

## Change without change

ANTHONY J. NARO

*Universidade Federal do Rio de Janeiro*

EDAIR GÖRSKI

*Universidade Federal de Santa Catarina*

EULÁLIA FERNANDES

*Universidade do Estado do Rio de Janeiro*

### ABSTRACT

Brazilian Portuguese possesses two forms used as 1st person plural pronouns: *nós* and *a gente*, both meaning ‘we’. The form *nós* has always been pronominal, whereas *a gente* is derived diachronically from the noun phrase *a gente* ‘the people’. In accord with this historical evolution, the standard language prefers the use of the 1st plural verb desinence *-mos* with *nós*, as in *nós falamos* ‘we speak’ or ‘we spoke’. The 3rd person desinence *o* is reserved for *a gente*, giving *a gente fala* ‘we speak’ as the preferred form. In popular speech both *nós fala* and *a gente falamos* are used frequently. We examine the use of these variable forms across four generations in Rio de Janeiro. In the older generations, phonic salience is the principal controlling factor for both *nós* and *a gente*. Since preterit desinences are stressed more frequently than present desinences, this induces a biased surface distribution, with *-mos* occurring more frequently with past tense reference. Nonetheless, for older speakers tense does not play a statistically significant role. In younger speakers, tense becomes statistically significant as a determining factor in the use of the desinences, with preterit favoring *-mos* for both subject forms. So far, there has been no change in the grammar itself, but the locus of determination of the use of *-mos* seems to have shifted from saliency to tense across the generations. One can speculate that some time in the future *-mos* may become a preterit marker.

In this article we discuss a shift in the distribution of 1st person plural pronouns, as well as changes in the patterns of use of the corresponding verb inflections, in spoken Brazilian Portuguese across four generations of speakers from Rio de Janeiro. Through this study in apparent time, we reach the conclusion that there is a change in progress underway, even though none of the forms under discussion has undergone any surface change. All surface structures used by one generation may also be used by any other without any difference in meaning. Thus,

This is a revised translation of our paper, “Uma mudança lingüística em curso: a concordância com *nós/a gente*,” which was presented at the Seminário sobre Variação em Sintaxe, Universidade Federal do Rio de Janeiro, in 1983.

the change consists strictly of a shifting distribution of tokens with respect to the grammatical and semantic contexts in which they are used, with no change on the level of type. By extrapolation, we predict the occurrence of categorical changes that may, in the future, cause modification of the distribution of types.

We make critical use of several statistical concepts that by now have a long tradition of use in quantitative linguistic studies: relative weights, generated by the VARBRUL/GOLDVARB program, and their statistical significance as well as their order of selection in the step-up procedure used by this program. In particular, we discuss in some detail data in which the percentages of occurrence of a variant in certain environments do not correspond to the relative weights. Since understanding how this situation can come about is important to understanding our line of reasoning, we present a preliminary explanation.

#### FREQUENCIES AND RELATIVE WEIGHTS

The central problem to which VARBRUL analysis addresses itself is separating out the effect of each individual category posited as relevant to the chances of realization of a given variant in a complex environment defined by a set of independent categories. It does this by using a mathematical model of the joint effect of the factors which, together, make up the environment. The model is known as “logistic” and, when represented as a graph, takes the form of an extended S, a curve typical of the replacement of one variant by another over time in an evolutionary process.

Consider, for example, an early study of variable agreement, or concordance, within the noun phrase in spoken Brazilian Portuguese carried out by Braga (1977) and Scherre (1978).<sup>1</sup> For a typical plural noun phrase such as *oS meuS paiS* ‘(the) my parents’ there is variation in the use of each of the plural *-s* morphemes shown in the example. In the initial stages of their studies, the researchers hypothesized that the chance of non-occurrence of the plural morpheme was greater in words in which the singular/plural opposition was simply *0/s* (*casa/casaS* ‘house/houses’) than in those in which the opposition was more complex, such as *milhão/milhÕES* ‘million/millions’ or *hotel/hotêIS* ‘hotel/hotels’.<sup>2</sup> However, this hypothesis seemed to be undermined by the first results they obtained from a preliminary study of all the marked and unmarked elements of the noun phrases found in an exploratory corpus consisting of hour-long recordings of two speakers: 49% of the simple opposition forms were found to have the plural marker, whereas the forms of the more complex opposition (where more plural marking was expected) had an explicit plural morpheme in only 36% of the cases.

In order to resolve this impasse, we must take into account a second category that is also relevant to the rate of plural marking in noun phrases: namely, the linear position of the element within the noun phrase. The following three positions were recognized:

Position 1: first markable element of the NP; in this position marking was present in about 98% of the cases (*a(s) minha(s) amiga(s)* ‘(the) my friends (fem. pl.)’).

