University of Pennsylvania Working Papers in Linguistics

Volume 21
Issue 2 *Selected Papers from New Ways of Analyzing Variation (NWAV) 43*

Article 18

10-1-2015

The Effect of Salience on Co-variation in Brazilian Portuguese

Livia Oushiro

Gregory R. Guy

The Effect of Salience on Co-variation in Brazilian Portuguese

Abstract
This paper analyzes cross-correlations among six variables of Brazilian Portuguese (the pronunciation of nasal /e/, coda r-retroflexion, coda r-deletion, NP agreement, 3rd person plural subject-verb agreement, and 1st person plural subject-verb agreement), with the objective of identifying constraints that promote the co-occurrence of sociolinguistic variants in individual speakers' speech. We focus on the perspective of structural cohesion, and show that co-variability is conditioned not only by structural similarities among dependent variables (such as agreement processes or coda weakening), but also by general linguistic constraints that operate across multiple variables, such as phonic salience (Naro 1981, Scherre 1988, Naro et al. 1999). Finally, we suggest that markedness may be a more general linguistic principle underlying co-variation.

The Effect of Salience on Co-variation in Brazilian Portuguese

Livia Oushiro and Gregory R. Guy*

1 Introduction

The analysis of co-variation is still a relatively unexplored area within language studies. Even though a number of sociolinguistic works have examined multiple variables in a given community, there haven't always been systematic analyses of how these variables may inter-relate to one another. The basic question is whether speakers who tend to employ variant *x* of variable A would also tend to employ variant *y* of variable B, or if variables are independently embedded in language and society; in other words, do specific variants tend to co-occur?

It might, of course, be expected that certain variables co-vary as a consequence of language-internal pressures such as for phonemic symmetry or paradigmatic regularization. Vowel chain shifts are a good example of how one variable phenomenon may affect other variables in the language system (e.g., Labov 1994). Furthermore, given the recurrent patterns of social stratification, stylistic variation, and changes in progress that are apparent in sociolinguistic research, we might also expect certain pairs of variables to co-vary, regardless of their structural relations. Sociolinguistic variables such as (-ING) (e.g., Trudgill 1974, Labov 2001a), coronal stop deletion (e.g., Wolfram 1969, Santa Ana 1991), and th-stopping (e.g., Labov 2006 [1996], Dubois and Horvath 1998, Newlin-Lukowicz 2013) have repeatedly been demonstrated to correlate with speakers' gender, ethnicity and/or social class in many different communities and in similar directions. Hence, it is reasonable to assume that they should be correlated with each other, but this assumption has rarely been put to test.

Strong evidence of the *structural* cohesion of sociolects was shown in Labov's (2006 [1966]) seminal work on New York City English, in which he demonstrates co-variability between certain pairs of phonetic variables. For instance, speakers who have a high index for (aeh)-raising, as in *bad*, also tend to have a high index for (oh)-raising, as in *law*, and conversely, those who have a low index for (aeh) also tend to have a low index for (oh). Labov concludes that the organization of the vowel space of New York City English follows Martinet's principle of functional economy, a tendency towards equal spacing of phonemic units.

A few subsequent works in this area have also looked into co-variation on the basis of *social* predictions. Notably, Guy (2013) analyzed four stable sociolinguistic variables which are sharply socially stratified in Rio de Janeiro Portuguese (coda –s deletion, denasalization, noun phrase plural agreement and 3rd person plural verb agreement); even though the social conditioning of these variables is very similar, in terms of gender, age and education, co-variation in this set is mostly found only between structurally related pairs, such as NP and VP number agreement. Tagliamonte and Waters (2011) investigated co-variation among changes in progress in Toronto English (quotatives, intensifiers, deontics, and stative possessives), with the hypothesis that innovators would share a repertoire of innovative variants in their speech. Their results, however, showed weak correlations in the leaders' speech, indicating that speakers who lead in one change do not necessarily lead in another. Thorburn (2014) looked at patterns of co-variation among four variables (verbal –s, intensifier *right*, voiced and voiceless interdental stopping) in the speech of 25 speakers of Inuit English in Canada, whose features are closely associated with the local dialect. Her results also show very weak lectal cohesion, as only one out of the six possible pairings among these variables is significantly correlated.

Thus studies of co-variation so far have shown that certain pairs of variables do co-vary in individual speakers' usage, but correlations are not as frequent or as strong as one might expect in terms of the prestige, innovative status or association with local identity that are indexed by the variables. It is still not clear which social and language-internal factors promote or constrain co-variation.

*This research was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2012/01930-9). The authors wish to thank the audience at NWAV43 for their useful comments and questions, and naturally assume full responsibility for any errors that may remain.

