Analyzing the production of a non-standard form: Variable use of preterit *andar* in *bogotano* Spanish

Maritza Nemogá and Matthew Kanwit

Abstract

Although the Spanish verb *andar* ‘to walk’ has a prescriptively irregular preterit conjugation (e.g., *anduve* ‘I walked’), adult native speakers may produce an innovative, regularized form (e.g., *andé*). Nevertheless, researchers have had difficulty documenting this form of variation. The current study presents the first systematic, variationist analysis of *andar* regularization. We sought to answer three research questions: 1) Do speakers in Bogotá, Colombia produce regularized forms of preterit *andar*? 2) If so, at what rate do *bogotanos* produce regularized forms? 3) Do any independent factors predict regularization? We designed a 25-item cloze task in which we manipulated person/number, with 164 *bogotanos* participating. In addition to providing the first systematic study of *andar* regularization, we demonstrate that the phenomenon is constrained by independent linguistic and social variables, while appealing to changes in the overall frequency of use of *andar* diachronically and to the relationship between the frequency of use of each person/number and regularization.

Keywords: NON-STANDARD FORMS, SOCIOLINGUISTIC VARIATION, *BOGOTANO* SPANISH, *ANDAR*
1 Introduction

Regular and irregular verbs exist in Spanish, as in other Romance, Germanic, and Semitic languages (Joseph, 2000). Although both types of verbs exist, regularization of irregular forms in Spanish is not common once speakers have fully acquired the language. That is, when children are acquiring the language, they tend to regularize irregular verbs, as with English (Batchelor and San José, 2010), but once the language has been acquired, this generally ceases to occur. It is not uncommon, for example, to hear a child say sabí\textsubscript{REG} ‘I knew’ instead of supe\textsubscript{IRREG} (Fernández and Cairns, 2011). Nevertheless, in certain regions, the regularization of the irregular Spanish verb andar ‘to go around’ or ‘to walk’ seems to be a possibility for adult speakers, although it has not yet been systematically studied. Speakers produce forms such as andé instead of anduve. Batchelor and San José explain, ‘It is not uncommon to come across the forms: andé, andaste, andó, etc. These are quite common in [colloquial] speech, notably in Iberian Spanish, and can be heard in Castile also’ (2010:169). The authors also state that there are corroborated uses of this regularized verb in Latin America, specifically in Colombia and Mexico. Seco (1998) affirms that there is proof of the use of regularized andar in the Middle Ages since the 12\textsuperscript{th} century and that nowadays its use is more commonly heard in Catalanian and Galician areas.

A Google search of the contrast andé / anduve yields over 4,000 results, mostly from blogs and forums that discuss the practice of regularizing andar. Some of these websites prescriptively explain that although people use the regularized form, the form should be avoided due to its perceived incorrectness. Other sites directly attack as uneducated those who use the regularized form in public, especially journalists or politicians. For instance, one of these blogs is titled ‘Yo andé, tú andaste, él andó’.\textsuperscript{1} The author starts the debate by saying ‘No crean que me he vuelto loco. […] escucho a Manu Sánchez, uno de los presentadores de deportes, decir […]: no se andaron con chiquitas’. ‘Don’t think I’m insane […] I hear Manu Sanchez, one of the sports anchors say […]: they didn’t go\textsubscript{REG} around with small ones’. Forum participants responded with comments such as ‘es lamentable y, por desgracia, demasiado frecuente, oír en boca de estos supuestos periodistas burradas semejantes’. ‘It’s regrettable, and unfortunately too frequent, to hear such nonsense in the mouths of these supposed journalists’. Nevertheless, in spite of its mention in online blogs, the variant has not captured the local imagination in ways that have been documented for other regional variants, such as the word-final bilabial nasal in the Yucatán, appearing on t-shirts and bumper stickers as ‘Yucatam’ (Michnowicz, 2008). Moreover, use of this variant has resulted in a rather popular joke that has been told for years about little Johnny
telling the story of the raising of Lazarus: ‘Cuando Jesús le dijo “levántate y anda”, Lázaro andó. Escandalizado el párroco le susurró con irritación “Anduvo, tonto, anduvo”. A lo cual Jaimito respondió “Al principio anduvo tonto, pero andó”’ (‘When Jesus said “get up and walk”, Lazarus walked\textsuperscript{REG}. In shock the chaplain told him with irritation “he walked\textsuperscript{IRREG}, silly, he walked\textsuperscript{IRREG}”. And little Johnny replied “At the beginning he walked\textsuperscript{IRREG} silly, but he walked\textsuperscript{REG}”’). This joke has several versions in different countries but they all poke fun at the fact that andar is regularized.\textsuperscript{2}

When we informally questioned several bogotanos about andar, they noted that they do use forms such as andé, and although they did not have strong intuitions about who might be most likely to use it, some individuals suspected that it would be younger speakers.

