The Role of Frequency in First-Person Plural Variation in Brazilian Portuguese: Nós vs. a gente

Catherine E. Travis & Agripino S. Silveira
University of New Mexico

Abstract

This paper investigates the conditioning of the variation between two first-person plural forms in contemporary spoken Brazilian Portuguese, an older pronoun nós used with first-person plural agreement and a newer pronoun a gente, derived from an NP meaning ‘the people’, used with third-person singular agreement. This is part of a broader change in the language involving the breakdown of verbal agreement as third-person marking extends to the domains of first- and second-person. We consider the conditioning of use of these forms in spoken Brazilian Portuguese from Fortaleza and uncover a phenomenon that as yet has not been noted in relation to this variation, namely frequency. We find that the high type frequency of a gente contributes to the spreading use of this form, and the high token frequency of nós in specific constructions slows it down. We predict that these highly frequent constructions (such as nos temos ‘we have’, digamos ‘let’s say’ and vamos + V-INF ‘let’s V’) may remain a last vestige of nós in Brazilian Portuguese as a gente comes to take over the realm of first-person plural.

1. Introduction

Brazilian Portuguese is currently undergoing a major restructuring of its pronominal system and, consequently, its verbal morphology. Specifically, the introduction of new pronouns derived from full Noun Phrases is resulting in the spread of third-person morphology to the realms of both first and second person. In the case of the second person, the forms vocé (from the NP Vossa Mercê ‘Your Grace’) and, more recently, o senhor / a senhora (from the NPs ‘the gentleman / lady’) have largely taken over from the older form tu. As these newer forms take third-person singular marking, this has resulted in the near loss of second-person marking in several parts of Brazil (cf. Azevedo 1981, Faraco 1996, Zilles 2004), including Fortaleza, the region under study here (cf. Lemos Monteiro 1994). In the case of the first-person plural, variation exists throughout Brazil between the older form nós and the newer form a gente, which derives from an NP meaning ‘the people’. As a gente occurs with third-person singular (V-3SG) and nós with first-person plural (V-1PL) agreement

Studies in Hispanic and Lusophone Linguistics
Volume 2, Issue 2
Fall 2009
marking in the Fortaleza dialect considered here, the use of *a gente* serves to further spread third-person agreement marking. The apparent interchangeability of (*a gente*) + V-3SG and (*nós*) + V-1PL\(^1\) can be seen in examples (1) and (2), where the speakers readily shift between them.

(1)  
*Nós fomos fazer um trabalho...*  
juntamente com o nosso professor da Universidade... Estadual  
*a gente foi* fazer um trabalho em Paracuru  
fomos fazer uma pesquisa sobre saneamento básico em Paracuru...  
e o resultado... foi assustador... não é Paracuru não tem esgoto...  
‘We went\(_{1\text{PL}}\) to do a job  

together with our professor from the State University  
*we went\(_{3\text{SG}}\) to do a job in Paracuru  
*(we) went\(_{1\text{PL}}\) to do some research on the sewer system in Paracuru  
and the results were terrifying, that is, Paracuru does not have a sewer system’  
(L52: 108)\(^2\)

(2)  
*Eu achei muito interessante Salvador e Maceió achei muito interessante...*  
João Pessoa *a gente passou assim...*  
*nós tivemos* o azar de passar em Salva em João Pessoa  
foi o primeiro lugar que *nós passamos...* e choveu o dia todo...  
pessoal até fazia muita hora no ônibus que *a gente foi...*  
‘I thought that Salvador and Maceió were very interesting, very interesting  
*we passed through\(_{3\text{SG}}\) João Pessoa just --  
*we had\(_{1\text{PL}}\) the misfortune of passing through Salva- um João Pessoa  
it was the first place *we passed through\(_{1\text{PL}}\) and it rained all day long  
people were even joking on the bus that *we went\(_{3\text{SG}}\) in’  
(I6: 87)

In this research we apply the variationist method in order to identify what factors condition this variation, that is, what determines whether a speaker will use (*a gente*) + V-3SG or (*nós*) + V-1PL in any given context. Given that these pronouns take standard agreement marking almost categorically in these data,\(^3\) the study of the conditioning of use of the pronouns allows us to simultaneously consider the conditioning of the co-occurring verbal agreement, and thus the spread of third-person singular (or zero) marking. Based on an analysis of approximately 2,600 verbs with first-person plural subjects in a corpus of spoken Brazilian Portuguese from the city of Fortaleza, we find that although there is a similar rate of use of the two forms, they pattern in very different ways. Much of this patterning has been described in the literature on this topic, and our results largely correspond with those of previous studies. However, we also uncover a phenomenon that as yet has not been discussed in relation to this variation, namely the role of frequency. While *a gente* has generalized such that it is found across a broad spectrum of first-person
plural environments, nós has become relatively restricted in use, and close to two thirds of the tokens in our data occur in just three environments: with the highly frequent verbs ter ‘have’ and ser ‘be’ (particularly in the present indicative, nós temos ‘we have’, nós somos ‘we are’); in hortative constructions, many of which have grammaticized to form discourse markers (e.g. digamos ‘let’s say’, vamos ver ‘let’s see’); and with the formally identical analytic future tense (vamos + V-INF).

We argue that the frequency with which these different forms are used plays a central role in the retention of nós + V-1PL, and hence places a restriction on the spread of a gente + V-3SG. On the one hand, we observe that the primary factor motivating the retention of nós is the high frequency with which it occurs in the three environments noted above, in accordance with the notion that high token frequency acts to slow analogical change. According to Bybee (2003, 2006, Bybee & Thompson 1997, inter alia), this is because high frequency forms (such as temos ‘we have’, digamos ‘let’s say’) are entrenched in the sense that they have a strong representation in memory and are accessed as whole units rather than formulated on-line. Thus, they resist restructuring from novel, more productive, forms. On the other hand, the generalized use of a gente is reflected in its occurrence with more verb types than nós, or its higher type frequency, which allows it to readily expand in use. That is, a gente is the more productive form, and this helps to advance the change, as it extends to more and more contexts (cf. Bybee & Thompson 1997, Smith 2001, Tottie 1991). Thus, the higher token frequency of certain constructions in which nós occurs and the higher type frequency of a gente function together to allow the spread of a gente to all but a limited set of environments, which we predict may be the only area where nós remains in the future.

2. Development of a gente as a pronoun

The pronoun a gente derives from a full NP consisting of the feminine singular definite article a and noun gente ‘people’, originally meaning ‘the people’. Over time, it lost its nominal properties and came to function as an indefinite pronoun, meaning something like ‘people in general’ (as seen in (3)), and from the 19th century, was used with a more specific first-person plural sense (cf. Borges 2004, Lopes 2003, Omena 2003, Zilles 2004, 2005), illustrated in (4).

(3) Geralmente o o pau...
    como a gente chama assim no no linguajar popular...
    ‘Usually the wood
    as we call_{3sg}, it in colloquial speech’ (C34: 305)
Interviewer: *O senhor pode lembrar assim o local... eh onde o senhor encontrou a sua esposa pela primeira vez?*

Interviewee: *A gente se conheceu por telephone*

Interviewer: ‘Can you remember the place um, where you met your wife for the first time?’

