019), and revealed no issues; Variance Inflation Factors (VIFs) ranged from 1.00 to 1.09. Further, there were no interactions found between predictor variables – thus, the final model presents main effects only. Reference levels for predictor variables are labeled "Ref".

Results from the mixed-effects analysis (Table 6) show that contact setting and lexical frequency significantly predict non-standard –s with children residing in contact areas and low-frequency verbs increasing the likelihood of producing the non-standard –s. Contrarily, there is no significant effect of age on children's 2sg preterit forms.

Discussion

Age results

The first research question explored the relationship between age and children's variable 2sg preterit production. Based on Lerner's (2016) results and caregiver's tendency to use fewer non-standard forms with younger children (Smith & Durham, 2019), H1a predicted that older children would produce the non-standard more than younger children while H1b predicted that as Spanish-speaking children associate 2sg word-final

https://doi.org/10.1017/S0305000921000507 Published online by Cambridge University Press

Table 6. Mixed Effects logistic regression analysis, 78 participants, AIC = 191.4

Fixed Effects		Estimate Std.	Std. Error	Z value	p
(Intercept)		−3.38	0.69	−4.9	< .0001
Age	1;10–3;11 (Ref)	–	–	–	–
Age	4;0–6;7	0.81	0.58	1.4	0.16
Age	7;1–12;4	1.42	.84	1.69	0.09
Contact setting	Noncontact (Ref)	–	–	–	–
Contact setting	Contact	1.18	0.58	2.01	0.04
Frequency	High (Ref)	–	–	–	–
Frequency	Mid	0.38	0.89	0.43	0.66
Frequency	Low	1.31	0.59	2.24	0.03

–s with 2sg subjects as early as 3;0 (Miller & Schmitt, 2012), young children would overregularize and produce more non-standard preterit –s. A mixed-effects regression analysis (Table 6) found no significant age effect suggesting that age is not a significant predictor of Spanish children's 2sg preterit variation. However, there were age groups with fewer data than others (see Table 4). For example, the oldest age group included 12 children all from non-contact settings. While this age group was underrepresented in the current study, they notably produced the highest rates of non-standard (31%).

Thus, it is possible that with more data from older participants, any age effects would become more discernible. Indeed, there are reasons to believe that as children grow older, they are exposed to more non-standard –s in the input as adults tend to use more non-standard forms with older children, which is especially true for stigmatized variants (Smith & Durham, 2019). Furthermore, Lerner's (2016) results demonstrate a developmental trajectory in this same direction with the older children producing more non-standard forms. Therefore, while Spanish children acquire imperfect and preterit tense forms at an early age (Pérez-Pereira, 1989), these results could signal a later developmental trajectory of 2sg preterit variation in relation to other sociolinguistic variants.

Contact results

The second research question explored the relationship between contact setting and children's 2sg preterit forms. Hypothesis 2a predicted that if 2sg preterit variation was influenced by language contact as has been observed in other variable morphosyntactic structures (Silva-Corvalán, 2014), children in contact settings would transfer the 2sg –s morpheme and produce higher rates of the non-standard forms in Spanish since 2sg preterit forms in Galician and Catalan end in –s. Contrarily, H2b predicted that all Peninsular speakers would produce low rates of the non-standard given the overall low rates in Peninsular Spanish: while Barnes (2012) found relatively low rates of non-standard -s in Spanish adults (13%), the results in Table 6 demonstrate that Spanish children in contact with Galician and Catalan

https://doi.org/10.1017/S0305000921000507 Published online by Cambridge University Press

produce the non-standard –s proportionately more than children in non-contact areas (e.g., Madrid), which corroborates H2a.

These preliminary findings highlight the importance of language contact on the production of 2sg preterit forms in Spanish. Although anecdotal evidence abounds suggesting non-standard –s is frequent in Galicia and Catalonia (e.g., Cousillas, 2019), we were unable to find empirical studies directly comparing rates of usage in these varieties with other varieties in Spain. Thus, the current study provides empirical support for what has heretofore been a widespread, but understudied phenomenon. The age and contact results taken together suggest that increased exposure to the non-standard –s, whether due to language contact with Galician and Catalan, or due to more exposure to verb forms containing the non-standard –s as children age, leads to higher rates of the non-standard –s in children's speech.