Position 2: second markable element of the NP; in this position marking was present in about 18% of the cases (*a(s) minha(s) amigas(s)* ‘(the) my friends (fem. pl.)’).

Position 3: third markable element of the NP; in this position only 10% of the cases had a plural marker (*a(s) minha(s) amiga(s)* ‘(the) my friends (fem. pl.)’).

Thus, there are at least two simultaneous factors at work in the variable use of a plural marker on an element of a noun phrase: the morphological class of the pluralizable element and its linear position within the noun phrase.

Let us examine the distribution of the data for the two morphological categories, beginning with the simple opposition (*0/-s*). The distribution of noun phrase elements belonging to this category with respect to the three linear positions (both with and without plural marking) is as follows:

position 1: 39%  
position 2: 55%  
position 3: 6%

In the case of the complex opposition the distribution of the linear positions occupied by noun phrase elements is as follows:

position 1: 0%  
position 2: 91%  
position 3: 9%

Since there are very few data in position 3, let us consider first positions 1 and 2. We find that, in the case of the simple opposition, there is a slightly higher concentration of tokens in position 2 than in position 1 (55% vs. 39%). For the more complex opposition there are no occurrences at all in position 1, and virtually all the data are concentrated in position 2 (0% vs. 91%). What is the reason behind this biased distribution? The fact of the matter is that position 1 in the noun phrase is normally occupied by a determiner, such as an article or a demonstrative, and these items all belong to the simple opposition (*a/as* ‘the (fem. sg./pl.)’, *o/os* ‘the (masc. sg./pl.)’, *este/estes* ‘this/these (masc.)’, etc.). In the data collected in the exploratory corpus, there were no occurrences of noun phrases without articles of the type *hotéis bonitos* ‘pretty hotels’, with the complex opposition in position 1. This construction type, although perfectly acceptable within formal grammar, is very infrequent in real usage and was not found in the initial corpus.

Recall that the simple opposition has a significant part (39%) of its occurrences in position 1, where the plural morpheme is almost always used by speakers (98% presence). On the other hand, all of the tokens of the complex opposition are to be found in positions 2 and 3, where the plural marker is almost always absent (18% presence for position 2 vs. 10% presence for position 3). These distributional facts cause a spurious inflation in the overall frequency of marking in the simple opposition class: the simple opposition shows high frequencies of marking simply because it is concentrated in position 1, where marking is nearly obligatory.

Since it seems that position 1 is causing the problems in our interpretation of the results, let us remove from the data all tokens in this position, retaining only the data from positions 2 and 3. Having done this, the new distribution of data becomes:

	Simple Opposition	Complex Opposition
Position 2	90% (55/61)	91%
Position 3	10% (6/61)	9%

The distribution of the complex opposition does not change when the position 1 data are removed because it had no occurrences at all in this position. But for the simple opposition the situation is radically different, since in this case 39% of the data are lost. As can be seen, the new distribution is very evenly balanced, eliminating the problems of analysis discussed earlier.

When we consider only the data from positions 2 and 3, the frequencies of presence of the plural marker for the two morphological categories become 36% for the complex opposition and 17% for the simple opposition. The frequency for the complex category (36%) does not change when the position 1 data are removed simply because this category did not occur in position 1. In the case of the simple category the frequency of marking falls from 49% to 17% after the removal of the frequently marked cases of position 1, which were inflating the results. The new frequencies, obtained from a nonbiased sample, confirm the initial hypothesis as to the greater favoring of marking for the complex opposition.

The moral of the facts just reviewed is that raw frequencies, although concrete and intuitively real, can be deceptive because their calculation does not take into account the relationships between the categories that influence the outcome of linguistic variation. When the data for noun phrase agreement were run in the VARBRUL program, the following logistic model relative weights were obtained for the morphological opposition group: 0.66 for the simple opposition and 0.84 for the complex opposition. These results, which confirm the initial hypothesis and agree with the frequencies for the nonbiased data of positions 2 and 3, were calculated using the complete set of data, including the data from position 1. This constitutes a practical demonstration that the logistic model is capable of overcoming difficulties of analysis caused by an uncritical use of raw frequencies.

The version of VARBRUL used to obtain the results reported here uses an iterative process to obtain the logistic weights that most closely approximate the reported frequencies of occurrence of the variants in each context for which we have empirical data. It does this using a step-up procedure, in which weights are first calculated for each postulated group of factors independently of all the other groups. The group that accounts for the largest proportion of variation is identified and selected. Then each of the remaining groups is run together with the first selected group, and the set of two groups that accounts for the greatest proportion of variation is selected. The procedure then continues selecting groups until the inclusion of new groups does not lead to a significant improvement in the proportion of variation accounted for. Thus, the factor groups not selected in this

procedure do not play a significant role in determining the chance of use of the variant under analysis.