This paper analyzes cross-correlations among six variables of Brazilian Portuguese (henceforth BP): the pronunciation of nasal /e/, coda r-retroflexion, coda r-deletion, NP agreement, 3rd person plural subject-verb agreement, and 1st person plural subject-verb agreement, with the objective of teasing apart which constraints promote co-variation. We focus on the perspective of structural cohesion, and show that co-variability is conditioned not only by structural similarities among dependent variables (such as the agreement phenomena), but also by general linguistic constraints that correlate with multiple variables, such as phonic salience (Naro 1981, Scherre 1988, Naro et al. 1999). Finally, we suggest that co-variation is conditioned by a more general linguistic constraint, namely markedness, and set forth hypotheses to be tested in future work.

2 Data and Methods

The data for the present study are drawn from 118 sociolinguistic interviews with native speakers from the city of São Paulo, all from the SP2010 Project (Mendes and Oushiro 2013). The speaker sample is balanced for sex/gender, three age groups (20-34 y.o., 35-59 y.o., 60 or more y.o.), two levels of education (up to high school *vs*. college), and two areas of residence in the city (central *vs.* peripheral). The data set includes the six aforementioned phonological and syntactic variables, exemplified in (1–6), some pairs of which are structurally related and some that are not.

(1) Nasal /e/: [fa.zẽ.dɐ] vs. [fa.zẽj̃.dɐ] *fazenda* ‘farm’
(2) R-realization: [puɾ.ke] vs. [puɻ.ke] *porque* ‘because’
(3) R-deletion: [puɾ.ke]/[puɻ.ke] vs. [pu.ke] *porque*
(4) NP-agreement: *meu-s menino-s* vs. *meu-s menino-∅* ‘my boys’; *o-s caminh-ões* vs. *o-s caminhão* ‘the trucks’
(5) 3PP VP-agreement: *eles comem* vs. *eles come-∅* ‘they eat’; *eles fizeram* vs. *eles fez* ‘they did’
(6) 1PP VP-agreement: *nós falamos* vs. *nós fala-∅* ‘we speak’; *nós vamos* vs. *nós vai* ‘we go’

One of the phonological variables is the realization of stressed nasal /e/ (1) as a monophthong [ẽ] (the most common pronunciation in most dialects of Brazilian Portuguese) or as a diphthong [ẽj̃]. Diphthongal nasal /e/ is considered to be local to the city of São Paulo and is a stereotype of Paulistano Portuguese. The second phonological variable is the realization of coda (-r) (2), whose variants are prominent features of social and regional identities in Brazilian Portuguese (Callou et al. 1996). In São Paulo, coda (-r) is most frequently pronounced as a tap [ɾ] (considered prototypical of the city) or as a retroflex [ɻ] (commonly associated with rural speakers from nearby areas of the countryside). Coda (-r) also happens to be subject to deletion, so rates of (-r) deletion were also analyzed for this study (3). This variable is not specifically associated with Paulistano speech; unlike coda (-r) retroflexion, (-r) deletion occurs in all dialects of Brazilian Portuguese.

The three syntactic variables investigated here are also found widely throughout Brazil, with regular social stratification. They all involve variable number agreement, in cases where the standard prescriptive variety requires redundant plural markers. Thus in a plural NP (4), all words may potentially bear plural marking, but in vernacular speech, plural marking is often found only on the first word in a plural NP. Similarly, verbs in plural VPs are standardly expected to bear a plural morpheme agreeing with the grammatical person (5–6), but they can also be realized with third person singular morphology. In these analyses, we treat the first and third plural cases as separate variables.[1]

Each of the variables was first analyzed separately in mixed-effects models in R, each of which included Speaker and Lexical Item as random effects. These analyses also included five social variables (sex/gender, age, level of education, area of residence, and social class) and rele-

[1]Note that in this case, second person forms are not considered, as the historical second person verbs (e.g., *vós fazeis* ‘you-PL do’) are rarely used in Brazilian Portuguese. The vernacular 2PP *vocês* is also rarely employed by the speakers in sociolinguistic interviews.

vant linguistic predictors for each variable as fixed effects.[2] From these analyses, we obtained each speaker's tendency of use in factor weights of the nonstandard variants (the second in each pair in 1–6), which were then used to calculate cross-correlations among the six variables through Pearson correlation coefficients using Baayen's (2011) languageR package. Pearson's r-values range from –1, which indicates a perfect negative correlation (the more *x*, the less *y*, proportionately), to +1, which indicates a perfect positive correlation (the more *x*, the more *y*); an r-value close to 0 indicates non-correlation, meaning that knowing the value of one variable predicts nothing about the value of the other.