Interviewee: *We met, 3sg, on the phone.’* (I8: 16)

The development of *a gente* has been analyzed as a case of grammaticization, defined by Bybee (2006, p.719) as “the creation of a new grammatical morpheme and new construction out of a particular instance of an old construction” (cf. Borges 2004, Lopes 2003, Omena 2003, Zilles 2004, 2005). Hopper (1991) lays out five features that are evident in grammaticization from the early stages, all of which can be seen in the pronominalization of *a gente*. First, the development of *a gente* has resulted in layering, or variation with *nós*. Second, it has diverged, or become autonomous (cf. Bybee 2006, inter alia), from its nominal counterpart meaning ‘the people’ from which it derived. Third, specialization is occurring, as *a gente* is coming to take over the domain of first-person plural, reflected in its higher type frequency and the restriction of *nós* to a limited number of contexts. Fourth, we find evidence of persistence in the retention of third-person singular agreement marking on the verb even when used with first-person plural meaning (at least for this community of speakers, if not for others; cf. Naro, Görski & Fernandes 1999). And fifth, the form has undergone decategorialization, as the component parts cannot be analyzed compositionally, that is, they no longer function in their original sense of *a* as a singular feminine definite article and *gente* ‘people’ as a singular feminine noun. The pronoun *a gente* has lost number marking (consistently occurring in the singular), as well as gender marking (collocating with both masculine and feminine predicate nominals and adjectives). It also does not allow any modification such as with adjectives, as opposed to its nominal counterpart (e.g. *a gente boa* ‘good people’, where *gente* is used as a noun).

We also see semantic bleaching accompanied by generalization as *a gente* has come to be used in a wide variety of contexts. As noted above, *a gente* developed as a pronoun through its use as a non-specific form to refer to people in general (including the speaker). It retains this use today, as seen in (3) above, however in our data *nós* is the preferred form in this context (illustrated in (5)), with *a gente* being preferred in more specific contexts (as seen in (4) above). Something similar was found by Omena (2003, p.78) and Zilles (2004, p.41) for data collected in 2000 and the 1990s, respectively, indicating that *a gente* has moved far from its source meaning.

(5) *O primeiro grau maior como nós chamamos aqui né?*

‘Secondary school, as we call, 1pl, it here, right?’ (C39: 160)
The generalization of the form *a gente* represents a further step in the ongoing spread of third-person agreement through the verbal paradigm, resulting in all but the first-person singular subject occurring with third-person marking. This “revolution of the third person”, as it has been termed by Faraco (1996, p.55) is summarized in Table 1 (cf. Zilles 2005).

<table>
<thead>
<tr>
<th>old system</th>
<th>emerging system</th>
</tr>
</thead>
<tbody>
<tr>
<td>1SG eu</td>
<td>canto</td>
</tr>
<tr>
<td>2SG tu</td>
<td>cantas</td>
</tr>
<tr>
<td>3SG ele / ela</td>
<td>canta</td>
</tr>
<tr>
<td>1PL nós</td>
<td>cantamos</td>
</tr>
<tr>
<td>2PL vós</td>
<td>cantais</td>
</tr>
<tr>
<td>3PL eles / elas</td>
<td>cantam</td>
</tr>
</tbody>
</table>

Table 1. Verbal agreement in Brazilian Portuguese

A similar spread of third-person morphology is seen in other Romance languages. For example, European Portuguese is undergoing the same kind of breakdown of agreement marking as that seen in Brazilian Portuguese (cf. Costa & Pereira 2003, Menucci 2000). In Latin American Spanish, the second-person plural form *vosotros* has been replaced by *ustedes* (from *Vuestra Merced* ‘Your Grace’) (Jonge & Nieuwenhuijsen in press), just as *vós* has been replaced by *vocês* in Brazilian Portuguese (cf. Azevedo 1981, Faraco 1996). And in French, the clitic pronoun *on* (from the noun *homme* ‘man’) has virtually taken over from the older first-person plural form *nous* (cf. Coveney 2000, Laberge & Sankoff 1979), paralleling the change under study here. However, these changes are largely complete and therefore the process by which they took place is only attestable through the analysis of historical documents. The change for the first-person plural in Brazilian Portuguese, on the other hand, is still very much underway, and thus the variation between *nós + V-1PL* and *a gente + V-3SG* in contemporary speech provides a valuable opportunity to study this change as it takes place, allowing us to better understand the process by which new pronouns generalize in use and verbal agreement breaks down.

3. Data

The data for this research are drawn from the Corpus of Educated Oral Portuguese from Fortaleza (*Corpus do Português Oral Culto de Fortaleza, PORCUFORT*), collected between the years of 1991 and 1994 under the direction of Lemos Monteiro (cf. Lemos Monteiro 1995). All participants are university-educated native speakers of Brazilian Portuguese who were born in Fortaleza, and whose parents were also from Fortaleza or the surrounding region. The portion of the
corpus used for this study comprises 58 speakers, relatively evenly distributed across gender and three age groups, of younger (20-35 years old), middle (36-50) and older (over 51) speakers (Y, M and O respectively in Table 2). Three different registers are represented, each varying in degrees of formality and interaction: two-party conversations; sociolinguistic interviews; and university lectures. For this study we used the entire set of conversational and lecture data in the corpus, and a portion of the interview data to provide a comparable number of tokens. The distribution of the data across registers and social groupings is summarized in Table 2.

<table>
<thead>
<tr>
<th>Register</th>
<th>Conversation</th>
<th>Interview</th>
<th>Lecture</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours</td>
<td>10 ½</td>
<td>12 ½</td>
<td>12 ¼</td>
<td>35 ¼</td>
</tr>
<tr>
<td>Words</td>
<td>130,000</td>
<td>135,000</td>
<td>92,000</td>
<td>357,000</td>
</tr>
<tr>
<td>Speakers</td>
<td>26</td>
<td>14</td>
<td>18</td>
<td>58</td>
</tr>
<tr>
<td>Gender (M / F)</td>
<td>11 / 15</td>
<td>7 / 7</td>
<td>11 / 7</td>
<td>29 / 29</td>
</tr>
<tr>
<td>Age (Y / M / O)</td>
<td>12 / 6 / 8</td>
<td>4 / 4 / 6</td>
<td>7 / 5 / 6</td>
<td>23 / 15 / 20</td>
</tr>
<tr>
<td>1 PL tokens</td>
<td>840</td>
<td>1,167</td>
<td>931</td>
<td>2,938</td>
</tr>
</tbody>
</table>

Table 2. Corpus constitution

4. Methodology

All verbs with first-person plural subjects were extracted, giving a total of 2,938 tokens. Of these, 335 tokens were determined to be unanalyzable and were excluded from the statistical analyses. We excluded all tokens of non-finite verbs (N = 129) (as in examples (6) and (7)), as these are well known to behave differently from other verbs (cf. Hopper & Thompson 1984), and furthermore we are interested here in agreement marking, something which is not relevant for these forms. Also excluded were the rare cases of non-agreement (either nós with a third-person singular verb (N = 14), as in (8)) or a gente with a first-person plural verb (N = 1) (as in (9)), as the low numbers do not permit a full analysis of these forms; truncated, or incomplete utterances (N = 155) (as in (10)); verbs with post-posed subjects (N = 17), e.g. somos nós ‘it is us’, a specific kind of construction that should be analyzed independently; and quotes from written material (N = 19), which do not represent the speech of the community under study here.

(6) É igual aqueles gato no meio de estrada,
    a gente passando em cima ai passou ploft.
    ‘It’s the same as those cats in the middle of the street,
    we pass-GER over them, pffft.’ (C47: 165)
(7) *Uma vergonha... a gente falar em termo em salário de governo.*
   ‘It’s shameful that we talk, in terms of the salary offered by the government.’
   (C34: 177)

(8) *Nós tava conversando*
   ‘We were talking.’ (C45: 386)

(9) *A gente tínhamos que ser polivalente.*
   ‘We had to be polyvalent.’ (I23: 19)

(10) *Nós estuda – Eu por exemplo estudei – fiz a minha Faculdade... gratuitamente.*
    ‘We studied – I for example studied – went to college for free.’ (C39: 378)

These exclusions left a total of 2,603 tokens for analysis, 1,310 tokens of *nós* and 1,293 tokens of *a gente*. Three hundred tokens of *nós* occurred in hortative constructions, an environment where *a gente* does not occur. As a non-variable context, this was excluded from the variable-rule analysis, though it is important for our understanding of the spread of *a gente* and will be discussed below (Section 7.2). The remaining 2,303 tokens were subject to multivariate analyses using the program GoldVarb X (Sankoff, Tagliamonte & Smith 2005), a program which uses multiple regression to identify the effect of individual factors on variant choice when a set of factors is considered simultaneously. We conducted independent analyses of the extra-linguistic and linguistic conditioning of the variation. We will first review these results, in order to obtain a full picture of the variation, and will then consider in more detail the patterning that is frequency driven.