#### 1ST PERSON PLURAL PRONOUNS AND VERB INFLECTIONS

In standard Portuguese the 1st person plural nominative pronoun is *nós* 'we', and the corresponding verb takes the *-mos* inflection categorically. A typical example is *nós falamos* 'we speak'. However, there is an alternative to the 1st plural subject pronoun: *a gente*, descended from the noun phrase of the same form, meaning 'the people'. In the standard language the verb used with *a gente* takes the 3rd person singular verb ending: *0*. A typical example is *a gente fala* 'we speak'. Furthermore, the use of subject pronouns, although frequent, is not obligatory, and, at least in the informal spoken language, the *-mos* inflection is often omitted with *nós* and used with *a gente*, despite the categorical rules to the contrary of the standard. Thus, the variants in (1) for 'we speak' can be found in spoken informal Brazilian Portuguese.

- (1) *nós*: *nós falamos* or *nós fala*  
*agente*: *a gente fala* or *a gente falamos*  
*0*: *fala* or *falamos*

The forms *nós falamos* and *a gente fala* are standard; *nós fala* and *a gente falamos* are nonstandard.<sup>3</sup> The form preferred for *0* subjects depends on which of the full pronouns is understood as the antecedent. Even though all of the forms listed in (1) are semantically 1st plural, we will refer to the *-mos* ending as 1st plural and the *0* ending as 3rd singular, based on their diachronic origin.

The nonstandard usage of *-mos* is undergoing a process of rapid expansion at the present moment, as can be seen by the comparison of younger speakers with their elders in (2).

- (2) Frequency of use of the *-mos* inflection with *nós* and *a gente*

	<i>-nós</i>	<i>a gente</i>
6–20 years	374/935 (40%)	427/2,673 (16%)
21+ years	609/924 (66%)	219/2,384 (9%)

Clearly, there has been a dramatic decrease in the use of *-mos* with *nós* and an expansion of *-mos* to *a gente*. The latter development is particularly striking in view of the long-term trend toward the loss of verbal inflections in Portuguese in particular and in the Romance languages in general.

For regular verbs the *-mos* forms are ambiguous between the present and the preterit tenses. Thus, *falamos* can mean either 'we speak' or 'we spoke',<sup>4</sup> *aprendemos* can be either 'we learn' or 'we learned', and *partimos* corresponds to both 'we leave' and 'we left'. The *0* ending forms, on the other hand, are unambiguously present tense: *fala* 'speak' is quite different from *falou* 'spoke', and the same is true of *aprende* 'learn' versus *aprendeu* 'learned' and *parte* 'leave' versus *partiu* 'left'. This fact turns out to be critically important in the developments studied here.

## THE SAMPLE

Previously collected corpora did not provide a sufficient amount of information on the realization of semantically 1st person plural subjects and the corresponding verb forms for our purposes. Similarly, random observation of natural speech within the amount of time available to us would not have provided enough data for statistical processing. For this reason we created and tested a questionnaire with the specific aim of encouraging speakers to use clauses containing *nós* or *a gente* or at least a semantically 1st plural verb form. Our questions typically included the following: *O que vocês fazem no trabalho?* ‘What do you (pl.) do at work?’ *Como você e sua família passaram o último domingo?* ‘How did you and your family spend last Sunday?’ *Como você e seus amigos vão organizar a próxima festa junina?* ‘How are you and your friends going to plan the next winter party?’ The questions were designed to elicit responses in different tenses, and the topics of conversation were varied, ranging from work, morals, and religion to daily life, leisure activities, social concerns, and love.

In the early 1980s, we conducted, recorded, and transcribed 64 individual interviews, each approximately 45 minutes in duration. The speakers we recorded were all born in the municipality of Rio de Janeiro and were from the lower socioeconomic levels, as determined indirectly from factors such as profession, employer, and type of residence. The place of residence of the speakers was distributed randomly throughout the municipality. At least two separate meetings were held with each speaker: the first, a preliminary session, was intended to obtain relevant social data and to familiarize the speaker with the interview situation, and the second, the interview itself, was conducted in accord with the questionnaire, modified to suit the personal interests of each person. Speakers were told that the aim of our research was to study the *carioca* (Rio de Janeiro) way of life and *carioca* opinions. No specific mention of language was made.

We organized the speakers into four age groups: 16 children were between 6 and 12 years of age, 16 young people were between 13 and 20 years, 16 younger adults were between 21 and 40 years, and 16 older adults were above 41 years old. In each age group, half the speakers had less formal education (at most, third grade), while the other half had spent more time at school (from the fourth to the eighth grade). Each group was equally divided between the two sexes. Thus, we obtained 16 groups of 4 speakers each, stratified for age, education, and sex.