3 Predictions and Previous Results

In terms of structural relations, the syntactic variables are clearly related. 3PP and 1PP agreement both refer to number morphology in the verbal system, which tends to form tight paradigms. It has been posited that changes in verb morphology in BP has led to a parametric change from a null subject to a non-pro-drop language, where person and number are now increasingly marked by the pronoun rather than verbal inflection, similar to English and French (Duarte and Varejão 2013). If variable VP plural marking is related to fundamentally different underlying grammars governing the morphosyntactic expression of person/number features, one could expect considerable language-internal pressure for 3PP and 1PP symmetry. Nominal number agreement involves a similar process whereby plural number is either redundantly marked across the NP, or is non-redundantly marked only on the first word; hence this might also be expected to correlate well with 3PP and 1PP verb agreement.

As for the phonological variables, coda (-r) retroflexion and deletion may also be structurally related. Coda (-r) variation in BP has been claimed to be part of a long historical process of coda weakening, in which the segment has gradually become more posterior and ultimately deleted (Callou et al. 1996). As retroflex /r/ is a relatively more posterior segment than the apical tap, one could raise the hypothesis that speakers who tend to employ the retroflex would be one step "further" in the –r *continuum* and thus should also exhibit greater rates of deletion.

There is also a partial structural overlap between coda (-r) deletion and variable NP agreement. In many cases the superficial plural morpheme is a coda /s/ (as in *meninos*, *menino-*∅ 'boys', but not in *caminh-ão* [ãw̃], *caminh-ões* [õjs] 'trucks'), so coda (-r) deletion and possibly many tokens of zero NP plural markers share a superficial phonological rule of coda simplification of the form CVC → CV (we further address this issue in Section 4 below).

All other variable pairs are, in principle, structurally unrelated. Nasal /e/ is the only variable that refers to a diphthongization or a lengthening phenomenon more generally, and coda (-r) retroflexion bears no structural similarity to any of the syntactic variables.

On the other hand, in terms of social distribution, all six variables share very similar patterns. Table 1 summarizes the results from separate multivariate analyses of each sociolinguistic variable by showing the social predictor favoring the application variant, as well as the range (difference between the highest and lowest factor weights) as an indication of the relative importance of each predictor variable. All six variables show significant associations with speakers' sex/gender, with women favoring [ẽj̃] and men favoring R-retroflexion, R-deletion, and the zero plural markers for the three syntactic variables. All variables also show significant effects of the speakers' social class and level of education, with upper class and more educated speakers favoring [ẽj̃] and working class and less educated speakers favoring the others. It is also worth noticing that the hierarchy of constraints for R-retroflexion and the three syntactic variables is widely similar: social class as the most significantly correlated variable, followed by education and sex/gender, and no correlation with age, which generally displays stable variation in apparent time. Age in fact only shows a significant association with nasal /e/, with younger speakers favoring diphthongal realizations.

[2]For instance, preceding and following phonological context for (ẽ); syllable stress, syllable position and preceding vowel both for R-retroflexion and R-deletion; phonic salience and parallelism for NP, 3PP-VP, and 1PP-VP. Style (defined as "attention paid to speech," Labov 2001b) was also analyzed for each of the variables, but since neither of the syntactic variables exhibit variation in more monitored styles, subsequent analyses discarded reading passage tokens. For the purpose of comparison here, we only considered conversational data for the phonological variables as well. See Oushiro (2014, 2015) for a more detailed account.

Hence, from this set, [ẽj̃] is the odd-one-out; however, this does not mean that this variable should not correlate in individual speakers' usage with the other five variables; rather, one could expect negative correlations between [ẽj̃] and the others: the more *x*, the less *y*.

	[ẽj̃]	**[ɻ]**	**R-∅**	**NP-∅**	**3PP-∅**	**1PP-∅**
Sex/Gender	women (28)	men (17)	men (14)	men (18)	men (13)	men (62)
Age	younger (26)	--	--	--	--	--
Education	college (13)	high school (22)	high school (14)	high school (53)	high school (42)	high school (70)
Social Class	upper (23)	working (41)	working (13)	working (64)	working (50)	working (81)
Area of Residence	--	peripheral (36)	--	--	--	--

Table 1: Predictors favoring the nonstandard variants in separate multivariate analyses and respective ranges (00) for each social variable. --: non-correlation; [ẽj̃]: diphthongal nasal /e/; [ɻ]: R-retroflexion; R-∅: R-deletion; NP-∅: nonstandard NP agreement; 3PP-∅: nonstandard 3PP agreement; 1PP-∅: nonstandard 1PP agreement.