We selected bogotano Spanish as our dialect of inquiry because Colombia is among the countries where regularization of andar is thought to occur (Batchelor and San José, 2010). Bogotá is a logical location to study because it houses the largest and most influential dialect in the nation. According to the 2005 census, with over 8 million inhabitants in the urban area and surrounding municipalities, Bogotá is the third largest Spanish-speaking city after Mexico City and Buenos Aires. Bogotano Spanish is a variety that is considered prestigious both in Colombia and the rest of the Spanish-speaking world (Montes, 1998; Schwegler, Kempff, and Ameal-Guerra, 2009). With these characteristics, Bogotá represents an optimal starting point to analyze systematically the regularization of the verb andar.

Although andar regularization is discussed in informal online environments, the current paper presents the first systematic, empirical investigation of this variable phenomenon. The remainder of the introduction is divided into two sections. Section 1.1 presents background information on Spanish verbal morphology and previous work on the study of other non-standard forms in Spanish. Section 1.2 contains the current study’s theoretical background, research questions, and hypotheses.

1.1 Spanish verbal morphology and the variable use of non-standard forms

Spanish verbal paradigms contain irregular forms in all tenses, moods, and aspects.\textsuperscript{3} Stem-changing verbs in the preterit, also called strong preterits (pretéritos fuertes, RAE, 2010: 64), are mostly those verbs that kept the Latin stem and whose first and third person conjugations become paroxytone words as opposed to the oxytone regular verbs. Some examples are decir ‘to say’, hacer ‘to do’, and querer ‘to want’. Their first person singular forms in the preterit are: dije,
Although there are conflicting accounts that explain the irregularity of andar (Commelerán y Gómez, 1897; Lanchetas, 1897; Monlau, 1856), Torres y Gómez (1899) groups andar with tener ‘to have’ and estar ‘to be’. These three verbs take [β] (<v>) as a characteristic of their preterit form; that is, tuve ‘I had’, anduve ‘I walked around’, and estuve ‘I was’.

Nonetheless, verbs with partial phonetic analogy with andar, such as mandar ‘to send’, are regular (e.g., mandé). Furthermore, semantically similar verbs are also regular in the preterit (e.g., caminé ‘I walked’). Although it might be common to overhear native speakers using andar as a regular verb in the preterit or the imperfect subjunctive (which is formed from the preterit), this is thought to occur less commonly for written forms (Seco, 1998). Though regularization of andar has not been studied empirically, other non-standard forms have received attention in the literature. We now provide a brief overview of this body of work.

Research on the regularization of irregular paradigms through analogy has primarily focused on the pluralization of the prescriptively singular existential haber, especially in the imperfect (habían ‘there were’) (Bentivoglio and Sedano, 1989, and D’Aquino-Ruiz, 2004 for Caracas Spanish; Claes, 2014 for Puerto Rican Spanish; Claes, 2015 for Dominican Spanish; Castillo-Trelles, 2007 for Yucatán Spanish; and Díaz-Campos, 2000, 2003 for Venezuelan and Chilean Spanish; see also De Mello [1991] and Fontanella de Weinberg [1992] for more general studies of haber pluralization). De Mello concludes from his study of 11 Latin American cities that pluralization is a characteristic found not only in the habla inculta ‘uncultured language or vernacular’ but also in the habla culta ‘cultivated language’. Claes (2015) reports pluralization rates of 47% in the Dominican Republic and notes that pluralization is conditioned by social class membership as an extra-linguistic factor and several cognitive factors (i.e., markedness of coding, structural priming, and statistical pre-emption). Claes (2014) also found robust pluralization (41%) in Puerto Rican Spanish, explained by social class, formality, and gender as a change from below (Labov, 2001).

1.2 Theoretical background, research questions, and hypotheses
The present study aimed to analyze the regularization of the Spanish verb andar. We now consider a range of linguistic and social factors.

1.2.1 Linguistic factor
To determine the extent of preterit andar use across persons and numbers, we consulted the original genre/historical section of Davies’ Corpus del Español, which contains 100 million words. The overall frequency of use of preterit andar according to subject in the 20th-century data in tokens per million is found in Figure 1. All such uses of preterit andar in the Davies corpus were prescriptively
irregular. Moving left to right in Figure 1, preterit *andar* occurs least frequently with tú (you) in the corpus, followed by nosotros/nosotras (we), ellos/ellas/Uds. (they/polite you), yo (I), and él/ella/Ud. (he/she/polite you).

![Frequency of preterit *andar* (in tokens per million words)](image)

**Figure 1.** Preterit *andar* in the 20th century according to person and number.

### 1.2.2 Social factors

Three constructs relevant for language variation and change are gender, socioeconomic status (SES), and age (Labov, 1972, 1994, 2001). According to Labov, any social class may introduce change, although higher SESs usually do not lead linguistic change. This is known as the ‘curvilinear hypothesis’, which states that ‘linguistic change from below originates in a central group, located in the interior of the socioeconomic hierarchy’ (2001:188). Gender also plays an important role in language change. Women tend to use innovative variants more frequently than men (Labov, 2001:292). However, this leads to the gender paradox: ‘women conform more closely than men to sociolinguistic norms that are overtly prescribed, but conform less than men when they are not’ (2001:293). In other words, if some level of stigma is associated with a non-standard form, women are generally less likely than men to adopt it. In relation to age, Labov presents the idea of apparent time, which refers to the ‘distribution of linguistic variables across age levels’ (1994:45). This is a straightforward method to study linguistic change, as it utilizes cross-sectional data to stand in for possible change over time (i.e., longitudinal data). Along the same lines, Labov affirms that preadolescents and adolescents seem to lead the change and therefore this population should be highlighted in studies of language change. Lastly, speakers from higher socioeconomic classes tend to have a higher degree of security in their use of the language based on contact with speakers of prestigious varieties (Labov, 1972). Although Labov’s principles have mainly been applied to phonetic change, they are also applicable to morphosyntax (as in Claes, 2015).
1.2.3 Research questions and hypotheses
The present study is the first to analyze systematically the regularization of Spanish \textit{andar}. With Labov’s tenets in mind, the following research questions guide the study:

1. Do speakers in Bogotá produce regularized forms of preterit \textit{andar}?
2. If so, at what rate do \textit{bogotanos} produce regularized forms?
3. Do any linguistic or social factors predict the production of regularized forms?

We hypothesize that \textit{bogotano} Spanish speakers will regularize \textit{andar}. Since the regularized form is innovative and non-standard, and since we do not have evidence that the form is particularly prestigious, we hypothesize, in line with Labov’s constructs, that participants in the lower edges of our sample (i.e., younger, lower SES) and men will be more likely to use the regularized form than those in the middle or higher edges (i.e., older, higher SES) and women. As for our linguistic factor, the person and number of the subject, we hypothesize that some persons and numbers will trigger more regularization than others, since the forms are not equally frequent. Given that less frequent forms are less resistant to regularization (e.g., Bybee, 2006a), we predict that the less frequent \textit{tú} should be most likely to co-occur with a regularized preterit \textit{andar} and that the most frequent \textit{él/ella/Ud.} should be least likely to do so.

If \textit{andar} is being regularized, an explanation for this phenomenon will be necessary. We propose that \textit{andar} is being regularized by analogy with other regular verbs such as \textit{mandar} (to order) and \textit{caminar} (to walk), which share phonetic and semantic similarity, respectively. According to Hopper and Traugott (2003), analogy and reanalysis are two mechanisms of general language evolution which apply to morphosyntactic change. Originally, analogy was defined simply and narrowly as the ‘process whereby irregularities in grammar, particularly at the morphological level, were regularized’ (2003:64). According to Hopper and Traugott, the product of change by analogy serves as evidence ‘for speakers (…) that change has taken place’ (2003:64). More importantly, language change by analogy ‘modifies the surface manifestation and in itself does not effect rule change, although it does effect rule spread either within the linguistic system itself or within the community’ (2003:39). In other words, regularization of \textit{andar} through analogy would account for a change in the morphology of the verb but not in its meaning. At the same time, this change by analogy has an effect on how the change is spread in the community. Change by analogy could be due to a reduction in the frequency of use of \textit{andar}. According to Bybee and Slobin (1982), verb regularization tends to occur for those verbs that are less frequently used by
speakers. On the other hand, words that are more frequently used are less likely to be changed by speakers (Bybee, 2006b). Accordingly, if the overall rate of use of andar has decreased diachronically, the verb would be a better candidate for regularization. In order to consider this possible diachronic change empirically, we consulted Davies’ corpus. We analyzed changes in the frequency of andar across all paradigms (i.e., not only the preterit) throughout the last 800 years. As seen in Figure 2, in the 20th century, the frequency of andar is the lowest of all centuries, with just 147 tokens per million words. This result is a good starting point to argue for the effects of the decreased frequency of andar, which will be considered in greater detail in section 4.5.

![Figure 2. Frequency of andar across all paradigms from the 1200s to the 1900s.](image)

If our data indicate that andar is being regularized, it will also be necessary to explain the form’s distribution in the population. Labov (2007) explains that transmission and diffusion are two of the mechanisms through which language change is spread. Transmission refers to children’s native-language acquisition and their ability to faithfully replicate their elders’ language. Diffusion is a simpler process that involves language contact between adults, and mostly occurs with lexical items or phonetic changes. Transmission usually occurs as change from below, whereas diffusion usually spreads through change from above, and the former is usually speech-community internal, whereas the latter involves contact between speech communities (Labov, 2007). Nevertheless, despite their differences, there are possible overlaps between the two processes, or a change may, for example, begin as transmission and then be spread through diffusion as contact occurs (Tagliamonte, 2012). The applicability of these concepts to the regularization of andar in bogotano Spanish will be further considered in the discussion (i.e., section 4).
Figure 3. Demographic data distribution. Raw values are in parentheses.
2 Method

2.1 Participants

The study includes 164 participants, all born in and currently residing in Bogotá, Colombia. We confirmed this information through the demographic questions in our online questionnaire, described in section 2.3. Participants provided their city of origin and the name of their current neighborhood in Bogotá, which allowed us to exclude any participants who were not born in Bogotá or not currently residing there. The social variables age, gender, and SES have been coded for each participant. The majority of our participants come from the intermediate age groups, as participants between 10 and 15 years of age represent 5% (N=9) of our sample, whereas 15% were between 16 and 20 years of age (N=24), 22% of participants were between the ages of 21 and 30 (N=37), 46% were between the ages of 31 and 40 (N=75), and 12% were 41 years of age or older (N=19). Ages 41–50 and 51+ were eventually combined given the small number of participants in the 51+ group and their comparable linguistic behavior to the prior group. Similarly, groups 10–15 and 16–20 were combined into a group labeled 10–20 (see Figure 3).