## INDEPENDENT VARIABLES

*Phonic salience*

Previous studies on variation in the use of agreement markers in subject, verb, and predicate phrases have shown that the degree of differentiation of the competing forms—known as “oppositional” or “phonic salience”—is of fundamental importance in determining which form is preferred in a given environment. For purposes of the present study we set up a salience hierarchy for the 1st plural *-mos* and 3rd singular *Ø* forms associated with the semantic 1st person plural reference (exemplified in (1)). The basic dimension in this hierarchy is stress: its lower

TABLE 1. *The hierarchy of oppositional, or phonic, saliency*

	Example	Description
1	<i>falava/falávamos</i> 'we spoke'	The opposition -V/-V- <i>mos</i> is unstressed in both forms
2	<i>fala/falamos</i> 'we speak' <i>trouxe/trouxemos</i> 'we brought'	The opposition -V/-V- <i>mos</i> is stressed in one of the forms
3	<i>está/estamos</i> 'we are' <i>tem/temos</i> 'we have'	The opposition -V/-V- <i>mos</i> is stressed in both forms
4	<i>comeu/comemos</i> 'we ate' <i>partiu/partimos</i> 'we left' <i>vai/vamos</i> 'we go' <i>foi/fomos</i> 'we went' or 'were'	The opposition -V/-V- <i>mos</i> is stressed in both forms, and the 3rd sg form shows a diphthong with an upglide that does not appear in the plural
5	<i>falou/falamos</i> 'we spoke' <i>é/somos</i> 'we are'	The opposition -V/-V- <i>mos</i> is stressed in both forms, and the stressed vowel changes

levels consist of morphological 3rd singular/1st plural oppositions in which the *-mos* inflection and the preceding vowel (usually the thematic vowel) are unstressed; on its upper levels this vowel is stressed. The hierarchy is set out in detail in Table 1.

Level 1 of the saliency hierarchy, its lowest level, consists of forms such as *falava/falávamos* [faláva/falávamus]<sup>5</sup> 'we used to speak', in which the opposition is realized entirely in final unstressed syllables in both forms. All forms on this level are in the imperfect. On level 2, *fala/falamos* [fála/falámus] 'we speak', the vowel is accented in only the *-mos* form. This level contains most of the present tense forms as well as certain preterit forms.

The remaining levels present a stressed opposition in both forms. The ordering of oppositions on the stressed levels is determined by means of the phonic substance that differentiates the forms. On level 3, *está/estamos* [ístá/ístámus] 'we are', and on level 4, *partiu/partimos* [partíw/partímus] 'we left', the stressed vowel is the same in both forms, but level 4 has a glide in the 3rd singular form that is not found in the 2nd plural. Level 3 contains only present tense forms, whereas level 4 has preterit as well as a few present tense forms. Finally, on level 5, *falou/falamos* [falów]/[falámus] 'we spoke', there is a very salient change in the stressed vowel in addition to the differences found on level 4. Level 5 also includes the case of suppletive forms, such as *é/somos* [ɛ]/[sómus] 'we are'. Most level 5 forms are preterit.

Although the oppositional hierarchy is determined exclusively by criteria of phonic salience, it induces a partially biased distribution with respect to tense. In particular, preterit forms can occur only on levels 2, 4, and 5 and are concentrated mainly in levels 4 and 5. The present, which can occur on all levels except the lowest, is concentrated in levels 2 and 3. Level 1 exhibits only imperfect forms, as we have already noted. The overall distribution, then, shows preterit forms mainly on the higher levels of the hierarchy, whereas the present is found mostly on the lower levels. In order to examine the consequences of this biased distribution, we set up a tense variable, opposing the preterit to the present.

In the quantitative analysis of the *-mos/0* alternation in our corpus, we did not include level 1 in the data because *-mos* had a very low frequency of occurrence, both for *nós* and for *a gente*, on this level. In our entire sample, the frequency of occurrence of *-mos* reached only about 3.8% on this level. Given this situation, we decided to eliminate level 1 from our quantitative analysis since our primary interest was to examine the interplay between present and preterit forms, which did not occur on this level. In the case of *nós*, we had to disregard the distinction between levels 4 and 5 for lack of sufficient relevant data. We were, however, able to retain all four levels of the hierarchy in the analysis of *a gente*.

*Position of the subject with respect to the verb*

For the position of the subject with respect to the verb we distinguished two categories: near and distant. We considered the subject to be near the verb when it is placed before the verb and is separated from it by not more than five syllables of phonic material.<sup>6</sup> Thus, in (3), the first verb is considered to have a near subject, and the second verb is classified as having a distant subject.

- (3) *A gente sempre reúne o pessoal, depois, fala com eles*  
 ‘We always meet with the group, then (we) speak with them.’

We did not distinguish between a distant and a *0* subject.