In short, from the perspective of social distribution, one would expect all six variables to co-vary, as they share very similar associations, but from the perspective of structural relations, not all pairs should co-vary. In a previous analysis (Oushiro and Guy 2013), the authors explored which of these pairs were correlated, and how strongly. Figure 1 is the correlation matrix containing all possible pairings among the six variables. The bottom left boxes display Pearson's r-coefficients and their respective significance values, and the top right boxes show the dispersion of speakers' factor weights for each pair of variables, along with a regression line.

This is a very dense and detailed display, but here we will focus on which of our predictions were born out, and which were not. In the figure, the gray-shaded boxes show significant correlations ($p < 0.05$). We predicted on social grounds that all pairs should be correlated, but this is clearly not the case; only 8 of 15 possible pairings reach significance. Let us consider, therefore, how the structural predictions fare. Diphthongal (ẽ) is not, in fact, correlated with any of the other variables, which is consistent with its absence of any structural connection. In the case of the two R-variables, however, no correlations are observed in the usage of individual speakers, despite the structural relationship between them; Pearson's r is very close to zero at 0.09 and the significance value is 0.33. On the other hand, the two coda (-r) variables were not expected on structural grounds to correlate with the syntactic variables, but in fact they do, although not very strongly, with values ranging from r=0.23 to r=0.37. Another correlation that doesn't exactly follow the predictions is the one between NP and 3PP agreement. This is the strongest correlation we find in this set with Pearson's r at 0.74; speakers cluster tightly along a neat regression line where both values increase together. While this is consistent with structurally-based predictions, it is unexpected that this correlation is even stronger than that between the two verb agreement variables (r=0.47).

Thus internal factors alone do not predict exactly which pairs of variables will co-vary nor the strength of such correlations. At the same time, purely social constraints also fail to account for the observed co-variation. In light of these results, we decided to run separate analyses in subsets of the data defined by different social and linguistic predictors. Pearson's r-coefficients were calculated by social groups (i.e., women *vs.* men, younger *vs.* older speakers, etc.) and by the principal linguistic predictors for each variable (e.g., for both coda (r) variables [+high] *vs.* [-high] preceding vowel and stressed *vs.* unstressed syllables; for NP, 3PP and 1PP agreement salient *vs.*

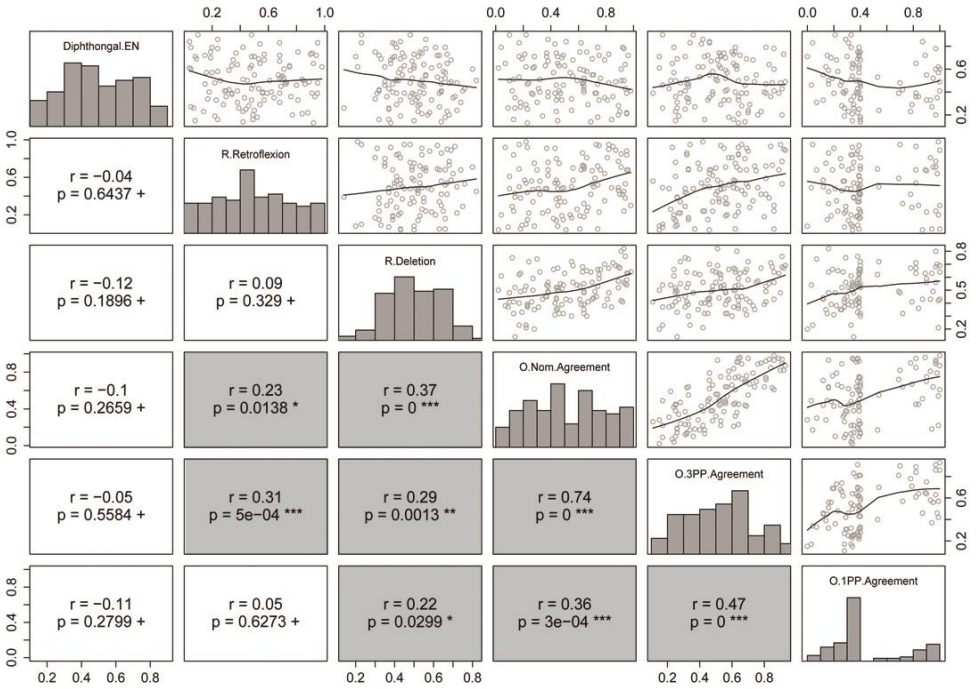

Fig. 1: Cross-correlations among six variables in BP (Oushiro and Guy 2013).

unsalient plural forms, etc.).[3] Differences between Pearson's r-values for different predictors were compared through the r.test function of the psych package (Revelle 2014). In the remainder of this paper we focus on the results for the linguistic predictors.[4]

4 Results and Discussion

In most cases, correlations between pairs of variables do not change by considering a specific subsample of tokens. For instance, preceding vowel height is one of the linguistic constraints on R-retroflexion. Table 2 shows the effect of subdividing the data into [-high] ([a, ɛ, ɔ]) and [+high] ([i, e, o, u]) preceding vowels. The second column shows the results for the overall analysis, as shown in Figure 1, as a reference for comparison with the third and fourth columns, which show the correlation coefficients when considering only low or only high preceding vowels. The last column is a *p*-value that indicates the probability that the difference between the data subsets is significant.