Somewhat typical of voluntary linguistic studies disseminated online (e.g., 67% participation by women in Agostini and Schwenter 2015), our sample included more women than men: 33% of our participants were men (N=54) and 67% were women (N=110). As for socioeconomic status, there were two factors that influenced participants’ placement in one of the three SES groups. Colombia has a social class system primarily determined by place of residence, which is highly dependent on income. Every house in Bogotá is assigned a social stratum from one (lowest) through six. The stratum appears in the utility bills and is familiar to most residents from a young age. In order to obtain an objective measure of SES without the intrusion of asking for people’s social stratum, respondents provided the neighborhood where they lived and their profession. The researchers found the social stratum of each neighborhood and compared it to the profession of each person; in the rare event of any discrepancies (N=2), SES was assigned according to profession, rather than neighborhood. For the younger school-age participants, SES was assigned solely based on the stratum of the neighborhood of residence, since we did not have access to the occupation of those participants’ parents. Low SES corresponded with social strata 1–2, whereas middle corresponded with 3–4 and high with 5–6. Accordingly, 37% of participants were classified as having a low SES (N=60), 56% were middle SES (N=93), and 7% were high SES (N=11). All participants were literate and had completed at least elementary school.
2.2 Cloze task
The data-collection instrument was a 25-item cloze task in which participants were given a sentence with a verb in parentheses and were instructed to complete the sentence by conjugating the verb in the past. We implemented the cloze task due to the aforementioned low frequency of *andar* in the preterit, since a traditional sociolinguistic interview would likely fail to elicit many relevant tokens, let alone across a variety of persons and numbers. Participants were first given an example, with a verb conjugated in the preterit in order to encourage the production of preterit rather than imperfect aspect. The task included verbs across five different persons: first, second, and third person singular, and first and third person plural. In addition to the five conjugations of *andar*, we included two irregular verbs and a subset of regular verbs, two with each thematic vowel (*-ar, -er, -ir*). The regular verbs within each paradigm were never repeated with the same person and number. For example, for the regular *-ar* paradigm, *preparar* ‘to prepare’ was used with the first and third person singular whereas *jugar* ‘to play’ was used in the second person singular and first and third person plural. Two irregular verbs, *tener* ‘to have’ and *estar* ‘to be’, both of which contain the morphological pattern *-uv-*, were also used across five forms each. This yielded a total of 25 verbs (five verbs × five conjugations = 25). For sample items, see the Appendix.

2.3 Procedure
The questionnaire for the present study was administered online via Qualtrics. The questionnaire link was distributed via email, social media, and electronic chat, and participants were able to respond to it on computers, tablets, or smart phones, which perhaps prevented some members of the lower SES from participating, although there were ultimately many more participants from the lower SES (37%) than from the higher (7%). Accordingly, individuals’ access to the questionnaire depended on whether they encountered the link through the above resources and whether they decided to participate. Prior to participation, respondents were asked whether they were minors (under 18 years of age) and taken to the corresponding consent pages. Once participants (and parents, in the case of minors) gave their consent, they answered a series of six demographic questions, which provided the information in Figure 3. After responding to the demographic questions, participants were reminded not to consult any outside sources when completing the questionnaire.

For the cloze task, subjects were instructed to complete the sentence with the verb in parentheses by conjugating it in the past tense. Although the term ‘preterit’ was never used in the questionnaire, the example provided occurred in that form. Instructions and the same example was displayed on the screen for the
twenty-five sentences that participants completed: Ejemplo: Ayer yo _______(visitar) a mi tío. Respuesta: ...visité... ‘Example: Yesterday I ______ (visit) my uncle. Answer: …visited…’ Sentences were displayed one at a time and participants could not go back to change their answers, so as to capture their initial responses. The order of the experimental sentences was held constant, since we were manipulating a number of other linguistic and social variables.7 Participants were not given a time limit.8

Data were also coded for one linguistic variable: the person and number of the conjugated verb. Once the tokens of andar were extracted, a mixed-effects binomial regression was performed using Rbrul in order to establish which independent variables influence regularization of the verb andar. We now turn to the results of these analyses.

3 Results

The study’s results begin with the overall distribution of regularization rates, followed by the results of the regression analysis, which selects only the factors that significantly improve the model that predicts regularization.

3.1 Overall rates of use

Our dependent variable, production of andar, included two variants: regularized andar (e.g., andé) and the canonical irregular form (e.g., anduve). Overall, participants produced regularized preterit andar 34.0% of the time and the canonical irregular form in 66.0% of cases (see Table 1). This indicates that, although the prescriptive irregular form is more commonly used, the use of the regularized form is not negligible.

Table 1. Distribution of preterit andar in bogotano Spanish.