THE USE OF *-mos* WITH *nós*

As we have already noted, in the case of the 1st plural pronoun *nós* the saliency hierarchy is made up of three levels: 2, 3, and 4–5. For our four age groups, from oldest to youngest, we obtained totals of 350, 616, 550, and 644 tokens, respectively, for a grand total of 2,160 occurrences of verb forms with *nós* as subject.

For the two oldest groups of speakers, VARBRUL chose only the oppositional saliency factor group as statistically significant. In the two youngest groups, the tense group was also selected, after saliency, on the second level. In Table 2 we show the relative weights for these two variables, with the nonsignificant results for tense for the older groups in boxes.<sup>7</sup>

Our results show that phonic saliency is statistically significant and is chosen by VARBRUL in first place for all four age groups. Both the frequencies and relative weights for each of the three levels of saliency increase progressively in each of the age groups. Although there are some fluctuations in the raw frequencies across the age groups, the relative weights increase in the same way for all groups. This means that, with subject *nós*, 1st plural *-mos* tends to appear most often when it is more salient with respect to the form with the 3rd person *0* desinence.

The tense variable is statistically significant only for the two youngest age groups, although the raw frequencies for all four groups exhibit reasonably large differences in percentages of actual use. The oldest speakers, for example, use *-mos* in only about 70% of all non-preterit contexts, but they exceed 95% for preterits. The VARBRUL program, nonetheless, attributes almost equal weights of



TABLE 2. *Results for use of -mos with subject nouns in four age groups*

Factor Groups	Frequency of -mos	Factor Weight
<b>Older Adults</b>		
Phonic salience (significance = 0.0)		
2	32/58 (55.2%)	0.21
3	17/23 (73.9%)	0.37
4-5	259/269 (96.3%)	0.87
Tense (significance = 0.76)		
Present	84/117 (71.8%)	<u>0.47</u>
Preterit	224/233 (96.1%)	<u>0.53</u>
<b>Younger Adults</b>		
Phonic salience (significance = 0.0)		
2	55/142 (38.7%)	0.07
3	46/51 (90.2%)	0.52
4-5	419/423 (99.1%)	0.93
Tense (significance = 0.34)		
Present	171/263 (65.0%)	<u>0.40</u>
Preterit	349/353 (98.9%)	<u>0.60</u>
<b>Young People</b>		
Phonic salience (significance = 0.0)		
2	13/331 (5.9%)	0.14
3	10/45 (22.2%)	0.42
4-5	230/254 (90.6%)	0.90
Tense (significance = 0.0)		
Present	74/333 (22.2%)	0.25
Preterit	179/187 (95.7%)	0.75
<b>Children</b>		
Phonic salience (significance = 0.0)		
2	13/268 (4.9%)	0.13
3	9/28 (32.1%)	0.60
4-5	325/348 (93.4%)	0.82
Tense (significance = 0.0)		
Present	48/340 (14.1%)	0.13
Preterit	299/304 (98.4%)	0.87

about 0.5 to both tense categories and classifies these weak results as statistically nonsignificant. We will examine this situation in greater detail because it is exactly this sort of lack of agreement between frequencies and relative weights that plays an important part in the development of our line of reasoning.

The explanation for the apparently inconsistent situation just outlined lies in the biased distribution of tense with respect to oppositional saliency. As we have already remarked, most preterit forms are located on the higher levels of saliency, whereas many non-preterits are on the lowest levels of the hierarchy. The first step of the VARBRUL program was to determine which of these two partially correlated variables accounts for the greatest proportion of variation. This calculation showed that saliency is much more powerful than tense, in the sense that,

in isolation, it accounts for more of the total variation. Once saliency enters the calculations, there is not much left for tense to explain—a high rate of occurrence of *-mos* is exhibited by preterits simply because they happen to possess high levels of saliency. In other words, the high rate of marking of preterits in comparison to non-preterits is illusory. At this stage of their historical development, preterits are marked because, by coincidence, they occur on high levels of the saliency hierarchy and not because of any inherent property of their own. We can show that this is true as follows. If the weights associated with the tense variable are weak because of the simultaneous force of the saliency variable, it follows that, if we remove saliency from the calculations, the polarization of the results for tense should increase. And, in fact, the results shown in Table 2 for the Older Adults change from 0.47/0.53 to 0.76/0.24 under these conditions. We wish to emphasize, however, that this second set of weights, as well as the raw frequencies themselves, are the gratuitous result of the biased distribution of tense with respect to saliency. It is saliency—not tense—that is the significant determining variable because saliency accounts for more of the overall variation. We conclude, then, that the use of the *-mos* desinence with subject *nós* by the Older Adults and Younger Adults is governed by saliency alone. Other apparent effects, such as that of tense, are merely consequences of their distribution with respect to saliency. We do not wish, however, to affirm that the phenomena referred to here as “apparent” and “gratuitous” are any less real. To the contrary, it is clear that they exist in the superficial distribution of the data and are capable of causing profound structural consequences, as we shall see.