Pearson's coefficients for low and high vowels do not differ much: e.g., zero and -0.05 for the correlation with [ẽj̃]; 0.05 and 0.09 for the correlation with R-deletion; 0.19 and 0.22 with NP agreement, etc. In the overall analysis, R-retroflexion only showed significant correlations with NP and 3PP, and the same is true when we consider just the low or just the high preceding vowels, which results in non-significant differences between the low and high subsets in the last column.

However, there is one linguistic variable that does have significant effects on co-variation, namely phonic salience (Naro 1981, Naro et al. 1999, Scherre 1988). This variable has been extensively analyzed in BP studies of NP and VP agreement, and refers to how different the singular and plural forms are. Studies invariably show more agreement with more salient forms, and less agreement with forms of lesser salience (see, e.g., Guy 1981, Dias and Fernandes 2000, Rodrigues 1987, Rubio 2012, *inter alia*).

[3]These were done in R by running scripts of the same respective models for each data subset.
[4]See Oushiro (forthcoming) for an account of the social predictors.

	R-retroflexion			
	All tokens	**[-high]**	**[+high]**	**[-high] vs. [+high]**
	N = 5,900	N = 2,177	N = 3,723	*p*
[ẽj̃]	-0.04	0	-0.05	0.10
R-∅	0.09	0.05	0.09	0.25
NP-∅	0.23*	0.19*	0.22*	0.30
3PP-∅	0.31***	0.29**	0.29**	0.50
1PP-∅	0.05	0.01	0.09	0.11

Table 2: Comparison of Pearson's r-coefficients and respective significance values (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$) for the concurrent usage of R-retroflexion and [ẽj̃], R-∅, NP-∅, 3PP-∅ and 1PP-∅ when considering all tokens or data subsets defined by preceding vowel height.

In the case of nominal agreement, previous studies coded a range of saliency categories as illustrated in (7) (cf. Scherre 1988). For the present analysis, regular plurals (7a-7b), which involve only the addition of a morpheme *-s* (*casa*/*casa-s*, *mão*/*mão-s*), were recoded as less salient forms, and irregular plurals (7c-7h), which involve more phonic material in plural formation ([is, õj̃s, ãj̃s] and root changes), were recoded as more salient forms.

(7) Nominal plurals
  a. Regular plurals
    i. Oral vowel + s: *casa*, *casa-s* 'houses'
    ii. Nasal vowel + s: *mão*, *mão-s* 'hands'
  b. Irregular plurals
    i. -r + /is/: *professor*, *professor-es* 'teachers'
    ii. -l + /is/: *azul*, *azu-is* 'blue'
    iii. -S + /is/: *vez*, *vez-es* 'times'
    iv. -ão /õj̃s/: *profissão*, *profiss-ões* 'professions'
    v. -ão /ãj̃s/: *pão*, *pães* 'bread'
    vi. metaphonic: *ovo* [ovu], *ovos* [ɔvus] 'eggs'

3PP verbs had been initially coded according to Naro's (1981:75) salience scale shown in (8), but were divided for this analysis into Level 1 and Level 2 verbs. In Level 1 (the less salient category), the opposition between singular and plural forms is marked in an unstressed syllable and generally involves minor phonetic changes, mainly nasalization of the final vowel (e.g., *come*/*comem* 'he eats, they eat'), while in the more salient cases in Level 2, the singular-plural distinction involves a stressed affix and greater phonological differences (e.g., *disse*/*disseram* 'he said, they said').

(8) 3PP verb forms
  a. Level 1: unstressed opposition
    i. *come*/*comem* [-i/-ĩ] 'they eat'
    ii. *fala*/*falam* [-a/-ũ] 'they speak'
    iii. *faz*/*fazem* [-∅/-ĩ] 'they do'
  b. Level 2: stressed opposition
    i. *dá*/*dão* [-ˈa/-ˈãw̃] 'they give'
    ii. *comeu*/*comeram* [-ˈew/-ˈerũ, -ˈ iw/-ˈ irũ, -ˈoj/-ˈorũ] 'they ate'
    iii. *falou*/*falaram* [-ˈo/-ˈarũ] 'they spoke'
    iv. *é*/*são* (unique case) 'is/are'
    v. *disse*/*disseram* [-∅/-ˈɛrũ] 'they said'

1PP verbs had been similarly coded according to the salience scale proposed by Naro et al.

