

# CONTRACTION, GRAMMATICALIZATION AND THE PHONOLOGY/SYNTAX INTERFACE: PROGRESSIVES IN BRAZILIAN PORTUGUESE

Ronald Beline Mendes  
University of São Paulo

Using recently collected data, this paper presents a new analysis of the process of grammaticalization of the periphrasis referred to as *estar + gerúndio*, in Brazilian Portuguese. This periphrasis is morphologically equivalent to the English progressive *be + gerund*, but its usage is somewhat different. Aside from semantic differences, there is an interesting morph-syntactic distinction in the comparison between the Brazilian *estar + gerúndio* and the English *be + gerund*, in that the contracted auxiliary verb may be pronounced as a clitic not to the verb subject, but to the gerund. The main goal of this paper is to show that such a cliticization (Heine 1993, Heine et al. 1991) is variable. It may or may not occur, when the auxiliary *estar* is pronounced in its phonetically reduced version. Furthermore, there is a correlation between the phonetic reduction of the auxiliary and the insertion of other syntactic elements (such as adverbials) between *estar* and the gerund.

## 1. Background and Corpus

Although the focus of this paper is on facts of the phonology/syntax interface, it is also useful to mention the semantic distinctions between the Brazilian Portuguese *estar + gerúndio* and the English *be + gerund*. The latter is essentially employed for the expression of progressive aspect (referring to events that are in progress at the time of the sentence utterance). In Brazilian Portuguese, aside from expressing progressive aspect, the former is also used to express iterative aspect (Ilari 2000, Ilari et al. 1983) and durative aspect (Mendes 2005). In the literature on aspect in Brazilian Portuguese, the iterative is also sometimes referred to as repetitive or habitual, and is understood as an intermittent repetition of an event or process through time – an aspectual concept that, in English, is prototypically conveyed by the *present perfect progressive*, as in *They have been cooking a lot lately*. As for the durative, it is an aspect similar to the progressive, with the difference that here the event is not necessarily in progress at the moment the sentence is being uttered. Therefore, the Brazilian Portuguese equivalent of the sentence *They are travelling* is not necessarily interpreted as something that is occurring *now*, at the moment the sentence is uttered; it can be interpreted as something that didn't use to be the case, something that has been occurring *lately*, differently from longer ago. These comments are important here for two reasons. The first is terminological in nature: the title of this paper calls the Brazilian construction *progressive*

because in English that term immediately brings up its morphological equivalent *be + gerund*; however, it is necessary to make the observation that this is not the only aspect that it expresses, as we have briefly seen above. The second reason is methodological, to the extent that *estar + gerund* is very productive and frequent in Brazilian Portuguese, which explains the number of tokens that have been analyzed for this paper.

The analysis presented here is said to be new because in Mendes 1999 I pursued a similar project, but with an older and much smaller corpus. For this paper, aside from a considerably larger corpus, the quantitative analysis performed is more detailed. Mendes 1999 quantitatively analyzed the tokens of *estar + gerúndio* found in 5 sociolinguistic interviews recorded in São Paulo, Brazil, in the 1970's by members of the project *Norma Urbana Culta*, or *Cultivated Urban Norm* (Preti et al. 1987) – a project that recorded and transcribed interviews in five capitals of Brazil for the purpose of linguistic analysis.

Following standard Labovian methodology for the study of variation, Mendes 1999 focused specifically on the variable pronunciation of the auxiliary *estar*, which can occur in its phonetically full or contracted form. For example, in the simple present, the full forms are *estou* and *está*, in the singular, and *estamos* and *estão*, in the plural. The corresponding contracted forms are *tô*, *tá*, *tamo* and *tão*, respectively. This example lists the forms for the present tense, but the phonetic contraction occurs in all tenses and moods.

In Mendes 1999, I analyzed the full and contracted forms of the auxiliary *estar* as variants of a variable. The basic questions were: given the alternative forms, what linguistic and social factors favor the reduced pronunciation of the auxiliary? Which factors were possibly inhibiting the use of the reduced form in Brazilian Portuguese, as spoken in São Paulo?