Turning our attention now to the results for the Young People and Children, we see that two variables were selected: first, oppositional salience, and second, tense. Relative weights and frequencies agree in attributing greater rates of occurrence of *-mos* to the preterit. The fact that the VARBRUL relative weight results for tense reach the level of statistical significance for the younger speakers shows that there is an important difference between their speech and that of the older speakers. In spite of the biased distribution, for the younger speakers the tense variable makes its own contribution to the chances of realization of *-mos*, favoring the presence of *-mos* in the preterit and disfavoring it for other tenses. In other words, the only factor for the older speakers that has a significant effect is salience—a salient form has a greater chance of showing *-mos*, while its tense category has no effect. For the younger speakers, however, even a salient form has less chance of occurring with *-mos* if it is not a preterit. This logic can be verified in Table 3, where we compare the behavior of certain equally salient forms that have different tenses.<sup>8</sup> In the first three lines of the table, we have the following types: *comeu/comemos* ‘ate’ (including data for all regular preterits of the second and third conjugations), *foi/fomos* ‘were’ or ‘went’ (preterit, data for this particular item only), and *vai/vamos* ‘go’ (present, data for this particular item only). These three categories have identical saliency: the 1st plural is formed by removing the final semivowel of the singular and adding *-mos*. The first two, however, are preterit, and the last is present. The fourth line contains *fala/falamos* ‘speak’ (all regular present tense forms of the first conjugation), a low saliency category.

TABLE 3. Comparison of rate of use of *-mos* with subject *nós* in four verb forms

	Adults	Young People	Children
<i>comeu/comemos</i> 'ate'	85/85 (100.0%)	27/28 (96.4%)	43/44 (97.7%)
<i>foi/fomos</i> 'were' or 'went'	114/117 (97.4%)	20/21 (95.2%)	43/43 (100.0%)
<i>vai/vamos</i> 'go'	95/101 (94.1%)	46/61 (75.4%)	23/42 (54.8%)
<i>fala/falamos</i> 'speak'	56/134 (41.8%)	8/131 (6.1%)	10/186 (5.4%)

For the adult groups, the first three verb forms, all equally salient, are close to categorical in the use of *-mos*. For the Young People, however, there is a clear distinction between the preterits of the first two lines of Table 3, close to 100%, and the present of the third line, at about 75%. For the Children, the difference between the salient preterits in the first two lines and the salient present in the third line is even greater, reaching about 50%. What we see, then, is a progressive weakening of the force of oppositional salience, with a concomitant strengthening of tense. For the older speakers high salience is sufficient to favor the use of *-mos*, but for younger speakers it is necessary to add preterit tense to the mix in order to reach the same result. The effect of salience remains valid even for the Children, since, for example, the level of 55% of marking attained by the high salience present tense form *vai/vamos* 'go' is well above that of the 5% associated with the low salience present tense *fala/falamos* 'speak' in the last line of the table. The overall picture we get from Table 2 is the promotion of the tense variable, reaching statistical significance in the Young People, with increasingly polarized results in the four age groups, and the stability of the salience variable.

#### THE USE OF *-mos* WITH *a gente*

We turn now to *a gente*, derived diachronically from a nominal form but used as a 1st plural subject pronoun alternating with *nós*. We find a somewhat curious situation. In the standard language, in accord with grammatical tradition, the 1st plural desinence *-mos* cannot occur with the historically 3rd singular form *a gente*. In fact, the Older Adults use *-mos* with *a gente* rather infrequently, reaching only about 10% on the average. Nonetheless, the desinence *-mos*, in and of itself, is favored by the standard language. We would expect that, although high oppositional salience may favor *-mos* with *a gente* in the same way as we have seen with *nós*, other structural features such as distance should work in the opposite way, disfavoring *-mos* when the verb form is close to *a gente*. This seems to be a typical situation in which a favored form is diffusing to a disfavored environment. For this reason, the type of structural features that favor *-mos* for *nós* may disfavor it for *a gente*, even though *-mos* itself is highly favored.

The data for *a gente* exhibit several characteristics that make it quite different from the data found with *nós*. In the first place, the promotion of the tense vari-

TABLE 4. *Results for use of -mos with subject a gente in two reduced age groups*