(1999:203), as shown in (9). Here we considered only the distinction between the less salient unstressed forms in Level 1 (as in *falava*/*falávamos* 'he/we used to speak'), and more salient types of forms with stressed inflections in Level 2 (e.g., *vai*/*vamos* 'he goes, we go', *falou*/*falamos* 'he/we spoke').

(9) 1PP verb forms
  a. Level 1: unstressed opposition
    i. *falava*/*faláva-mos* [-∅/-mʊs] 'we used to speak'
  b. Level 2: stressed opposition
    ii. *fala*/*fala-mos* [-a/-ˈa.mʊs] 'we speak' – stressed in one of the forms
    iii. *está*/*esta-mos* [-ˈa/-ˈa.mʊs] 'we are' – stressed in both forms
    iv. *vai*/*vamos* [-ˈaj/ˈa.mʊs] 'we go' – stressed in both forms, 3PS form shows an upglide
    v. *falou*/*falamos* [-ˈow/ˈa.mʊs] 'we spoke' – stressed in both forms, stressed vowel changes

When comparing less and more salient tokens of NP plural marking (Table 3), correlational strength changes only with R-deletion, which gets significantly weaker in the more salient forms (r=0.38*** vs. 0.22*, $p = 0.01$). Thus correlation between R-∅ and NP-∅ is driven mostly by less salient forms of NP, which are the regular plurals marked only by a suffixed *-s* (as in 7a–7b above). This suggests that co-variation between R-∅ and NP-∅ may also be motivated by a phonological rule of coda deletion affecting both coda (-r) and coda (-s) of CVC syllables (cf. *casas*/*casa*, *porque*/*poque*). Note, however, that the correlation between R-∅ and NP-∅, even though much weaker, is still significant when considering only more salient plurals ([-is, -õjs, -ãjs], and metaphonic); this cannot be straightforwardly explained by the same phonological rule, without postulating rule ordering or more abstract representations and derivations.

	NP-∅			
	All tokens	**[-salient]**	**[+salient]**	**[-salient] vs. [+salient]**
	N = 19,884	N = 16,460	N = 3,424	*p*
[ẽj̃]	-0.10***	-0.10	-0.12	0.39
[ɻ]	0.23*	0.21*	0.21*	0.50
R-∅	0.37***	0.38***	0.22*	0.01*
3PP-∅	0.74***	0.72***	0.65***	0.08
1PP-∅	0.36***	0.35***	0.31**	0.28

Table 3: Comparison of Pearson's r-coefficients and respective significance values (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$) for the concurrent usage of NP-∅ and [ẽj̃], [ɻ], R-∅, 3PP-∅ and 1PP-∅ when considering all tokens or data subsets defined by phonic salience.

In the case of 3PP-∅ (Table 4), the more salient forms again show significantly weaker correlation coefficients with R-∅ (r=0.34*** vs. 0.17, $p < 0.01$) and with NP-∅ (r=0.77*** vs. 0.55***, $p < 0.001$). Hence the stronger correlations are found with the unsalient 3PP forms, those that mainly involve marking an unstressed final vowel with nasality. This could be considered another case of coda simplification like deletion of final *-r* and *-s*, although at a more abstract level (postulating a plural marker in the form of an abstract [nasal] suffix).

Finally, a similar phenomenon is observed when comparing less and more salient forms of 1PP-∅ (Table 5). Correlation coefficients for the more salient forms become significantly weaker with R-∅ (r=0.33** vs. 0.02, $p < 0.03$), NP-∅ (r=0.50*** vs. 0.11, $p < 0.01$), and 3PP-∅ (r=0.54*** vs. 0.21*, $p < 0.01$). Here for the unsalient cases, the nonagreeing forms differ from the agreeing forms by truncation of a final syllable (-mos).

	3PP-∅			
	All tokens	**[-salient]**	**[+salient]**	**[-salient] vs. [+salient]**
	N = 9,480	N = 5,309	N = 4,171	*p*
[ẽj̃]	-0.05	-0.07	-0.05	0.39
[ɻ]	0.31***	0.28**	0.31***	0.33
R-∅	0.29**	0.34***	0.17	0.007**
NP-∅	0.74***	0.77***	0.55***	0.00***
1PP-∅	0.47***	0.44***	0.48***	0.26

Table 4: Comparison of Pearson's r-coefficients and respective significance values (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$) for the concurrent usage of 3PP-∅ and [ẽj̃], [ɻ], R-∅, NP-∅ and 1PP-∅ when considering all tokens or data subsets defined by phonic salience.