5. Extra-linguistic conditioning

Table 3 presents the results of a Variable-rule analysis of the contribution of extra-linguistic factors selected as significant to the choice of *nós* in those contexts where *nós* and *a gente* exist in variation. In this and following tables, the ‘input’ indicates the overall likelihood that the variant (in this case, *nós*) will occur. The fact that the input given here (.435) is just below .5 indicates that there is a slightly lower possibility that *nós* will be used rather than *a gente*. The first column lists the factor groups that were found to have a significant effect on pronoun choice, register and age, while gender, also included in the analysis, was not selected as significant. In the second column, the probabilities represent the weight that each factor contributes to the occurrence of the variant: the closer to 1, the more likely that *nós* will be used, and the closer to 0, the less likely that it will be (and the more likely that *a gente* will be used). Thus, for register, *nós* is most favored in the lecture data
(with a weight of .68), less favored in the interview data (.45), and most strongly
disfavored in the conversational data (.39), and for age, speakers over 50 favor
(.65), and those under 50 disfavor (.43), the use of \((nós) + V\text{-1PL}\). The range
provides an indication of the relative strength of each factor group in the analysis
(cf. Poplack & Tagliamonte 2001, p.93, Sankoff 1988), which are presented in order
of their strength: in these results, register (presented first in the table) has the
strongest effect, with a range of 29, and age has a slightly weaker effect, with a
range of 22. The third column shows the percentages of \((nós) + V\text{-1PL}\) in each
case, the fourth column the total number of tokens in each factor, and the fifth
column the percentage of the data each factor makes up. The most important result
for our purposes here are the relative weights, as reflected in the probabilities
provided in the second column, which indicate the constraint hierarchy, or direction
of effect, and it is on these results that we will focus our discussion.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Probability</th>
<th>% nós</th>
<th>N</th>
<th>% of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Register</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lecture</td>
<td>.68</td>
<td>61</td>
<td>695</td>
<td>28</td>
</tr>
<tr>
<td>Interview</td>
<td>.45</td>
<td>39</td>
<td>953</td>
<td>41</td>
</tr>
<tr>
<td>Conversation</td>
<td>.39</td>
<td>33</td>
<td>655</td>
<td>30</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td><strong>29</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51+ yrs</td>
<td>.65</td>
<td>58</td>
<td>684</td>
<td>70</td>
</tr>
<tr>
<td>20 – 50 yrs</td>
<td>.43</td>
<td>38</td>
<td>1619</td>
<td>30</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td><strong>22</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

log likelihood = -1475.387; chi-square/cell = 2.2229
Factor group not selected as significant: gender

Table 3. A Variable-rule analysis of the contribution of extra-linguistic factors
selected as significant to the probability of the use of nós

Considering in more detail these results, we note that register exhibits a neat
cline from the most formal (and hence conservative) lecture data, to slightly less
formal interview data, and then to the informal conversational data in their favoring
of the novel form \(a\ gente\) (cf. Albán & Freitas 1991, Lemos Monteiro 1994, Lopes
effect). This indicates that \((nós) + V\text{-1PL}\) remains the prestige form, but the fact that
\((a\ gente) + V\text{-3SG}\) is used close to one third of the time in the lecture data shows
how far it has spread across different domains of use. To further test these results,
we conducted independent analyses of the linguistic conditioning of this variation in each of the three registers. Interestingly, nearly identical results were obtained in each, suggesting that the significant difference we observe here is not due to different linguistic conditioning across the registers, but a different distribution of the factors within each.8

The results presented here for age consist of just two age groups, of 50 and below, and over 50 years of age. Although originally we considered three age groups, our analysis revealed that the ‘young’ (20-35 years) and ‘middle’ (35-50) age groups patterned identically, indicating that it is not meaningful to treat them as two distinct categories and thus they were collapsed (cf. Zilles 2004, for a similar breakdown of categories of age). The results show that speakers over 50 favor, and those under disfavor, the use of (nós) + V-1PL, but again, even older speakers use the novel form over half of the time.

Gender was not selected as significant in our data, though the ranking parallels that found in other studies, namely a favoring of the older form nós by men over women (e.g. Lemos Monteiro 1994, Lopes 1998, 2003, Omena 1996a, Zilles 2005). These results are consistent with a change in progress, as more formal registers and older speakers tend to represent more conservative varieties of the language (cf. Labov 1972, 2001). They also suggest a change from below (as proposed by Zilles 2005, p.51), and studies in socially stratified samples confirm this, finding a higher rate of use of a gente among groups of lower socio-economic status (cf. Zilles 2004, p.26, for a review).

6. Linguistic conditioning

Table 4 presents the results of a Variable-rule analysis of the linguistic factors selected as significant to the choice of nós. As can be seen, the factor groups that have a significant effect on pronoun choice are previous subject, tense, verb class, subject expression, meaning, clause type and presence of a modal verb. These results concord with those of previous studies (e.g. Lopes 1998, 2003, Naro, Görski & Fernandes 1999, Omena 1996b, 2003, Zilles 2005), though the linguistic conditioning has not received the same attention as the extra-linguistic conditioning, and thus it deserves more consideration. We will focus on the first five factor groups given in Table 4, as these have the strongest effect (as reflected in their ranges) and are the most relevant for our purposes.
### Total N

<table>
<thead>
<tr>
<th></th>
<th>2,303</th>
</tr>
</thead>
</table>

### % nós

<table>
<thead>
<tr>
<th></th>
<th>44%</th>
</tr>
</thead>
</table>

### Input

<table>
<thead>
<tr>
<th></th>
<th>.423</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th></th>
<th>probability</th>
<th>% nós</th>
<th>N</th>
<th>% of data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Previous subject</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nós</td>
<td>.83</td>
<td>80</td>
<td>337</td>
<td>15</td>
</tr>
<tr>
<td>other</td>
<td>.54</td>
<td>44</td>
<td>1551</td>
<td>67</td>
</tr>
<tr>
<td>a gente</td>
<td>.14</td>
<td>15</td>
<td>415</td>
<td>18</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td>.83</td>
<td>80</td>
<td>337</td>
<td>15</td>
</tr>
<tr>
<td><strong>Tense</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>preterit</td>
<td>.74</td>
<td>59</td>
<td>387</td>
<td>17</td>
</tr>
<tr>
<td>future</td>
<td>.67</td>
<td>55</td>
<td>229</td>
<td>10</td>
</tr>
<tr>
<td>present</td>
<td>.46</td>
<td>44</td>
<td>1264</td>
<td>55</td>
</tr>
<tr>
<td>imperfect</td>
<td>.30</td>
<td>23</td>
<td>423</td>
<td>18</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td>.14</td>
<td>15</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td><strong>Verb class</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>possession (ter ‘have’)</td>
<td>.76</td>
<td>71</td>
<td>483</td>
<td>21</td>
</tr>
<tr>
<td>copula</td>
<td>.66</td>
<td>55</td>
<td>174</td>
<td>8</td>
</tr>
<tr>
<td>speech</td>
<td>.46</td>
<td>39</td>
<td>212</td>
<td>9</td>
</tr>
<tr>
<td>other</td>
<td>.40</td>
<td>35</td>
<td>925</td>
<td>40</td>
</tr>
<tr>
<td>cognition/perception</td>
<td>.39</td>
<td>34</td>
<td>311</td>
<td>13</td>
</tr>
<tr>
<td>motion</td>
<td>.33</td>
<td>29</td>
<td>198</td>
<td>9</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td>.33</td>
<td>29</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td><strong>Subject expression</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>unexpressed</td>
<td>.80</td>
<td>59</td>
<td>1939</td>
<td>16</td>
</tr>
<tr>
<td>expressed</td>
<td>.44</td>
<td>41</td>
<td>364</td>
<td>84</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td>.39</td>
<td>34</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td><strong>Meaning</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>general</td>
<td>.61</td>
<td>55</td>
<td>938</td>
<td>42</td>
</tr>
<tr>
<td>specific</td>
<td>.42</td>
<td>35</td>
<td>1285</td>
<td>58</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td>.42</td>
<td>35</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td><strong>Clause type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>main</td>
<td>.53</td>
<td>47</td>
<td>1623</td>
<td>70</td>
</tr>
<tr>
<td>relative</td>
<td>.49</td>
<td>40</td>
<td>294</td>
<td>13</td>
</tr>
<tr>
<td>subordinate</td>
<td>.39</td>
<td>34</td>
<td>386</td>
<td>16</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td>.39</td>
<td>34</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>
The Role of Frequency in Brazilian Portuguese: Nós vs. a gente