Factor Groups	Frequency of -mos	Factor Weight
Older Adults		
Phonic salience (significance = 0.0)		
2	18/525 (3.4%)	0.17
3	4/93 (4.4%)	0.22
4	78/152 (51.3%)	0.77
5	83/107 (77.6%)	0.84
Tense (significance = 0.0)		
Present	59/719 (8.2%)	0.29
Preterit	124/158 (78.5%)	0.71
Position (significance = 0.0)		
Far	96/326 (30.1%)	0.64
Near	88/561 (16.7%)	0.36
Younger Adults, Young People, and Children		
Tense (significance = 0.0)		
Present	230/2,826 (8.1%)	0.22
Preterit	730/861 (84.8%)	0.78
Phonic salience (significance = 0.0)		
2	74/2,081 (3.6%)	0.17
3	27/286 (10.1%)	0.40
4	350/735 (47.6%)	0.69
5	509/603 (84.4%)	0.76
Position (significance = 0.0)		
Far	434/1,187 (36.6%)	0.61
Near	526/2,500 (21.0%)	0.39

able is much more advanced for *a gente* than for *nós*. Thus, even for the oldest speakers in our sample, we find the present tense high saliency form *vai/vamos* 'go' with a much lower rate of realization of *-mos* than that of the corresponding preterit high saliency form *foi/fomos* 'were' or 'went', as shown in (4).

(4) Frequency of use of *-mos* with *a gente*: Oldest age group

*foi/fomos* 18/23 (78%)  
*vai/vamos* 30/100 (35%)

Let us recall that we found this sort of differentiation between the two forms only beginning with the Young People in the case of *nós*. Because of the more advanced state of evolution of the process with *a gente*, the behavior of the three youngest groups is quite uniform, and there is no need to study them separately as we did with *nós*. For this reason, in Table 4 we distinguish only two age groups, opposing the oldest group to the three youngest groups. The larger number of data obtained for *a gente* permits us to reintroduce the distinction between levels 4 and 5 of oppositional saliency. We must eliminate level 1, however, because *-mos* is nearly categorically absent on this level, as already noted in connection with

Table 1. The results for *a gente* are found in Table 4. The factor groups are listed in the order in which they were selected by VARBRUL.

For the Older Adult group, we obtained a total of 877 tokens for *a gente* (183 with *-mos*), resulting in an overall rate of marking of 21%. For these data, VARBRUL chose phonic salience in first place, followed by tense and position. The weights for the first two factor groups are similar to those found for *nós*.

The results for the salience variable show that *-mos* occurs more frequently in the higher levels of the hierarchy and confirms the distinction between levels 4 and 5, postulated in Table 1.<sup>9</sup> Tense, significant here even for the oldest speakers in our sample, exhibits highly polarized weights. Finally, the positional variable was selected by VARBRUL in third place. This variable shows lower chances of occurrence of *-mos* when subject *a gente* is nearby. The results we obtained for *nós*,<sup>10</sup> although not statistically significant, would seem to go in the opposite direction, favoring the presence of *-mos* when *nós* is near. This is exactly the situation we would expect, since *nós*, but not *a gente*, traditionally takes the *-mos* desinence.

In the combined group of our three youngest age groups, VARBRUL chose the tense variable in first place for *a gente*. This is the only circumstance under which tense reaches first place, indicating that it accounts for a greater proportion of the overall variation than saliency. Furthermore, the tense weights for the younger group are slightly more polarized than those for the older speakers. Nonetheless, saliency is still chosen in second place and continues to exhibit the same sort of typical values as in our other calculations.

The results for *nós* and *a gente* can be summarized as follows. For *nós*, the most important variable is phonic salience in all four age groups, but tense gains strength in each succeeding generation, reaching statistical significance beginning with our third age group, the Young People. In the case of *a gente*, the main variable is phonic salience, but only for the Older Adults. For the remaining three age groups, tense becomes the principal variable and exhibits well-polarized values. We conclude that there is a tendency in spoken Brazilian Portuguese to shift the locus of determination of the use of *-mos* from saliency to tense. In other words, in the old system the *-mos* desinence was semantically redundant, controlled basically by the principle of salience, in the environment of a 1st plural subject. In the new system, to the contrary, *-mos* is associated in its own right with a semantically important category (i.e., past tense) and is acquiring a role as the preterit morpheme.

Two factors may be cited as underlying this change. First, a form such as *falamos* 'we speak' or 'we spoke' is, at the present time, ambiguous in the standard language of Brazil, as there is no surface differentiation between the present and preterit forms. Second, as a necessary consequence of the effect of the salience variable, there is an apparent concentration of *-mos* in the preterit. Even though this concentration is, as we have already observed, merely a result of a biased distribution and is not statistically significant, it is nonetheless quite real in the linguistic output. The new system inverts the statistical significance of the two variables, ascribing status as the principal independent determining variable to tense and relegating salience to a secondary role. The basis for this change is

the circumstance that the preterit is marked on the surface unambiguously for all grammatical persons except the 1st plural, and that *-mos* is in fact more frequent in the preterit than in other tenses, even in the speech of the older speakers. We can postulate that this situation leads to a change in the analysis of *-mos* in the younger generation. Since there is no other element to mark the preterit in the 1st plural, and since *-mos* occurs frequently in this environment, *-mos* is apparently being reanalyzed as a marker of the preterit in this context. It is possible to foresee a future period in which *-mos* may come to be categorically preterit and *0* categorically non-preterit in the 1st person plural.