Lexical frequency results

The final research question explored the relationship between lexical frequency and children's 2sg preterit forms. Hypothesis 3a predicted that children would produce non-standard –s more often with low-frequency verbs. In contrast, H3b predicted that a strong tendency to overregularize could override children's lexical learning of frequent forms, in which case children who produce non-standard –s would do so with both high- and low-frequency verbs. The results in Tables 5 and 6 confirm H3a evidencing that low-frequency verbs such as *pintaste* 'you painted' and *abriste* 'you opened' occurred more with the non-standard -s while high-frequency verbs such as *viste* and *fuiste* favored the standard forms. These results lend support to the Conserving Effect (Bybee, 2007) suggesting that children learn high-frequency verbs as entrenched forms; the more entrenched a form is in the child's mental repertoire, the less likely it is to be overregularized. These findings further suggest that when children overregularize, they do so by extending a productive morphological pattern to the low-frequency verbs. That is, low-frequency forms like *pillaste* 'you caught' are targeted during the regularization process (Bybee, 2015; Clahsen et al., 2002).

The lexical frequency effect is further exemplified by the 38 non-standard tokens, of which 30 were low-frequency verbs, three mid-frequency, and five high-frequency. It is also worth noting that none of these non-standard tokens was a collocation, such as *te diste cuenta* 'you realized'. In fact, there was only one lexically bound collocation, *tuviste miedo* 'you were afraid', which occurred once and with *tuviste* in the standard form. While there were no other lexically bound collocations, the verb *viste* was frequently followed by *que* 'that' and *como* 'how'. *Viste* in these contexts was never produced as *vistes*. Similarly, *dijiste que* 'you said that' occurred four times, always with the standard variant (8). Meanwhile, when used without *que*, *dijiste* occurred with the non-standard –s on two occasions (9).

(8) *No me dijiste que me ibas a llevar.* (Luis, m, 6;07; Corpus: BecaCESNo)
'You didn't tell me that you were going to take me.'

(9) *Pero como no me dijistes el otro día.* (Virginia, f, 8;09; Corpus: BecaCESNo)
'But since you didn't tell me the other day.'

While there were insufficient collocations to arrive at firm conclusions, the lack of non-standard –s observed in units like *dijiste que* supports the idea that highly frequent strings of words are learned as a chunk and therefore resist

https://doi.org/10.1017/S0305000921000507 Published online by Cambridge University Press

overregularization of a form internal to the unit (Bybee & Hopper, 2001). In this sense, frequent collocations show an effect that is similar to the one found for high-frequency verbs. In summary, the lexical frequency results support the conclusion that high-frequency, entrenched forms are protected from overregularization.

Closing remarks and future research

This preliminary study investigated Spanish children's 2sg preterit variation examining the effects of age, contact setting, and lexical frequency. Results demonstrate that Spanish children, like adults, do in fact vary between non-standard and standard 2sg preterit forms with non-standard –s produced at a rate of 17%. Results from mixed-effects regression analyses indicate that contact setting significantly affects children's non-standard –s production. Specifically, those in contact settings (Galicia, Catalonia) are likelier to produce non-standard –s. A question that remains open for exploration, however, is whether these findings reflect CLI from the children's own L1 (Galician or Catalan) or increased exposure to non-standard –s in the Spanish spoken in contact areas. Additionally, lexical frequency predicts children's 2sg preterit forms: low-frequency verbs are likelier to occur with non-standard –s than high-frequency verbs. Contrarily, mixed-effects results (Table 6) indicate no age effect. While age was not a significant predictor of 2sg preterit variation in the current data set, there were limitations regarding the distribution of participants across age groups and contact settings as the oldest age group included 16 children who all resided in the non-contact group (Table 6). Since contact setting was a significant predictor of non-standard –s, it is important to include data from older children in contact settings for more representative data for contact setting and to better perceive the effects of age. Indeed, based on Lerner (2016), there is reason to believe that 2sg –s variation is more prevalent among older children.

The current study's findings provide areas of interest for future research – in addition to incorporating more children across age groups, future studies should address the effect of register and the age at which children acquire adult-like constraints on 2sg preterit variation by incorporating a balance between formal and informal speech. Due to the disparity between formal and informal settings, it was difficult to address this question in the current study. The majority of corpora utilized consisted of more informal speech situations such as child play or conversation between family members and researchers. However, one corpus incorporated more formal speech acts such as narrating a story, and there was a preference for the standard form in this setting with all but one token being the standard variant (N = 28).