Altogether, the scales of phonic salience for the syntactic variables show that, in all cases, the less salient forms are the ones that have stronger correlations with other variables. Furthermore, there is an implicational order among the variables. The results show that the salience scale for NP only affects correlations with R-∅, the one for 3PP affects correlations with both NP and R-∅, and saliency in the 1PP case affects 3PP, NP, and R-∅. This may be explained by a more general salience scale involving all these variables. Each of them may be seen as involving some kind of reduction: R-deletion involves loss of a segment, and non-agreement in the syntactic variables involves omission or deletion of a suffix. But the degree of reduction differs across the cases. The phonic material marking agreement in the 1PP is an entire syllable (-mos); in the unsalient 3PP case it involves a vocalic nasal feature, which may trigger accompanying changes in the vowel quality ([-i/ĩ, -a/ãw̃]). Nominal plural marking involves the *-s* suffix, sometimes accompanied by root vowel changes ([-s, -õj̃s]), and R-deletion consists exclusively of the deletion of just one segment. Hence, the implicational ordering of the interactions we have observed depends on how big a difference there is between the full and reduced forms of a variable. If interactions occur where this difference is big (i.e., salient), it will occur with all less salient variables as well. One possible interpretation of this finding is that the correlations between variables are affected by the extension of some abstract process of reduction; this occurs straightforwardly when only segments are involved (*-r*, *-s*), but is harder to generalize to abstract affixes (3PP nasality), or to entire syllables. Hence the implicational relation among these variables: generalizing the final reduction process so far as to include the removal of a syllable (1PP) implies having also generalized it to the removal of segments and features, but this is not true in the reverse direction.

	1PP-∅			
	All tokens	**[-salient]**	**[+salient]**	**[-salient] vs. [+salient]**
	N = 1,074	N = 297	N = 777	*p*
[ẽj̃]	-0.11	-0.19	0.07	0.051
[ɻ]	0.05	0.10	-0.13	0.076
R-∅	0.22*	0.33**	0.02	0.023*
NP-∅	0.36***	0.50***	0.11	0.004**
3PP-∅	0.47***	0.54***	0.21*	0.009**

Table 5: Comparison of Pearson's r-coefficients and respective significance values (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$) for the concurrent usage of 1PP-∅ and [ẽj̃], [ɻ], R-∅, NP-∅ and 3PP-∅ when considering all tokens or data subsets defined by phonic salience.

5 Final Remarks

We have discussed cross-correlations among six variables of Brazilian Portuguese. Neither structural nor social constraints alone give an adequate account of the observed co-variation, or lack thereof. On the social level, this suggests that the indexicality of linguistic variables evoked by their general social distribution is not necessarily strongly predictive of their relative rates of use by individual speakers; this is consistent with other recent work on co-variation (e.g., Guy 2013). But at the level of linguistic structure, we propose in this paper a more detailed analysis that subdivides the dataset by different predictors. From a language-internal perspective, we show that co-variation is promoted not only by structural similarity between dependent variables (for instance, 3PP and 1PP verb agreement) but also by more general linguistic constraints that operate across multiple variables, such as phonic salience. Less salient forms tend to co-vary more frequently and more strongly with other variants, and the salience scale affects co-variation with all other less salient forms.

It may be the case that there is an even more general principle operating here, which is that lectal cohesion is conditioned not only by the less salient forms, but unmarked forms more generally. Perhaps these are the forms that are subject to more automatic and consistent treatment in the grammars of individuals, while more salient forms are more readily available for manipulation for stylistic or identity-performing purposes. So future analyses will need to look more carefully into the effect of other variables such as subject position, animacy, and parallelism/priming, which correlate with multiple sociolinguistic variables in BP and other languages and may shed some light into the hypothesis of markedness.

References

Baayen, Harald. 2011. LanguageR. R package version 1.4.1.

Callou, Dinah, João Moraes, and Yonne Leite. 1996. Variação e diferenciação dialetal: a pronúncia do /r/ no português do Brasil. In *Gramática do Português Falado Volume 6*, ed. I. Koch, 465–493. Campinas: UNICAMP.

Dias, Juçá F. V., and Marisa Fernandes. 2000. A inter-relação da concordância nominal e da concordância nos predicativos/particípios passivos, sob enfoque da teoria da variação e mudança linguística. *Organon* 14:115–131.