<table>
<thead>
<tr>
<th>Andar</th>
<th>Number of tokens</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regularized</td>
<td>236</td>
<td>34.0</td>
</tr>
<tr>
<td>Irregular</td>
<td>459</td>
<td>66.0</td>
</tr>
<tr>
<td>Total</td>
<td>695</td>
<td>100</td>
</tr>
</tbody>
</table>

We also note that the other two irregular verbs included in our task (i.e., estar and tener) were never regularized by our participants, who produced the canonical irregular forms in all 695 opportunities. Although overall rates of production are
interesting, an analysis of the role of independent factors sheds further light on possible constraints on regularization. Variation in the use of andar according to the independent social and linguistic factors will now be considered.

### 3.2 Significant factors
As explained in section 2.1, due to the small number of participants in certain age groups, we combined smaller groups that displayed similar linguistic behavior. This resulted in four evenly distributed age groups (10–20, 21–30, 31–40, and 41+).

Once the data were recoded, a mixed-effects regression analysis was performed in Rbrul (Johnson, 2009), which evaluates the effect of the independent linguistic variables on the dependent variable as main effects, while considering each individual participant as a random effect. The regression measures the probabilistic weight of the impact of each independent variable in relation to the application value (i.e., regularization), with the possibility of selecting our four independent variables: SES, age, and gender, as well as the person and number of the conjugated verb (see Table 2). As the regression selects an independent variable for inclusion in the predictive model, we have statistical confirmation that the observed differences in andar regularization across the categories of that variable are significantly different. The regression continues to select independent variables in the order in which they aid its predictive power, until selection no longer improves the model. One benefit of the method that we have implemented over traditional methods, such as the use of regression analysis performed in Goldvarb X, is that the latter does not account for the role of individual participants (Johnson, 2009; Tagliamonte, 2012). Factor weights above 0.50 (in the rightmost column) indicate a favoring effect for regularization of andar, whereas values below 0.50 indicate that regularization was disfavored (i.e., less likely). Brackets around factor weights indicate that a factor group was not significant, which is also revealed by a p-value for that factor group of at least .05. Log odds, in the leftmost column, are similar to factor weights except that positive values indicate the favoring of regularization, whereas negative values indicate disfavoring effects. Values further from 0 indicate stronger effects, in the same way that factor weights further from 0.50 do.
Table 2. Factors contributing to production of regularized *andar* in Bogotá.

<table>
<thead>
<tr>
<th>Age (p = 0.000113)</th>
<th>Log odds</th>
<th># of tokens</th>
<th>% (of regular)</th>
<th>Factor weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>10–20</td>
<td>2.516</td>
<td>33</td>
<td>72.7</td>
<td>0.925</td>
</tr>
<tr>
<td>21–30</td>
<td>1.832</td>
<td>89</td>
<td>65.2</td>
<td>0.862</td>
</tr>
<tr>
<td>31–40</td>
<td>-1.909</td>
<td>155</td>
<td>32.3</td>
<td>0.129</td>
</tr>
<tr>
<td>41+</td>
<td>-2.439</td>
<td>418</td>
<td>24.9</td>
<td>0.080</td>
</tr>
</tbody>
</table>

Range 85

<table>
<thead>
<tr>
<th>Socioeconomic Status (p = 0.000287)</th>
<th>Log odds</th>
<th># of tokens</th>
<th>% (of regular)</th>
<th>Factor weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower</td>
<td>2.167</td>
<td>243</td>
<td>51.9</td>
<td>0.897</td>
</tr>
<tr>
<td>Middle</td>
<td>-0.419</td>
<td>403</td>
<td>26.1</td>
<td>0.397</td>
</tr>
<tr>
<td>Higher</td>
<td>-1.748</td>
<td>49</td>
<td>10.2</td>
<td>0.148</td>
</tr>
</tbody>
</table>

Range 75

<table>
<thead>
<tr>
<th>Verb Person/Number (p = 0.000706)</th>
<th>Log odds</th>
<th># of tokens</th>
<th>% (of regular)</th>
<th>Factor weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tú</td>
<td>1.013</td>
<td>140</td>
<td>42.9</td>
<td>0.734</td>
</tr>
<tr>
<td>Nosotros</td>
<td>-0.008</td>
<td>139</td>
<td>34.5</td>
<td>0.498</td>
</tr>
<tr>
<td>Uds./Ellos</td>
<td>-0.047</td>
<td>133</td>
<td>33.1</td>
<td>0.488</td>
</tr>
<tr>
<td>Yo</td>
<td>-0.143</td>
<td>148</td>
<td>33.1</td>
<td>0.464</td>
</tr>
<tr>
<td>Él/Ella/Ud.</td>
<td>-0.815</td>
<td>135</td>
<td>25.9</td>
<td>0.307</td>
</tr>
</tbody>
</table>

Range 42

<table>
<thead>
<tr>
<th>Gender (p = 0.642)</th>
<th>Log odds</th>
<th># of tokens</th>
<th>% (of regular)</th>
<th>Factor weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women</td>
<td>0.190</td>
<td>465</td>
<td>34.8</td>
<td>[0.547]</td>
</tr>
<tr>
<td>Men</td>
<td>-0.190</td>
<td>230</td>
<td>32.2</td>
<td>[0.453]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Speaker (random)</th>
<th>Log odds</th>
<th># of tokens</th>
<th>% (of regular)</th>
<th>Factor weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random St. Dev.</td>
<td>3.218</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total N = 695 Overall rate 34.0%