Furthermore, while the current study did not analyze input effects, this factor merits future attention. This is especially important considering that children's production of the non-standard –s presents an interesting conundrum: it could be attributed to an input effect, but could also be due to children's tendency to regularize irregular forms (Clahsen et al., 2002; Johnson, 1995). Previous research on developmental sociolinguistics showing that children's variation mirrors their input predicts that children exposed to both standard and non-standard 2sg preterit forms in the input will produce both forms and learn the social and linguistic factors that constrain adults' variation (Lerner, 2016; Miller, 2013; Smith & Durham, 2019). While it is expected that children's variation between 2sg preterit forms is influenced by the variation found in their input (e.g., Lerner, 2016) analyses of input were beyond the scope of the current study due to the nature of the data.

https://doi.org/10.1017/S0305000921000507 Published online by Cambridge University Press

One limitation of corpus studies like the one reported here is that 2sg forms are not very frequent, and 2sg past tense forms are even rarer. Although the spontaneous conversations included CDS, the scarcity of 2sg preterit forms directed at the children prevented meaningful analysis of input effects. The lack of input data in the current study leaves us with the question of whether children's non-standard 2sg preterit –s forms are the result of acquisition of sociolinguistic variation – that is, learning from the variation present in the input – or the outcome of the general tendency to regularize irregular forms during childhood. Children's general tendency to regularize may result in the production of non-standard –s even among children who are not exposed to this form in the input, provided they have acquired 2sg –s in other verbal paradigms such as the present simple (e.g., *caes* 'you fall') and imperfect tenses (e.g., *caías* 'you were falling'). Future studies that include input data in addition to child language production data will help tease apart the effects of input and regularization on Spanish-speaking children's 2sg preterit variation.

Acknowledgements. I would like to extend my deepest gratitude to Dr. Naomi Shin for all her support throughout this process. Thank you, Naomi. Your guidance and encouragement have been invaluable in helping me prepare this manuscript. I would also like to thank Dr. Eva Rodríguez-González, Dr. Jill Morford, and Dr. Karen Miller for their input and thoughtful comments on this manuscript. Lastly, I would like to thank two anonymous reviewers for their helpful feedback.

References

Aguado-Orea, J., & Pine, J. M. (2015). Comparing different models of the development of verb inflection in early child Spanish. *PloS one*, *10*(3), e0119613.

Aguirre, C. (2000). *La adquisición de las categorías gramaticales en español*. Ediciones de la Universidad Autónoma de Madrid.

Aleza, M., & Enguita, J. M. (2010). *La lengua española en américa: Normas y usos actuales*. Valencia: Universitat de València.

Barnes, S. (2012). ¿Qué dijistes?: A variationist reanalysis of non-standard –s second singular preterit verb forms in Spanish. *Selected proceedings of the 201 Hispanic Linguistic Symposium, Somervile*, MA: Cascadilla press, 38–47.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bybee, J., & Hopper, P. (2001). *Frequency and the emergence of linguistic structure.*

Bybee, J. L. (2007). *Frequency of use and the organization of language*. Oxford: Oxford University Press, 2007. 5–22.

Bybee, J. L. (2015). *Language change*. Cambridge: Cambridge University Press.

Cappelli, G., Marrero-Aguiar, V., & Albalá, M. J. (1994). Aplicación del sistema MORFO a una muestra de lenguaje infantil. *Procesamiento del Lenguaje Natural. 14*, 23–32.

CHILDES Database BecaCESNo Corpus. doi:10.21415/T5ZG77

CHILDES Database LlinasOjea Corpus. doi:10.21415/T5K01Q

CHILDES Database FernAguado Corpus. doi:10.21415/T5MK75

Clahsen, H., Aveledo, F., & Roca, I. (2002). The development of regular and irregular verb inflection in Spanish child language. *Journal of Child Language*, *29*(3), 591–622.

Cousillas, J. (2019). *El castellano de Galicia. Una propuesta didáctica para el aprendizaje de ELE*. Masters thesis. University of Santiago de Compostela.