Duarte, Maria Eugênia Lamoglia, and Filomena Varejão. 2013. Null subjects and agreement marks in European and Brazilian Portuguese. *Journal of Portuguese Linguistics* 12:101–124.

Dubois, Sylvie, and Barbara M. Horvath. 1998. Let's tink about dat: interdental fricatives in Cajun English. *Language Variation and Change* 10:245–261.

Guy, Gregory R. 1981. *Linguistic Variation in Brazilian Portuguese: Aspects of the Phonology, Syntax, and Language History*. Ann Arbor: University Microfilms International.

Guy, Gregory R. 2013. The cognitive coherence of sociolects: How do speakers handle multiple sociolinguistic variables? *Journal of Pragmatics* 52:63–71.

Labov, William. 1994. *Principles of Linguistic Change, Vol. I, Internal Factors*. Malden: Wiley-Blackwell.

Labov, William. 2001a. *Principles of Linguistic Change, Vol. II, Social Factors*. Malden: Blackwell.

Labov, William. 2001b. The anatomy of style-shifting. In *Style and Sociolinguistic Variation*, ed. P. Eckert and J. Rickford, 85–108. Cambridge: Cambridge University Press.

Labov, William. 2006 [1966]. *The Social Stratification of English in New York City*. New York: Cambridge University Press.

Mendes, Ronald Beline, and Livia Oushiro. 2013. Documentação do Projeto SP2010: Construção de uma amostra da fala paulistana. URL http://projetosp2010.fflch.usp.br/producao-bibliografica.

Naro, Anthony J. 1981. The social and structural dimensions of a syntactic change. *Language* 57:63–98.

Naro, Anthony J., Edair Görski, and Eulália Fernandes. 1999. Change without change. *Language Variation and Change* 11:197–211.

Newlin-Łukowicz, Luiza. 2013. TH-stopping in New York City: substrate effect turned ethnic marker? In *U. Penn Working Papers in Linguistics 19.2: Selected papers from NWAV 41*, ed. A. Freeman, 151–160.

Oushiro, Livia. 2014. O apagamento de (-r) em coda nos limites da variação. *Veredas* 18:251–266.

Oushiro, Livia. 2015. Identidade na Pluralidade: Avaliação, Produção e Percepção Linguística na Cidade de São Paulo. Doctoral Dissertation, Universidade de São Paulo, São Paulo.

Oushiro, Livia. Forthcoming. Social and structural constraints in lectal cohesion. *Lingua*, ed. G.R. Guy and F. Hinskens.
Oushiro, Livia; Guy, Gregory R. 2013. The social clustering of linguistic variables in São Paulo Portuguese. Paper presented at *New Ways of Analyzing Variation (NWAV) 42*, Pittsburgh-PA.
Revelle, William. 2014. Package 'psych' v.1.4.5. Procedures for Psychological, Psychometric, and Personality Research. URL http://personality-project.org/r/psych.manual.pdf.
Rodrigues, Angela Cecilia de Souza. 1987. A Concordância Verbal no Português Popular em São Paulo. Doctoral Dissertation, Universidade de São Paulo, São Paulo.
Rubio, Cássio F. 2012. Padrões de Concordância Verbal e de Alternância Pronominal no Português Brasileiro e Europeu: Estudo Sociolinguístico Comparativo. Doctoral Dissertation, UNESP, São José do Rio Preto.
Santa Ana, Otto. 1991. Phonetic Simplification Processes in the English of the Barrio: A Cross-Generational Study of the Chicanos of Los Angeles. Doctoral Dissertation, University of Pennsylvania, Philadelphia.
Scherre, Maria Marta Pereira. 1988. Reanálise da Concordância Nominal em Português. Doctoral Dissertation, Universidade Federal do Rio de Janeiro, Rio de Janeiro.
Tagliamonte, Sali, and Cathleen Waters. 2010. Innovators across innovations: Exploring co-variables in linguistic change. Paper presented at *New Ways of Analyzing Variation (NWAV) 39*, San Antonio,TX.
Trudgill, Peter. 1974. *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.
Thorburn, Jennifer. 2014. Dialect Development in Nain, Nunatsiavut: Emerging English in a Canadian Aboriginal Community. Doctoral Dissertation, Memorial University, St. John's (Newfoundland).
Wolfram, Walt. 1969. A Sociolinguistic Description of Detroit Negro Speech. Washington DC: Center for Applied Linguistics.

Livia Oushiro
Departamento de Linguística
Universidade de São Paulo
Av. Prof. Luciano Gualberto, 403
Cidade Universitária – São Paulo – SP 05508–900
livia.oushiro@usp.br

Gregory R. Guy
Department of Linguistics
New York University
10 Washington Place
New York, NY 10003
gregory.guy@nyu.edu