According to the regression, there were three significant factor groups that predicted the regularization of *andar*: age, SES, and person/number. For the factor groups, lower *p*-values indicate greater confidence that the variable was not selected by chance, along with larger ranges (i.e., the difference between the largest and smallest factor weight within the factor group). We now consider the significant factor groups in the order in which they were selected by the regression.
3.2.1 Age
There was a significant prediction of regularization by age at the \( p < .001 \) level. Bogotanos between the ages of 10 and 20 favored regularized \textit{andar} most strongly (with a factor weight of 0.925). Similar to the youngest group, participants from 21 to 30 years of age also strongly favored regularization (factor weight 0.862). Participants in the 10–20 age group (72.7%) and 21–30 group (65.2%) each regularized \textit{andar} more than twice as often as speakers between the ages of 31 and 40 (32.3%). Speakers older than 41 years of age produced regularized \textit{andar} at the lowest rate (24.9%) and thus they strongly disfavored this variant (factor weight of .08). We note, nevertheless, that regularization was present in all age groups.

3.2.2 Socioeconomic status (SES)
Socioeconomic status also had a significant effect on the production of regularized \textit{andar} at the \( p < .001 \) level. Participants in the lower SES were the most likely to produce regularized instances of \textit{andar} (51.9%, factor weight of 0.897), followed by those of the middle SES (26.1%), who disfavored the form (factor weight of 0.397). Speakers from the higher SESs most strongly disfavored regularization (factor weight 0.148), producing the form 10.2% of the time. Regularization of \textit{andar} was again found in all three groups.

3.2.3 Person and number
The person and number of the conjugated verb was the final significant predictor of regularization of \textit{andar} (\( p < .001 \)). Regularized \textit{andar} was produced at higher rates with 2SG \textit{tú} ‘you’ (42.9%) than all other persons, creating the most favorable context for regularization (factor weight 0.734). \textit{Nosotros} (1PL), \textit{Uds./éllos} (3PL), and \textit{yo} (1SG) all yielded rates of regularization just below 35%. Third person singular (3SG) \textit{él/ella/usted} ‘he/she/you’ (25.9%) was the context that most disfavored regularization (factor weight 0.307). We further explore the link between regularization and 2SG in section 4.

We now briefly consider the possible role of the ordering of the task items on regularization of \textit{andar} by person and number. Recall that one version of the task was created to avoid introducing further confounding variables in this first analysis of variable \textit{andar} regularization. Our first item that targeted \textit{andar} was in 2SG and was the seventh task item, following six items that contained prescriptively regular verbs. Our item for 3SG \textit{andar}, which had the lowest rate of regularization, was elicited following an item that included the irregular verb \textit{estar}. Accordingly, relatively high regularization of 2SG \textit{andar} (43%) following a series of regular verbs and relatively low regularization of \textit{andar} (26%) following irregular \textit{estar} point to the possibility that the arrangement of task items
influenced rates of regularization. Nevertheless, participants did not simply produce fewer regularizations of *andar* as they gained experience with the task. Regularization rates were higher for the final three *andar* conjugations than for 2SG, and these three forms’ rates of regularization were quite similar to each other (i.e., 33% for 1SG, 35% for 1PL, and 33% for 3PL). Overall, the ordering of task items likely had some effect on responses, but participants did not simply produce more standard, non-regularized forms as the task continued.

### 3.3 Non-significant factor: Gender

As was shown in Table 2, regularization rates of *andar* were very similar for participants of each gender. Men regularized *andar* 32.2% of the time, and women demonstrated a similar rate of 34.8%. Accordingly, gender was not selected as significant ($p = .642$).

### 4 Discussion

We now return to the research questions that guided the current study. The first two questions were concerned with whether residents of Bogotá, Colombia regularized the Spanish verb *andar*, and if so, at what rate. Results showed that *bogotanos* regularized *andar* in the preterit 34% of the time. Although the prescriptively canonical irregular form was used in the majority of cases, the regularization rate is quite substantial. Furthermore, the 34% rate of regularized *andar* is noteworthy given the instrument used to collect data, for which participants were required to type written responses. Thus, *andar* is being regularized at a high rate even in a relatively formal context, so we might expect even higher rates in less formal contexts, as predicted by Labov’s (2001) patterns of stable variation, for which increases in formality are accompanied by decreased uses of innovative variants. Although a less formal context may yield higher rates of regularization, we selected a more controlled task due to the relatively low frequency of preterit *andar*, since traditional sociolinguistic interviews would likely fail to elicit many such tokens. Furthermore, due to the lack of previous variationist work on innovative *andar* use, we have aimed to provide an initial baseline of the extent of regularization in an elicitation task. Nonetheless, our results still indicate that *andar* is in fact being regularized.