<table>
<thead>
<tr>
<th>Presence of modal verb</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>absent</td>
<td>.5246</td>
<td>2005</td>
</tr>
<tr>
<td>present</td>
<td>.4029</td>
<td>298</td>
</tr>
</tbody>
</table>

*Range 12*

log likelihood = -1168.808; chi-square/cell = 1.4427

Factor group not selected as significant: polarity

Table 4. A Variable-rule analysis of the contribution of linguistic factors selected as significant to the probability of the use of nós

6.1 Previous subject

The strongest effect we observed regards the previous subject: where there is a first-person plural subject in the immediately preceding clause, speakers tend to reuse that same form, as has also been found in a number of other studies (Lopes 1998, 2003, Omena 1996b, 2003, Zilles 2005). The common pattern is illustrated in (11) and (12), and that seen in (1) and (2) where speakers switch between the two forms is in fact relatively uncommon.

(11) a gente → a gente

* A gente luta a gente batalha a gente faz uma Faculdade pra estudar.
  ‘We struggle, we fight, we go to university to study.’ (C 39: 344)

(12) nós → nós

* Nós não recebemos este aumento então nós estamos com o salário defasado.
  ‘We didn’t receive this raise, so we have (lit. ‘are with’) our salary outdated.’ (C 33: 228)

Both expressed and unexpressed subjects followed this pattern, suggesting that this repetition effect is based on the verbal morphology and not the pronoun itself. This can be understood as a kind of structural parallelism, or a priming effect, that is “the unintentional and pragmatically unmotivated tendency to repeat the general syntactic pattern of an utterance” (Bock & Griffin 2000, p.177), which has been widely observed in the literature (cf. Bock 1986, Cameron & Flores-Ferrán 2003, Poplack 1980, Scherre 2001, Travis 2007, Weiner & Labov 1983).

6.2 Tense

The effect for tense is also very strong in these data. Here, we collapsed the various Tense / Aspect / Mood categories into four, namely present, future, imperfect and preterit (each category including indicative and subjunctive moods as well as...
progressive and perfect aspects\textsuperscript{11}). Our results demonstrate that the preterit and future favor nós, the present shows little effect, and the imperfect favors a gente. Examples (13)-(16) illustrate the preferred patterns.

(13) *Preterit* → nós  
*Foi a primeira turma de supervisão aqui da U.E.C.E.*  
nós terminamos se não me engano foi em s:: setenta e:: setenta e oito.  
‘It was the first class of supervisors here at U.E.C.E.  
*We finished,\textsubscript{1pl}, if I am not mistaken, it was in ’70 and ’78.’ (C47: 26)

(14) *Future* → nós  
*O treinamento do dia que nós vamos passar naquele dia...*  
‘The daily training that we are going to show,\textsubscript{1pl} on that day’ (C7: 21)

(15) *Present* → a gente / nós  
*A gente faz uma média de oitenta a cem artigos por mês.*  
‘We do,\textsubscript{3sg} an average of 80 to 100 articles a month.’ (C30: 247)

(16) *Imperfect* → a gente  
*Aí a gente sentia apavorada às vez a gente queria comprar mas tinha medo de abrir a bolsa*  
‘Then we felt,\textsubscript{3sg} fear and sometimes we wanted,\textsubscript{3sg} to buy something  
but (we) were afraid,\textsubscript{3sg} of opening our purses’ (I6: 25)

These results are consistent with those from previous studies (Lopes 1998, pp.8-10, Lopes 2003, p.105, Naro, Görski & Fernandes 1999, p.205, Omena 1996b, p.202, Omena 2003, p.70), and have been explained in terms of phonic salience, or the degree to which differences in the phonetic forms of the first-person plural and third-person singular verb forms are “salient”. The hypothesis is that tenses for which the forms differ only minimally (such as the present in regular verbs and the imperfect) have readily adopted a gente, while tenses for which the forms differ in more marked ways (such as the preterit and future as well as the present in irregular verbs, e.g. é ‘(we) are,\textsubscript{3sg}’ / somos ‘(we) are,\textsubscript{1pl}’) are slower in adopting the change, and show a greater use of nós (cf. Naro & Lemle 1976, on the role of phonic salience in morphological change). As has been pointed out by Poplack (1992, 2001), however, the most phonically salient forms (such as the copula ser ‘be’, included in a category of its own by Naro, Görski & Fernandes (1999)) tend to be irregular verbs of very high frequency which offers an alternative account for the conservative nature of such forms, as we shall discuss below (Section 7). Even for the regular verbs there are other possible explanations for the patterning across the different tense and aspects, such as the discourse function of the preterit and
imperfect, and the grammaticization of the analytic future. We shall explore these alternative explanations here.

The very strong favoring of nós by the preterit leads Naro, Görski & Fernandes (1999) to suggest that this form is coming to be used as a preterit marker, in part to resolve the homophony between the first-person plural preterit and present with regular verbs (e.g. falamos could mean ‘we speak’ or ‘we spoke’). That is, (nós) + V-1PL comes to be understood as preterit and (a gente) + V-3SG as present (e.g. (nos) falamos ‘we spoke’ vs. (a gente) fala ‘we speak’). The range of 28 between the preterit (with a probability of .74) and present (.46) in our data does support such a differentiation. In order to investigate this further it would be necessary to consider potential ambiguity of the tokens in context, and whether person morphology is indeed used alongside other features (such as temporal adverbials) to resolve such ambiguity.

It is interesting to compare the preterit with the imperfect, which strongly favors a gente, with a weight of .30. This means that when faced with a past tense, speakers tend to choose either the preterit with (nós) + V-1PL, or the imperfect with (a gente) + V-3SG. We propose that this may be due to the discourse function of these two past tenses in conjunction with the origins of a gente as an indefinite pronoun. The preterit is generally considered a narrative tense, used to advance the story line in a narrative, while the imperfect is used for presenting background information (cf. Givón 1984, Hopper 1982, Silva-Corvalán 2001). As such, the imperfect, more than the preterit, is compatible with indefinite subjects (cf. Company Company & Pozas Loyo in press, on the grammaticization of indefinite forms in the history of Spanish). Thus, it may be the case that the indefinite pronoun a gente was rarely used as the subject of verbs in the preterit, while it was readily used with verbs in the imperfect, and that the distribution we see today is a hangover from that (despite the fact that a gente is no longer restricted to indefinite reference) (cf. Deutscher 2000, Traugott 2003, on the retention in contemporary language use of productive patterns of an older stage of the grammar). This proposal, which may present a fuller account than that of resolving ambiguity for the preterit and phonic salience for the imperfect, requires diachronic testing, which we leave for future research.12

The strong favoring of (nós) + V-1PL with the future tense we propose has its origins in the grammaticization of the analytic future. Our future tense category comprises several forms used to express future meaning (including the synthetic future and the future subjunctive), but the analytic future accounts for close to 80% (178/229) of future tokens, and thus, the results for this tense most strongly represent the behavior of this form.13 The analytic future is a ‘go’-based future, that is, it derives from the verb ir ‘go’ used with an infinitival verb (as seen in (14), nós vamos passar ‘we are going to show’). This is a relatively new future form in Brazilian Portuguese, which still shows some variation with the older synthetic future, though the latter has largely been lost (cf. Poplack & Malvar 2007). The
analytic future began to develop in the 15th Century (cf. Silva 2006, p.143), the same time at which *a gente* was coming to be used as an indefinite pronoun, and long before it came to be used as a first-person plural form in the 19th century (cf. Lopes 2003). Thus, at the time at which the analytic future was emerging, \((nós) + V-1PL\) was the primary first-person plural form in use, and hence the construction that developed for the analytic future was \((nós) vamos\). The continued favoring of *nós* in this context today we believe is related to the categorical use of *nós* in the formally identical hortative construction, e.g. *vamos dizer* ‘let’s say’. This will be discussed in detail below.