The contracted form of the auxiliary proved to be the preferred variant in the vast majority of contexts. In 184 tokens, 69% were represented by *tá + gerúndio* with the reduced auxiliary. Aside from this overall distribution, I observed a correlation between the use of the phonetically contracted auxiliary and age, the verb tense and the occurrence of other elements between the auxiliary and the main verb. The reduced form was favored among younger speakers and in the most frequent verb tenses (the simple present and the imperfect preterit). On the other hand, the contracted form was less frequent when there were other syntactic elements between the auxiliary and the gerund.

This present paper focuses more specifically on that last factor group. Almost a decade after Mendes 1999, I return to *estar + gerund*, in order to analyze the effect of elements intervening between the auxiliary and the gerund on the occurrence of the contracted form of *estar*. My main question concerns the grammaticalization of the periphrastic construction as a unit, that is, the process of coalescence (Hopper et al. 1993) between the contracted form of the auxiliary and the gerund.

As mentioned before, the *corpus* for this new analysis is much larger and is composed of 30 sociolinguistic interviews recorded by students of Variationist Sociolinguistics at the Department of Linguistics of the University of São Paulo,

between 2005 and 2007. From these interviews, I extracted 817 tokens, of which 78% feature the contracted forms of the auxiliary – 10 percentage points more frequent than in the previous work.

### 1. Qualitative Analysis – Developing the Hypothesis

With such a large amount of tokens, I have actually been able to look into the correlation between the phonetic reduction of the auxiliary and the bondedness between the auxiliary and the gerund by a reliable quantitative method. In contrast to Mendes 1999, where tokens were not so numerous, I describe and explain that correlation in the form of relative weights (with VARBRUL, version Goldvarb X), and not just percentages of occurrence of a form in the set of tokens.

The basic premise here is that the more grammaticalized a periphrastic construction becomes, the more cohesion there is between its components. Therefore, the hypothesis is that the auxiliary *tá* (in its different forms, according to the tense and mood) may turn into an affix in relation to the gerund. In fact, it is in that sense that I'm not simply looking into the phonetic contraction of *estar*, but rather into the grammaticalization of the whole periphrastic compound.

Therefore, I'm proposing that the grammaticalization of *estar* + *gerund* is an ongoing process similar to the formation of future forms in Portuguese and other romance languages. In the history of Portuguese, the Vulgar Latin auxiliary *habere* may have gone through a process of phonetic reduction and semantic bleaching, becoming a suffix to the infinitival form that it followed (Heine 1993, Hopper et al. 1993).

- (1) a. Amare        habeo  
       To love-inf. have-pres.1p.s.  
       'I will love'
- b. Amare        hei  
       To love-inf. have-contractedAUX  
       'I will love'
- c. Amarei  
       Love-future1p.s.  
       'I will love'

In the sequence of examples in (1) above, adapted from Castilho 1995, the meaning of the sentences is the same, before and after the affixation of the contracted form of the auxiliary (*habeo*>*hei*). However, the suffix *-ei* in (1c) is no longer interpreted as an auxiliary, and it can't be separated by other syntactic constituents, such as adverbials, from the verb stem – *amar-*.

Returning to the grammaticalization of *tá* + *gerúndio*, I present the possibility that the phonetically reduced auxiliary *tá* may function as a clitic,

which could ultimately become a prefix. In such a case, we would expect a progressively weaker tendency to insert other syntactic constituents between the reduced auxiliary and the gerund.

The variationist perspective recognizes that language use is not necessarily categorical. In examples (2) and (3), we see tokens where the auxiliary and the gerund were not separated by the underlined constituents, and the auxiliary is contracted. These examples can be considered as evidence in favor of the hypothesis described above. In (4), the auxiliary is in its plain phonetic form, and there are adverbials between the auxiliary and the gerund.