Our third research question targeted whether any linguistic or social factors predicted the production of regularized *andar*. We hypothesized that younger and lower SES participants, as well as men, would be more likely to regularize *andar*. Our results partially align with our hypotheses. The youngest speakers (ages 10–20) most strongly favored the regularized form, followed closely by speakers...
aged 21 to 30, whereas our two older groups strongly disfavored regularization. Nevertheless, use of the regularized form was not progressive enough among age groups to indicate whether this is a change in apparent time or if it is an age-grading phenomenon (Labov, 1994). That is to say, if this is a relatively new development, then we might hypothesize that the future will show a more step-like progression across age groups, with less regularization from the youngest group to each older group. On the other hand, if age grading is occurring, *andar* regularization would particularly affect the youngest speaker group and would continue to set that group apart from the others. In the latter scenario, in essence, younger speakers would be using the innovative, non-standard form more frequently, but as they get older and become more educated, their use of the canonical irregular form would increase. A further possibility, which would result in similar patterns to age grading and which would affect speakers slightly younger than those of our study, would be that relics of child speech remain and cause greater regularization of *andar*. For example, for child speakers of English under the age of eight, the strength of the regularity of the schema for verbs ending in *-t/-d* triumphs over that of the suffixation rule, whereas later such regularization gives way to the suffixation rule (Bybee and Slobin, 1982). Similarly, child speakers of Spanish develop a grammar that is constrained by a number of factors that increase with age (e.g., Shin, 2016). By age 10, for example, more factors constrain children’s subject pronoun use, as compared to younger children (ages 6–9), indicating a system that is more adultlike and less reflective of relics of child speech. Since there was no previous wide-scale documentation of regularized *andar*, future data collection will enable us to better disentangle which scenario better reflects *bogotano* speech. Regardless, the effects for age here are different from some previous findings for the production of non-standard pluralized *haber* (Claes, 2014, 2015; D’Aquino-Ruiz, 2004; Fontanella de Weinberg, 1992). In those studies, age was not a significant factor, which led the researchers to conclude that the pluralization process is taking place slowly enough to be undetectable in apparent time, whereas our results indicate potentially more notable differentiation by our youngest group.

With respect to socioeconomic status, we hypothesized that participants in the lower SES would be most likely to regularize *andar*. The results support our hypothesis since participants in the lowest SES strongly favored the innovative form, whereas individuals in the middle and higher SESs disfavored regularization. As the middle SES participants were 2.5 times more likely than the higher SES to produce regularized *andar*, the use of the non-standard form was inversely proportional to SES, which is similar to findings for *haber* pluralization (Bentivoglio and Sedano, 1989; Díaz-Campos, 2003). In those studies, the pluralized
form appeared to be the norm among lower class speakers, whereas the opposite occurred with the higher class, which mirrors our data, as our low SES speakers actually produced the regularized variant more (52%) than the standard form.

The results for gender were somewhat counter to our expectations. We hypothesized that men would use the non-standard form more frequently than women if regularization carried some level of stigmatization. Previous findings on the effects of gender on the use of non-standard forms are conflicting, even from the same authors. For instance, Díaz-Campos (2000) and Claes (2014) found gender to be a significant contributor, but for Díaz-Campos (2003) and Claes (2015) gender’s effects were not significant. Cases where gender does not achieve significance are consistent with Labov’s (2001) claim that gender differences are becoming less noticeable, which may be related to women’s increasing participation in the workforce. In our sample, for instance, fewer than 10 women were stay-at-home mothers. Furthermore, that both men and women produced regularized andar at similarly high rates in a relatively formal task likely indicates that the form is not particularly stigmatized, which is further supported by the spread of the form to all classes and age groups.

The independent linguistic variable coded was the person and number of the conjugated verb. We hypothesized that andar would demonstrate differential regularization across persons, and the results supported this hypothesis. Specifically, our earlier search of Davies’ 20th-century data (see Figure 1) indicated that preterit andar occurred least with tú and most frequently with él/ella/Ud. In the current study, the informal second personal singular tú triggered more regularization than the other persons. In fact, tú was the only favorable context for regularization. A possible explanation for the higher regularization rate for 2SG is the different frequencies of use of different verbal conjugations. We can compare the previously reported rates of use of preterit andar in Davies’ 20th-century corpus data with use of regularized andar by person and number in our data. As illustrated in Figure 4 (an update of Figure 1), the frequency ranking of each form of andar in the 20th-century corpus is inversely proportional to the regularization rates of the same form in our data. Though the frequency in Davies’ corpus is given in tokens per million, whereas the current data list percentages of use, the comparison reveals that the more frequent a form is, the less it tends to be regularized. These data also match the general prediction that more frequent forms better withstand formal irregularities (i.e., undergo conservation), whereas less frequent forms tend to be regularized (Bybee, 2006a). Among the more accepted explanations for the conservation of frequent forms is that frequency bolsters the memory representations of words and phrases, enabling easier access to them as wholes (Bybee, 2006a; Hooper, 1976). Our findings would be consistent with such a process.
Due to a decrease in *andar*’s overall frequency of use (see again Figure 2), possible regularization by analogy with phonetically or semantically similar verbs might further account for regularization. In our data, the distribution of regularized *andar* occurs in an upward fashion, in which the younger and less wealthy are using the form more than the rest of the population, leading to paradigmatic simplification. Similarly, the inverse relationship between *andar*’s increasing regularization and decreasing frequency of use in the last century is consistent with claims that regularization occurs with less frequent forms (Bybee and Slobin, 1982), which can further be seen according to the frequency of each individual person/number of *andar*, indicating a microcosm of this process more generally.