6.3 Verb class

Verb class was included to test whether there was any relation between the semantics of the verb and one or the other pronoun. Though verb class has not been considered in prior research on the variation between *nós* and *a gente*, studies in other areas have demonstrated a close relationship between subject and verb class. It has been found that certain subjects tend to pattern with specific verb classes, such as first-person singular with cognitive verbs and third-person singular with copulas in English conversation (Scheibman 2002), as well as for various registers of spoken Spanish (Vázques Rozas & García-Miguel 2006), and Brazilian Portuguese conversation (Silveira 2007).

We identified six verbal categories in the data, based primarily on Bentivoglio (1987) and Travis (2007), with the addition of one further category (possession) that consists of one verb (*ter* ‘have’), due to the extremely high frequency of this verb in the data (constituting over 20% of the analyzed tokens of first-person plural, see Table 4). Other categories are: copula (with three lexical types, *ser, estar, ficar* ‘be’), speech (e.g. *dizer* ‘say’, *falar* ‘speak’), cognition/perception (e.g. *ver* ‘see’, *saber* ‘know’), motion (e.g. *ir* ‘go’, *chegar* ‘arrive’), and a final category consisting of verbs that did not fit into any of the others (e.g. *fazer* ‘do’, *dar* ‘give’).

Verb class was found to have a significant effect on pronoun choice, with possession and copula verbs favoring *nós*, speech verbs having little effect and other, cognition/perception and motion verbs favoring *a gente*. A careful analysis of the different semantic classes reveals that not all verbs in the classes behave in the same way, and the results for some of the classes are better accounted for as a lexical effect of highly frequent verbs, rather than of the class as a whole.

This is most apparent with the so-called class of possession verbs, which consists of just one member, *ter* ‘have’, but it is also so for the copula and “cognition/perception” verbs. In the case of the copula verbs, of the three verbs that make up this class, two, *ser* and *estar*, occur most frequently with *nós* (70% and 56% of the time respectively), while *ficar* tends to occur with *a gente* (occurring with *nós* just 25% of the time). For cognition/perception verbs, the two most frequent verbs in this class, *ver* ‘see’ and *saber* ‘know’, occur with *nós* an average
The Role of Frequency in Brazilian Portuguese: Nós vs. a gente

of 42% of the time (40% and 47% respectively), but the remaining verbs in the class occur with this form just 18% of the time. This demonstrates that the patterning observed for a semantic class as a whole should be considered in relation to any lexical effects of frequent members of that class, as has been found by several other scholars (including Torres Cacoullos & Walker (2009) for the variable use of the complementizer that in English, and Travis (2006, 2007) and Silveira (2008) for subject expression in Spanish and Portuguese, respectively).

To further test for these lexical effects, an independent analysis of the five most frequent verbs in the corpus (ter ‘have’, ser ‘be’, fazer ‘do’, ver ‘see’ and ir ‘go’) was conducted. This revealed that ter ‘have’ and ser ‘be’ are the only two verbs in the corpus to favor nós, with identical weights of .67, while all other verbs showed a favoring of a gente (ver ‘look’ .35, ir ‘go’ .28, fazer ‘do’ .25). We believe that an explanation for this can be found in the fact that ter ‘have’ and ser ‘be’ are not just highly frequent with first-person plural subjects, but are overwhelmingly the two most frequent verbs in Brazilian Portuguese. This suggests that the conservative nature of high frequency forms (Bybee 1985, 2006, Bybee & Thompson 1997, inter alia) need not apply across the board, but may be restricted to cases of exceptionally high frequency (cf. Bybee 2006, for discussion of the role of extreme high frequency in grammaticization). We will return to this point below (Section 7.1).

6.4 Subject expression

Brazilian Portuguese has variable subject expression, though this is in the process of being lost, as subjects are coming to be expressed at a higher and higher rate (cf. Azevedo 1981, Duarte 1993, Silveira 2008). This being the case, we might expect the newer form to be obligatorily expressed, and the older form to retain some degree of variability. Furthermore, as a gente takes third-person singular verbal morphology, we may expect it to avoid non-expression as this may give rise to ambiguity in some contexts. This is precisely what was found, with unexpressed subjects very strongly favoring the first person plural form and expressed subjects favoring the third-person singular form. Indeed, the unexpressed tokens of a gente that did occur (accounting for 12% of the total number of its uses, 149/1,293) all occurred in contexts where a gente was either the subject of the immediately preceding clause, or was the subject in the near vicinity (as can be seen in (16) above), demonstrating that speakers avoid leaving a gente unexpressed where ambiguity is not readily resolved by the context, and that its non-expression is thus very limited, as has been noted in the literature (Albán, Cruz, Oliveira, Passos & Rapp 1991, p.109, Lopes 1998, p.6). Nós, on the other hand, does not show this same restriction – not only does it have nearly double the rate of non-expression than a gente (21%, 215/1010), but less than one half (44%, 95/215) of the unexpressed tokens have (nós) + V-1PL as the subject of the preceding clause (see
(1), where we have an unexpressed first-person plural verb with *a gente* as the subject of the preceding clause).

6.5 Meaning

Given its origins as an indefinite pronoun, we may expect *a gente* to be favored for more generic reference and *nós* for specific reference. While it appears that this was indeed the case for the use of *a gente* at least in some parts of Brazil 20 or 30 years ago, its use in specific contexts seems to be on the rise. In a comparison between data collected in Rio de Janeiro in the 1980s and 1999-2000, Omena reports that, while in the 1980s, *a gente* was used more than *nós* in indeterminate (or generic) contexts, this was no longer the case in the 2000 data (Omena 2003, p.68). Zilles observed something similar in Porto Alegre, namely a marked increase in the use of *a gente* with specific reference in the 1990s as compared with data from the 1970s (2004, p.39). In order to determine whether *a gente* retained a more general semantic sense, we tested for the effect of general vs. specific meaning. We defined general meaning as reference to a nonspecific group of people (such as "we Brazilians", "we teachers" and so on), and specific meaning as reference to a specifiable group of people including the speaker and others, possibly (but not necessarily) including the addressee. Examples of each are given in (17) and (18).

(17) **Specific → a gente**

Aí eu perguntei assim ‘P. ... e quanto é que é?’... ele disse  
’não se preocupe não que *a gente* paga’...  
*era cento e sessenta dólares... eu me lembro que ele me falou.*  
‘Then I asked, “P. and how much is it?” he said  
“don’t worry because **we will pay**-3SG for it’

it was a hundred and sixty dollars. I remember he told me.’ (C2: 654)

(18) **General → nós**

*O Brasil chegasse pro pessoal dos Estados Unidos*  
e cortasse a relação com os Estados Unidos relação político-econômica\(^{18}\)  
se o Brasil fizesse isso o que ia acontecer?  
**Nós** íamos ter um problema muito grave.  
‘If Brazil went to the United States  
and broke ties with them, political and economical ties,  
if Brazil did that, what would happen?  
**We would have**, **1PL** a very serious problem.’ (L17: 193)

Our results for meaning are contrary to what the etymology would predict, but in accordance with those of Omena (2003) and Zilles (2004): (nós) + V-1PL is favored in more general environments, and (*a gente*) + V-3SG in more specific
The Role of Frequency in Brazilian Portuguese: Nós vs. a gente

contexts. Also note that meaning has a comparatively weak effect in our data: with a range of just 19, it has among the lowest magnitudes of effect of the significant factor groups, indicating that the meaning distinction is not a driving force in the distribution of these pronouns.