However matters may turn out in the future, no categorical change had occurred up to the moment our research was carried out. What happened was a shift in the patterns of distribution, with use of *-mos* decreasing in non-preterit environments and remaining more or less constant, at values reasonably close to 100%, in preterit environments. There has been no real grammatical change. At least for the meantime, the different generations use exactly the same surface forms but distribute them differently in the relevant contexts. The fundamental difference between the old and the new usage is a gradual redistribution of *-mos* in accord with new variable weights that favor the occurrence of *-mos* in preterit environments.

#### CONCLUSIONS

The results reviewed here show how a grammatical system can undergo restructuring. In the speech of the older speakers we have a typical case of confusion of desinences historically associated with each of the two forms used as semantically 1st plural subjects. At this stage, the controlling variable is salience, with *-mos* favored in salient environments. Due to the biased distribution of tense with respect to salience, the distribution of surface forms in the older speakers' sample exhibits the desinence *-mos* with considerably greater relative frequency in preterit than in non-preterit verb forms. On the other hand, no other morphological marker exists to distinguish the preterit from the present in the 1st person plural, despite the existence of such a morpheme in the rest of the paradigm. This, in turn, has led to a gradual change in the relative weights of the variable constraint groups that influence the appearance of *-mos*. Thus, while there has been no change in the inventory of morphological surface types or of the construction types in which they could appear, there has been a change in the distribution of tokens. In fact, this distributional change has progressed in such a way that one can reasonably speculate that the change will become categorical some time in the future, setting up *-mos* as an unambiguous preterit marker.

#### NOTES

1. For details, see Naro (1994, 1998). For more general considerations, see Sankoff (1988) and Guy (1998).
2. This was based upon the results of Naro and Lemle (1976), who showed that, in subject-verb agreement with 3rd person plural subjects, formal plural marking on the verb is used more frequently when the difference between the singular and plural forms is greater. For details, see Naro (1981).

3. Variation in use of the type *a gente fala* versus *a gente falamos* for 'we speak' also occurs in Portugal, where *a gente falamos* may actually be more frequent than in Brazil. The status of the type *nós fala* in Portugal needs to be investigated.
4. There is said to be a difference in the degree of openness of the thematic vowel of the first conjugation forms in Portugal, but this distinction does not exist in Brazil.
5. The transcriptions given here in square brackets are roughly phonetic, without regard to fine detail or variation. They are intended only as a guide to readers who are not familiar with Brazilian Portuguese.
6. The limit of five syllables is arbitrary; it was chosen to facilitate comparison with other research, such as Naro and Lemle (1976), in which this same limit had been used.
7. The notation "significance = 0.0" is used by the VARBRUL program to indicate a number smaller than the limit allowed by the forming.
8. In Table 3 we have combined the two older groups because their behavior was similar in the previous table.
9. We were obliged to combine levels 4 and 5 in the case of *nós* due to a lack of a sufficient number of data.
10. Not cited in this article.

## REFERENCE

- Braga, Maria Luiza. (1977). *A concordância de número no sintagma nominal no Triângulo Mineiro*. Master's thesis, Pontifícia Universidade Católica do Rio de Janeiro.
- Guy, Gregory R. (1998). VARBRUL: Análise avançada. *Cadernos de Tradução* 1:27–49. Originally published in K. Ferrara et al. (Eds.), *Linguistic change and contact* (Texas Linguistic Forum, vol. 30). Porto Alegre: Universidade Federal do Rio Grande do Sul.
- Naro, Anthony J. (1981). The social and structural dimensions of a syntactic change. *Language* 57:63–98.
- (1994). Modelos matemáticos. In Cecília Mollica (Ed.), *Introdução à Sociolinguística Variacionista*. Rio de Janeiro: Editora da Universidade Federal do Rio de Janeiro. 17–25.
- (1998). Quantitative models for the study of language use. *Language Design: Journal of Theoretical and Experimental Linguistics* 1:41–57.
- Naro, Anthony J., & Lemle, Miriam. (1976). Syntactic diffusion. *Papers from the Parasession on Diachronic Syntax*. Chicago: Chicago Linguistic Society. 221–240.
- Pintzuk, Susan. (1988). VARBRUL programs. Unpublished manuscript, University of Pennsylvania.
- Sankoff, David. (1988). Variable rules. In U. Ammon, N. Dittmar, & K. J. Mattheier (Eds.), *Sociolinguistics: An international handbook of the science of language and society*. Berlin: de Gruyter. 984–998.
- Scherre, Maria Marta Pereira. (1978). *A regra de concordância de número em português*. Master's thesis, Pontifícia Universidade Católica do Rio de Janeiro.