The distribution of regularized *andar* in our sample seems to more closely approximate Labov’s (2007) notion of transmission, although it follows aspects of both transmission and structural diffusion, which is the product of language contact between adults. Our data are consistent with Labov’s description of transmission in the sense that children may be advancing a linguistic change to an extent beyond their caretakers (1994), since both adults and children produced regularized *andar*, but our youngest group did so to a greater extent. Our data also meet the description of transmission in that transmission usually corresponds with change from below, which would correspond with the greater regularization that we see by our lower SES speakers. This would also be consistent with the directionality reported in recent work on the pluralization of *haber* in Puerto Rican Spanish (Claes, 2014). Nevertheless, characteristics of both processes may overlap (Labov, 2007). Language diffusion usually involves lexical items (or phonetic changes), displays weakening of original patterns (e.g., lower use of a
standard variant), and a loss of structural features. The loss of the inflectional morphological pattern -uv- that expresses the irregularity of andar in the preterit seems to follow some of the predictions of structural diffusion. Aspects of diffusion are also present in our data in that our higher SES speakers do produce regularized andar. Nonetheless, diffusion usually involves change from above, which does not appear to be supported by our data, as our highest SES speakers are less likely to produce regularized andar. Although transmission and diffusion are complex and may overlap (Tagliamonte, 2012), the present data point more to transmission as the type of change currently occurring for andar regularization in Bogotá.

5 Conclusions and implications

The present study contributes to a growing body of research of the use of non-standard morphosyntactic variants in Spanish. By using a controlled task, we were able to elicit tokens of preterit andar across speaker groups and persons and numbers, which enables an analysis of independent predictors that is not generally possible in a sociolinguistic interview for a relatively infrequent form. Our results indicate that bogotano Spanish speakers do regularize andar, and that younger individuals of the lower socioeconomic status produced the highest rates of this non-standard form. Nevertheless, regularized andar was found in all age, SES, and gender groups, as well as with all persons and numbers.

One of the main contributions of our study is that it is the first systematic documentation of the regularization of andar. Furthermore, our use of a cloze task, rather than a preference task, allowed us to avoid priming participants, which would have occurred by having respondents select from forms that were provided for them. The study also sheds light on possible explanations for the regularization of andar in bogotano Spanish. The tendency towards regularization seems to be a paradigm simplification process in which speakers change a verb’s form but not its meaning, in analogy with similar phonetic or semantic forms. Documented differences in regularization rates among persons and numbers provide further support to established links between higher frequency and lower regularization (Bybee, 2006a). We have also proposed that the regularization of andar exhibits characteristics of both transmission and diffusion, although the higher rates of regularization by our lower SES and our youngest participants are more representative of the process of transmission. Furthermore, time will indicate more definitively whether differences between our youngest groups and the older groups are the result of a change in apparent time or age grading.
The current study has demonstrated that bogotanos produce regularized andar, even in a relatively formal task, and that regularization is constrained by independent linguistic and social factors. Despite the study’s contributions, future work will benefit from analyzing regularization rates in more informal registers, across additional dialects, and through longitudinal data collection.

About the authors

Maritza Nemogá is a doctoral student at the University of Pittsburgh, United States. She is originally from Bogotá, Colombia, and her research has appeared in Bilingualism: Language and Cognition and Sociolinguistic Studies.


Appendix: Sample items

6. Ayer tú ... (jugar) ajedrez.
7. Las vacaciones pasadas tú ... (andar) por la playa.
8. La semana pasada usted ... (estar) enfermo.

Notes

3. For a detailed explanation of the Spanish verbal paradigm, see chapter 4 of the Nueva gramática de la lengua española manual (2010:71–92).
4. Although the imperfect subjunctive is formed from the same stem as the preterit, we have constrained our focus to the preterit in the current study in order to limit the duration of the task and to provide a first step in analyzing innovative use of preterit forms.
5. Although changes in overall rates of use of preterit andar are meaningful, we agree with an anonymous reviewer that diachronic sociolinguistic data from a range of speaker groups would shed important light on the development of andar and would enable firmer conclusions.
6. Although participants were given an example in the preterit, a total of 58 tokens were produced in the imperfect and are therefore excluded from the analysis. There were also 67 items that participants did not complete.

7. Nevertheless, we note that the ordering of the items may have an effect on participant responses. We consider the possible relationship between ordering and person/number in section 3.4.2, although future research will help determine to what extent *andar* regularization may be affected by the order of task items.

8. As one reviewer noted, an untimed task opens the possibility that participants may perform searches about forms online. This is a possibility, although rates of regularization were still quite high, and recall that participants were not able to return to previous task items. Nevertheless, rates would likely be even higher for a timed task.

References


(Received 1st March 2017; revision received 8th July 2017; accepted 2nd November; final revision received 14th December 2017)