7. Frequency effects

Scholars working within the usage-based model have observed that frequency of use is a primary mechanism in the formation of grammar (Bybee 1985, Bybee & Thompson 1997, Du Bois 1985, Givón 1979, Haiman 1994, Hopper 1987, Langacker 2000). Within this model, where grammar is understood to derive from the speaker’s experience with language, it is the frequent repetition of forms that gives rise to structure (cf. Bybee 2006). This is particularly evident in language change, where it has been shown that frequency of use plays a fundamental role in the way in which changes spread through a language (cf. Bybee 2003, 2006, Bybee & Thompson 1997, Bybee & Torres Cacoullos 2009, Poplack 1992, Torres Cacoullos 2006, Torres Cacoullos & Walker 2009, Tottie 1991, and papers in Bybee & Hopper 2001).

There are two types of frequency which act in different ways in language change: token and type frequency. Token frequency refers to the number of occurrences (tokens) of a given item in the corpus. In our corpus, (nós) + V-1PL has a total of 1,310 tokens and (a gente) + V-3SG 1,293; thus, they have a similar token frequency. Type frequency refers to the number of different patterns in which a form occurs. For the forms (nós) + V-1PL and (a gente) + V-3SG, we can consider the verbs with which they co-occur. In the data under study here, (nós) + V-1PL occurs with 181 different verbal types, while (a gente) + V-3SG occurs with 255 different types; thus a gente has a higher type frequency than nós. High token frequency can promote change such as phonological reduction, loss of internal structure and semantic bleaching, as production becomes automated for forms that occur together repeatedly (e.g. gonna for the future tense in English) (cf. Bybee 1985, 2003, 2006, Bybee & Thompson 1997). This is seen in the grammaticization of a gente from an NP to a pronoun, as discussed above (Section 2). At the same time, high token frequency can slow change, as forms that are frequently used together come to be accessed as a single unit and thus become entrenched and resistant to novel, more productive, forms (e.g. the English past tense, where more frequent verbs maintain the irregular pattern, e.g. keep / kept, while less frequent verbs regularize, e.g. weep / weeped) (cf. Bybee 1985, 2003, 2006, Bybee & Thompson 1997, Poplack 1992, Tottie 1991). Bybee & Thompson (1997) refer to these as the “Reduction” and “Conserving” Effect, respectively. Type frequency, on the other hand, affects language change because it is related to productivity: the greater the number of types with which a construction occurs, the more general (or productive) its use and thus its availability to extend to new contexts (Bybee &
This is evident in Smith’s (2001) work on the history of the *have* and *be* auxiliaries for the resultative / perfect in English, which existed in variation in Old English. Smith found that the high type frequency of the *have* construction allowed it to generalize (and thereby increase in use) over time, while *be* came to be restricted to highly frequent verbs (due to the Conserving Effect) and eventually lost, today being maintained in American English only in the highly frequent and entrenched form *is gone*. This is precisely what we observe for the variation between (*nós* + V-1PL and (*a gente*) + V-3SG: (*a gente*) + v-3SG, with a higher type frequency, is the more productive form. It therefore does not demonstrate the same strong lexical associations we observe for *nós*, and is more readily applied to new contexts. The continued use of *nós*, on the other hand, appears to be primarily due to the Conserving Effect, or its retention in highly frequent contexts. This is evident in its use with the two most frequent verbs in Brazilian Portuguese, *ter* ‘have’ and *ser* ‘be’, as well as in the hortative constructions and in the related future tense. We consider each of these forms to be prefabs (Erman & Warren 2000), or conventionalized collocations that represent the standard way of expressing these concepts, and therefore, if not wholly immune, at least strongly resistant to change. Below we discuss each of these contexts individually and address in more detail the role of frequency in their conservative behavior.

### 7.1 *Ter* ‘have’ and *ser* ‘be’

We discussed above the very strong favoring of *nós* by *ter* ‘have’ and *ser* ‘be’ and how this relates to the extremely high token frequency of these verbs. In a token frequency analysis in a corpus of 3¾ million words of spoken Brazilian Portuguese, *ser* ‘be’ and *ter* ‘have’ emerged as the two most frequent main verbs, *ser* ‘be’ with a standardized frequency of use of 142 tokens per 10,000 words, and *ter* ‘have’ with 87 tokens per 10,000 words. The next most frequent verbs after these two are *fazer* ‘do’ at 54 tokens, and *saber* ‘know’ at 47. Thus, although the difference in frequency between *ser* ‘be’ and *ter* ‘have’ is great (142 vs. 87), there is also a large gap between *ter* ‘have’ and the next most frequent verb *fazer* ‘do’ (87 to 54), from which the verbs cluster together relatively closely. It is not surprising, then, that we see effects for *ser* ‘be’ and *ter* ‘have’ that we do not see for the other verbs. Furthermore, although *ser* ‘be’ is overwhelmingly the most frequent verb in Brazilian Portuguese, *ter* ‘have’ is overwhelmingly the most frequent verb to occur with first-person plural subjects: as was noted above, *ter* ‘have’ alone accounts for 21% of the first-person plural tokens analyzed (Table 4), and furthermore, it accounts for 34% of all tokens of (*nós* + V-1PL where it exists in variation with (*a gente*) + V-3SG (345/1,010) (i.e. excluding the hortative constructions). The high frequency of these verbs (both in general and with first-person plural subjects) means that the constructions *nós* + *ser* ‘we + be’ and *nós* + *ter* ‘we + have’ have such a strong representation in the minds of speakers that they function as prefabs,
or as entrenched single units which are therefore resistant to change. That is, just as the irregular past in English is being lost for *weep* but retained for *keep*, (nós) + V-1PL is disfavored with most verbs, but not with *ter* ‘have’ and *ser* ‘be’. Indeed, it is not just nós and any form of the verbs *ser* ‘be’ and *ter* ‘have’, but, more specifically, *nós somos* ‘we are’ and *nós temos* ‘we have’, as both these verbs occur in the present indicative with nós approximately two thirds of the time (35/54 tokens for *ser* ‘be’ and 248/345 for *ter* ‘have’). Examples (19) and (20) illustrate these constructions.

(19) *Nós somos* talvez os dois mais novo do SINE.
    ‘We are-1PL probably the two newest at SINE (name of organization).’
    (C30: 948)

(20) *Nós temos* belas praias...
    *que nós temos* um bom caranguejo e uma boa água de coco.
    ‘We have-1PL beautiful beaches,
    we have-1PL good crab and good coconut milk.’ (C28: 223)

The fact that *a gente* has fully spread across the verb types of lower frequency of use, such that even verbs of relatively high token frequency (such as *fazer* ‘do’, *ir* ‘go’ and *ver* ‘see’) favor *a gente* over nós, demonstrates the advanced nature of this change. This was confirmed by a further independent analysis in which we tested for the effect of frequency using three categories of exceptionally high frequency (*ser* ‘be’ and *ter* ‘have’), high (*ir* ‘go’, *fazer* ‘do’, *saber* ‘know’, *dizer* ‘say’, *ficar* ‘be’) and mid/low (all other verbs). This analysis showed a significant difference between the exceptionally high frequency verbs *ser* ‘be’ and *ter* ‘have’ and the verbs that made up the high frequency factor, but no significant difference between high and mid/low frequency verbs, both groups showing a strong disfavoring of (nós) + V-1PL.