- (2) *cê tá indo sempre pra frente, né?*  
 you be-1p.s.-aux going always to-the front tag  
 ‘You’re always going ahead, aren’t you?’
- (3) *quem tá precisando realmente de um pedaço de chão?*  
 who be-3p.s.-aux needing really of a piece of ground  
 (lit.) ‘Who’s really needing a piece of land?’
- (4) *eu estou já faz dezoito anos trabalhando nessa área*  
 I be-1p.s.-aux already for 18 years working in-this area  
 ‘I’ve been working in this area for eighteen years now.’

These tokens are presented here as they were found in the interviews. However, the corresponding examples (5) to (7) are also possible in spoken Brazilian Portuguese. This has to be noted, even though they potentially constitute a counterargument to my hypothesis. Examples (5) and (6) show us that the contracted auxiliary does not behave, categorically, as a clitic or prefix that can’t be detached from the gerund – at least not yet. On the other hand, (7) illustrates that, when the auxiliary is plainly pronounced, it’s grammatically acceptable to put the adverbials to the right of the whole periphrasis, and not necessarily between *estar* and the gerund.

- (5) *cê tá sempre indo pra frente, né?*  
 you be-1p.s.-aux always going to-the front tag  
 ‘You’re always going ahead, aren’t you?’
- (6) *quem tá realmente precisando de um pedaço de chão?*  
 who be-3p.s.-aux really needing of a piece of ground  
 (lit.) ‘Who’s really needing a piece of land?’
- (7) *eu estou trabalhando já faz dezoito anos nessa área*  
 I be-1p.s.-aux working already for 18 years in-this area  
 ‘I’ve been working in this area for eighteen years now.’

Therefore, there is variation among all forms, which leads to the need to check if there is at least a tendency in spoken Portuguese to prefer sentences like

(2) to (4) and to avoid sentences like (5) to (7). In the framework of variationist sociolinguistics, the most efficient way to approach facts such as these is to pursue a multivariate analysis of the data. All the results presented in the following section were obtained with a multivariate analysis in which were included both social factors – such as sex/gender and age group – and linguistic factor – such as verb tense, verbal person and the nature of the syntactic constituents that occurred or could have occurred between the auxiliary *estar* and the gerund. Here, I will only report the results on the linguistic factor group named “intervening constituents” and the correlation with age group.

## 2. Quantitative Analysis

For the hypothesis of the correlation between the phonetic contraction of *estar* and the tendency to avoid inserting other syntactic constituents between the auxiliary and the gerund, the most important factor group in this analysis is the one that looks into the type of constituents. Initially, the data were coded according to the factors below:

- (8) No intervening constituents  
 na verdade eu to fazendo toda essa explicação  
 in-the truth I be-1ps doing all that explanation  
 ‘Actually, I’m doing all the explanation’.
- (9) Clitic pronouns  
 igual eu tô te falando, ele era meu superior  
 equal I am you-OBJ speaking he was my superior  
 ‘As I’m telling you, he was my superior’.
- (10) Adverbs  
 ele tava já chamando a polícia  
 he was already calling the police  
 ‘He was already calling the police’.
- (11) Adverbs in *-mente* (*-ly*)  
 a gente tá facilmente sendo manipulado pela mídia  
 1p.pl. PRO is easily being manipulated by-the media  
 (lit.) ‘We’re being easily manipulated by the media’.
- (12) Locatives  
 eu estava no ponto esperando ele  
 I be-past 1ps at-the point waiting he-ACC  
 ‘I was waiting for him at the bust stop’.
- (13) Others  
 a gente tava né conversando  
 1p.pl. PRO be-1p.s.past tag conversing

‘We we were talking, right?’

- (14) Others  
 tá todo mundo sabendo  
 be-1p.s.pres. all world knowing  
 (lit.) Everybody’s knowing it.