Davies, M. (2018). Corpus del Español NOW: 6.2 billion words, 21 countries. Available online at https://www.corpusdelespanol.org/now (2012-present)

Diez-Itza, E. (1995). Procesos fonológicos en la adquisición del español como lengua materna. In J.M. Ruiz, P. Sheerin, & E. González-Cascos (Eds.), *Actas del XI Congreso Nacional de Linguistica Aplicada*. Valladolid:Universidad de Valladolid.

https://doi.org/10.1017/S0305000921000507 Published online by Cambridge University Press

Escalante, C. (2015). ¿Qué twiteastes tú? Variation in second person singular preterit –s in Spanish tweets. *Entrehojas: Revista de Estudios Hispánicos 5*(1), 1–18.

Fernández Pérez, M. (2015). Lenguaje infantil y medidas de desarrollo, *ENSAYOS* (monográfico sobre Educación infantil), *30*/2, 53–69.

Fernández-Dobao, A., & Herschensohn, J. (2020). Acquisition of Spanish verbal morphology by child bilinguals: Overregularization by heritage speakers and second language learners. *Bilingualism: Language and Cognition* 1–13.

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression*, Third edition. Sage, Thousand Oaks CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Johnson, C. M. (1995). Verb errors in the early acquisition of Mexican and Castilian Spanish. In E.V. Clark (ed.) *The proceedings of the twenty-seventh annual child language research forum, Center study language & information*, 175–186

Labov, W. (2001). *Principles of Linguistic Change: Social Factors*. Oxford: Blackwell.

Lerner, M. (2016). The acquisition of sociolinguistic variation in a Mexican immigrant community. *Publicly Accessible Penn Dissertations*. 2420.

Liceras, J. M., Fuertes, R. F., Perales, S., Pérez-Tattam, R., & Spradlin, K. T. (2008). Gender and gender agreement in bilingual native and non-native grammars: A view from child and adult functional–lexical mixings. *Lingua*, *118*(6), 827–851.

Linaza, J., Sebastián, M. E., & del Barrio, C. (1981). Lenguaje, comunicación y comprensión. La adquisición del lenguaje. *Monografía de Infancia y Aprendizaje*, 195–198.

Lipski, J. M. (2014). *El Español de América*. Octava Edición. New York: Longman Group Limited.

López Ornat, S. (1994). *La adquisición de la lengua Española*. Madrid: Siglo XXI.

MacWhinney, B. (Ed.). (2000). *The CHILDES Project: Tools for analysing talk: Vol. 1*. Mawah, NJ: Erlbaum.

Miller, K. (2013). Acquisition of variable rules: /s/-lenition in the speech of Chilean Spanish-speaking children and their caregivers. *Language Variation and Change*, *25*, 311–340.

Miller, K., & Schmitt, C. (2012). Variable input and the acquisition of plural morphology. *Language Acquisition: A Journal of Language Development 19*, 223–261.

Paradis, J., & Genesee, F. (1996). Syntactic Acquisition in Bilingual Children. *Studies in Second Language Acquisition*, *18*, 1–25.

Pérez-Pereira, M. (1989). The acquisition of morphemes: Some evidence from Spanish. *Journal of Psycholinguistic Research*, *18*(3), 289.

Penny, R. J. (2000). *Variation and change in Spanish*. Cambridge: Cambridge University Press.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Shin, N. L. (2016). Acquiring constraints on morphosyntactic variation: children's Spanish subject pronoun expression. *Journal of Child Language*, *43*(4), 914–947.

Silva-Corvalán, C. (2014). *Bilingual language acquisition: Spanish and English in the first six years*. Cambridge University Press.

Smith, J., & Durham, M. (2019). *Sociolinguistic Variation in Children's Language: Acquiring Community Norms (*Studies in Language Variation and Change). Cambridge: Cambridge University Press.

Tinoco, A. R. (2017). Variation of the second person singular of the simple past tense in Twitter: Hiciste vs. hicistes 'you did'. *Dialectologia. Special issue*, *7*, 149–167.

Vila, I. (1990). *Adquisición y desarrollo del lenguaje*. Barcelona: Graó.

Cite this article: Baker E (2022). ¡Casi te caístes!: Variation in second person singular preterit forms in Spanish Children. *Journal of Child Language* **49**, 1256–1267. https://doi.org/10.1017/S0305000921000507

https://doi.org/10.1017/S0305000921000507 Published online by Cambridge University Press