Note also that even *ser* ‘be’ and *ter* ‘have’ are not wholly resistant to *a gente*, occurring approximately 30% of the time with this pronoun Nevertheless, the lagging behind of these verbs is a prime example of the Conserving Effect of high frequency forms (Bybee & Thompson 1997).

7.2 Hortatives

The hortative constructions represent the only domain in our data to which *a gente* has not spread. This domain consists of two different formal types, *vamos* + V-INF and the present subjunctive. We consider these to be constructions in that they show a degree of fixedness (they categorically take the first-person plural form and occur with unexpressed subjects), but they also have an open “slot”, which is filled by
different verb types (cf. Croft 2001). The most frequent types are listed in Table 5, with representative examples and the distribution of each.

<table>
<thead>
<tr>
<th>Hortatives</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <strong>vamos</strong> (+ V-INF) ‘let’s V’</td>
<td></td>
</tr>
<tr>
<td>vamos dizer</td>
<td>‘let’s say’</td>
</tr>
<tr>
<td>vamos ver</td>
<td>‘let’s see’</td>
</tr>
<tr>
<td>vamos</td>
<td>‘come on’</td>
</tr>
<tr>
<td>Other (48 lexical types)</td>
<td></td>
</tr>
<tr>
<td><strong>Total vamos</strong> (+ V-INF)</td>
<td></td>
</tr>
<tr>
<td>2. present subjunctive</td>
<td></td>
</tr>
<tr>
<td>digamos</td>
<td>‘let’s say’</td>
</tr>
<tr>
<td>vejamos</td>
<td>‘let’s see’</td>
</tr>
<tr>
<td>Other (2 lexical types)</td>
<td></td>
</tr>
<tr>
<td><strong>Total present subjunctive</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Distribution of hortative constructions

The first construction, **vamos** (+ V-INF), is largely formulaic, and its most frequent types (namely, **vamos dizer** ‘let’s say’, illustrated in (21), **vamos ver** ‘let’s see’ (22), and **vamos** (23)) function as discourse markers. That is, they play the pragmatic role of managing the interaction with the interlocutor, for example hedging an utterance by indicating that the wording is non-precise, as in (21), or drawing attention to the upcoming discourse as in (22) and (23), and so on.21 This construction is not limited to discourse marker uses, but is in fact relatively productive, as can be seen in the fact that, outside of these three most frequent constructions, it occurs with a further 48 lexical types. An example of its non-formulaic use can be seen in (24) (where, note, in making a metalinguistic comment about the upcoming utterance, it could also be said to function as a hedge). In all cases, the construction is autonomous from the source verb *ir* ‘go’, and any sense of motion has been lost.22

(21) *Existe uma:: política:: de:: vamo dizer de controle ambiental né?*  
‘There is a policy of let’s say-1PL of environmental control, isn’t there?’ (L52: 141)

(22) **Vamo ver qual é o que dá melhor.**  
‘Let’s see-1PL which will work better.’ (L55: 86)
The Role of Frequency in Brazilian Portuguese: Nós vs. a gente

(23) _Vamos pessoal comé que era a poesia da época?_  
‘Come on (lit. let’s go. 1pl) everyone, how was the poetry of that time?’  
(L36: 444)

(24) _Mas vamos ser franco... eu num tenho antipatia nem simpatia._  
‘But let’s be. 1pl frank, I don’t dislike it nor do I like it.’  
(L53: 536)

The second construction, the present subjunctive in main clauses, is almost wholly formulaic with a full 95% (103/108) of its tokens being limited to the one form, _digamos_ ‘let’s say’, illustrated in (25). This form functions similarly to _vamos dizer_ discussed above, and could also be considered a discourse marker used to hedge the material it marks. Other, non-formulaic, forms do occur, as illustrated in (26) and (27).

(25) _Eu fumava por exemplo digamos... três cigarro por dia._  
‘I used to smoke for example let’s say. 1pl three cigarettes a day.’  
(C48: 517)

(26) _Então vejamos esse soneto._  
‘Let’s look. 1pl at this sonnet.’  
(L3: 359)

(27) _Bom... voltemos à questão da seca aí._  
‘OK, let’s return. 1pl to the question of the drought’  
(I10: 30)

Key to the use of these forms is that they occur with an unexpressed subject. That is, while the form with an unexpressed subject, _vamos dizer_, could be interpreted either as the hortative ‘let’s say’ or as the future ‘we will say’, the form with an expressed subject, _nós vamos dizer_, only allows the future interpretation ‘we will say’. It is no doubt the case that the limited occurrence of unexpressed _a gente_ (i.e. third-person singular verb forms with first-person plural reference, see Section 6.4) inhibits the use of this form in this context. Nevertheless, it is interesting that, despite the increase in rates of subject expression in Brazilian Portuguese (cf. Duarte 1993), the hortative retains the older, less common form of the implicit subject, and it is worth asking why this should be the case.

As noted above, the most frequent forms to occur (_digamos / vamos dizer_ ‘let’s say’, _vamos ver_ ‘let’s see’ and _vamos ‘come on’_) can be considered discourse markers in the sense that, functionally, they play a primarily pragmatic role in terms of managing the interaction between interlocutors, and formally, they are invariable (always taking the first-person plural form _-mos_ and occurring only with unexpressed subjects) and commonly undergo phonological reduction (namely loss of word final /s/) (cf. Company Company 2006, Schiffrin 1987, Travis 2005, for
features of discourse markers). Through their high frequency of use they have become lexicalized and have lost their verbal status, and thus are immune to the change to the norm of expressed subjects, and thus also to that of \((a \text{ gente}) + V-3SG\) (cf. Albán et al. 1991, p.110, on the retention of \(tu\) in formulaic contexts).

It is interesting that even non-formulaic uses of the hortatives, as seen in (24), (26) and (27), show the same invariability in occurring with unexpressed subjects in the first-person plural form. This suggests that these forms represent the same construction which has given rise to the fixed discourse markers, and thus they are subject to the same effects. That is, the non-formulaic hortatives are formed by speakers on the basis of analogy with the more frequent fixed forms, and hence follow that same pattern (unexpressed subject + \(vamos\)), rather than the more widespread pattern evident in the language (which would lead to \(a \text{ gente vai}\)).

7.3 Future tense

As noted above, we attribute the favoring of \((nós) + V-1PL\) by the analytic future to the fact that, at the time of the development of the analytic future, \(nós\) was the first-person plural form in primary use, and thus the construction developed as \(vamos + V-INF\). What is interesting is that, although \((a \text{ gente}) + V-3SG\) has crept into the future domain, this is one of the few contexts where \(nós\) remains the preferred form. The analytic future is historically related and formally identical to the hortative \((ir + V-INF)\). We consider it to be a distinct construction because of its different semantics and its slightly different behavior (cf. Croft 2001): unlike the hortative, although it favors \((nós) + V-1PL\) it does also occur with \((a \text{ gente}) + V-3SG\), and it shows variable subject expression. However, it seems that it has enough of a link to the \(vamos\) hortative in the minds of speakers that it is influenced by the behavior of this form, resulting in a favoring of \(vamos\). Like the non-formulaic hortatives, then, the future would appear to be formed on analogy with the highly frequent, fixed discourse markers, rather than with the more productive, novel form, \(vai\).

8. Conclusions

The analysis of the variation between \((nós) + V-1PL\) and \((a \text{ gente}) + V-3SG\) outlined here indicates that this change is very advanced. \((A \text{ gente}) + V-3SG\) has generalized such that original meaning distinctions between the two forms have been lost, and it has spread across almost all domains of use, as was seen in the variable-rule analyses. We noted the high type frequency of \(a \text{ gente}\), indicating that it is the more productive of the forms and therefore the one that spreads most readily to novel contexts. \((Nós) + V-1PL\), on the other hand, exhibits a relatively restricted distribution, occurring primarily in a small number of environments, namely with \(ser\) ‘be’ and \(ter\) ‘have’, in the present subjunctive and \(vamos\) hortative constructions, and the related future tense. These constructions occur with sufficiently high
frequency that they have come to represent conventionalized expressions which are autonomous from the more productive pattern in the language involving the use of \((a \text{ gente}) + V-3SG\). More specifically, we have proposed that the retention of \(nós\) with \(ser\) ‘be’ and \(ter\) ‘have’ is attributable to the exceptionally high token frequency of these verbs, and have noted that the spread of \(a \text{ gente}\) is so advanced that the Conserving Effect of high token frequency (Bybee & Thompson 1997) is only evident in such cases of extreme high frequency. For the hortatives, we observed that the most frequent forms (\(\text{digamos} / \text{vamos dizer} \) ‘let’s say’, \(\text{vamos ver} \) ‘let’s see’ and \(\text{vamos} \) ‘come on’) are wholly formulaic, rendering them immune to the change both to expressed subjects and to \(a \text{ gente}\). These forms in turn affect the patterning of other, non-formulaic forms, such that all hortatives occur categorically with unexpressed first-person plural forms, and the \(\text{vamos}\) future tense shows a strong favoring of this form.

This study contributes to a growing body of research that highlights the role of frequency in language change, demonstrating that it is a factor that needs to be taken into account if we are to fully understand the way in which new forms come into use, old forms are lost, and new and old forms co-exist in variation. The role frequency plays in conditioning the \(nós + V-1\text{PL} / a \text{ gente} + V-3\text{sg}\) variation is evidence that grammar primarily consists of patterns of use that have become conventionalized through repetition. While conventions are not impervious to change, the stronger a convention, or the more entrenched a form, the greater the resistance it will show. Our results suggest that \(a \text{ gente}\) will continue to rise in use, taking over more and more domains where \(nós\) is still found. The highly frequent contexts we have identified are likely to be the last to see the loss of the first-person plural form, or indeed may remain a final vestige of \(nós\) indefinitely.

Acknowledgements

We are grateful to José Lemos Monteiro for making the PORCUFORT materials available to us, and Joan Bybee and Rena Torres-Cacoullos for valuable comments on earlier versions of this work.

Notes

1. Parentheses around the pronouns capture their variable expression, and this formulaic representation is intended to indicate that the variation we study here is first-person plural and third-person singular verbal morphology (with first-person plural reference) with both expressed and unexpressed subjects.

2. This indicates the register from which the example is drawn (Conversation (C), Interview (I), Lecture (L)), the number of the transcript, and the line number of the token under consideration.
3. Out of the total of 2,938 first-person plural tokens initially extracted for this study, there was just one token of *a gente* with the first-person plural verbal marking *-mos*, and 14 tokens of *nós* with third-person singular, or zero, marking. That is, non-standard agreement represents less than 1% of first-person plural tokens in the data. (Note that these tokens were excluded from the statistical analysis; see Section 4). While Zilles also found similarly low rates of non-standard agreement in her Porto Alegre data (2004, 2005), much higher rates have been reported for other Brazilian dialects. Naro, Görski & Fernandes (1999), for example, observe first-person plural agreement with *a gente* at a rate of 16% for young speakers from Rio de Janeiro, suggesting that a different kind of change may be going on in this variety.

4. Note that this is so even excluding hortative expressions, which categorically occur with *nós* (see Section 7.1). Given that, when used as discourse markers (e.g. *digamos* ‘let’s say’, *vejamos* ‘let’s see’) these expressions always have a general meaning, their inclusion would give an even higher rate of general *nós*.

5. We include here both specific and non-specific uses. Although it could be argued that the non-specific tokens of *a gente*, as in (3), do not have a first-person plural sense, given that *nós* is also used in this way, as in (5), we deemed it appropriate to include all such tokens.

6. Gerunds occurred categorically with *a gente*, and infinitive verbs with *a gente* 84% of the time.

7. It was necessary to conduct independent analyses, as the 2,303 tokens subject to the variable-rule analyses were not sufficient to analyze together the 10 factor groups that resulted from the three external and seven internal factor groups studied (see Sections 5 and 6 below). This is common practice in Variationist research, and is not problematic when, as noted by Poplack, analyses that include both internal and external factor groups usually obtain similar results to those that consider them independently (1997, p.304).

8. For example, the lecture data showed a higher proportion of use of the verbs *ter* ‘have’ and *ser* ‘be’, which highly favor (*nós*) + V-1PL.

9. This includes both those produced by the same speaker and those produced by another speaker, though it should be noted that only 75 out of the 2,303 tokens under analysis occurred with the immediately preceding clause produced by a different speaker, and of those 75, only 13 have a first-person plural subject.

10. A separate analysis in which preceding expressed and unexpressed subjects were kept distinct obtained the following results: preceding *nós* + V-1PL .87, preceding V-1PL .71, preceding other .54, preceding *a gente* + V-3SG .15, preceding V-3SG (with first-person plural reference) .06. There was no significant difference obtained between preceding expressed and unexpressed subjects, neither for (*nós*) + V-1PL nor for (*a gente*) + V-3SG.
11. The overwhelmingly most frequent TAM for each was the simple indicative, making up between 80% (for the future) and 99% (for the preterit) of each category.

12. We are grateful to Joan Bybee for drawing our attention to this possible explanation.

13. There are a further 44 tokens of the future subjunctive (of which 70% occur with a gente), five of the synthetic future (all of which occur with nós); two of the analytic and synthetic future combined (iremos + V-INF, both of which occur with nós) and two of the future progressive (one with a gente and one with nós). The analytic future tokens occur with nós 60% the time (105 / 178).

14. This class is made up of a total of 239 lexical types. The most frequent verb, fazer ‘do’, occurs 198 times and accounts for 21% of the class, but no other verb occurs more than 30 times. It is because of this large number of verb types with relatively low token frequency that this class accounts for 40% of the data.

15. These are the only verbs to occur more than 60 times in the corpus, and thus the only ones which permit independent analysis. The analysis was run on a total of 994 tokens with a corrected mean of .596. The same factor groups as those included in the overall analysis were run; all were selected as significant, and the constraint hierarchy for each was identical to that found in the overall analysis (log likelihood = -500.159).

16. As do all the figures given in this section, this excludes the hortative constructions, where the subject is categorically unexpressed.

17. There were 80 tokens for which it was not possible to determine whether they were being used with general or specific meaning, including 36 tokens of (a gente) + V-3SG and 44 of (nós) + V-1PL. These tokens were excluded from the analysis of the effect of meaning.

18. Note the lack of agreement here, as is common in Brazilian Portuguese.

19. This figure includes the hortative constructions, which were excluded from the variable-rule analysis as a non-variable context; it does not include the exclusions noted in Section 4.

20. This corpus was based on the full PORCUFORT corpus of 500,000 words and the spoken portion (comprising conversations and interviews) of the LAEL corpus of 3¼ million words (cf. Berber Sardinha 2005). LAEL is a constantly updated 223 million-word corpus of Brazilian Portuguese that originated in São Paulo in 1999. The corpus consists of several genres of written material and several registers of oral language. All verb forms were included in our frequency counts.

21. This is not unlike the use of vamos as a discourse marker in Peninsular Spanish (cf. Chodorowska-Pilch 1999, Fuentes Rodríguez 1988).

22. Vamos is also used on its own where a sense of motion is maintained to mean ‘we go / we will go’. This differs from the discourse marker use seen in (23), which functions to gain the attention of the interlocutor.
References


Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. Language 82, 711-733.


The Role of Frequency in Brazilian Portuguese: Nós vs. a gente


Scherre, Maria Marta Pereira. 2001. Phrase level parallelism effect on noun phrase number agreement. Language Variation and Change 13, 91-107.