In the coding for this factor group, adverbs like *já* (*already*) and *facilmente* (*easily*) were initially analyzed separately because the latter are phonetically heavier than the former. In fact, adverbials like *já* can be pronounced without stress (as if they were unstressed clitics), which is certainly not the case of polysyllabic adverbials formed with the suffix *-mente*. As for the locatives, one might argue that they are adverbial adjuncts as well, and therefore they could have been analyzed as adverbials, as in (11). However, locatives are especially interesting to the testing of the hypothesis, since they relate to older, locative uses of the verb *estar*. In other words, it was interesting to at least observe if the use of the contracted form of the auxiliary would present some special correlation to locatives, considering their semantic relation to one of the basic functions of the verb *estar* in Portuguese. All other cases of intervening constituents, such as tags and discourse markers (13), and verb subjects (14) were coded as *others*.

Table 1 – Distribution of tokens *estar* + *gerund*

	contracted aux		plain aux		TOTAL	
	N	%	N	%	N	%
no interven. constituent	554	78,1	155	21,9	709	86,8
clitics	33	86,8	5	13,2	41	4,7
adverbials	9	69,2	4	30,8	13	1,6
adverbials in <i>-mente</i>	2	100	-	-	2	0,2
locatives	9	69,2	4	30,8	13	1,6
others	26	61,9	16	38,1	42	5,1
TOTAL	633	77,5	184	22,5	817	

Table 1 shows that approximately 87% of the data are represented by tokens in which there is no constituent placed between the auxiliary and the gerund. This means that, out of 817 tokens, 108 featured other constituents between the auxiliary *estar* and the gerund. The distribution of these tokens over five different categories of intervening constituents makes the cells relatively small – a common fact in works on grammatical variation. For a further analysis of the data, in order to test the correlation between the phonetic contraction of *estar* and its coalescence with the gerund, the data had to be recoded, as shown in table 2. All tokens in which there is an intervening constituent between the auxiliary and the gerund were regrouped in one category, except for clitic pronouns.

	contracted auxiliary			<i>plain aux</i>		TOTAL	
	N	%	weight	N	%	N	%
clitics	33	86,8	0.612	5	13,2	41	4,7
no intervening constituents	554	78,1	0.509	155	21,9	709	86,8
with intervening constituents	46	65,7	0.352	24	34,3	70	8,6
TOTAL	633	77,5		184	22,5	817	

Tokens with intervening clitics were kept separate from the ones with other intervening constituents because when it comes to periphrastic constructions and the placement of clitics in spoken Brazilian Portuguese, these occurrences don't represent a counter-argument to the hypothesis. When the post-verbal auxiliary *habere* got reduced and became a suffix in Portuguese, clitic pronouns could still be placed between the former auxiliary verb and the main verb (nowadays, a verb stem) (Castilho 1995). So, in the future tense forms, breaking the verb morphology to insert the clitic pronoun between the tense morpheme and the verb stem is not only still possible, but also considered grammatically correct, especially in written Portuguese, despite the fact that it is rather infrequent in spoken Brazilian Portuguese.

- (1) a. Amá-lo-ei.  
 Ama lo ei  
 Love-stem you-ACC FUT-1<sup>st</sup>.p.sing.  
 'I will love you'.
- b. Fá-lo-ei.  
 Fa lo ei  
 Do-stem it-ACC FUT-1<sup>st</sup>.p.sing.  
 'I will do it'.

Considering that, table 2 presents very interesting results. First, the low relative weight on the third line of the table means that placing morphologically heavy syntactic constituents between *estar* and the gerund is strongly disfavored when the auxiliary is phonetically reduced. Aside from that, by keeping the clitics separate from the other intervening elements it's possible to realize that they do behave differently; the higher relative weight in the table shows that speaker do not tend to avoid inserting clitic pronouns between *estar* and the gerund, even when the auxiliary is pronounced in its phonetically contracted version.

## 2.1 The Correlation with Age

In Mendes 1999, I observed that the use of *tá + gerund*, that is, the usage of the contracted form of the auxiliary in the periphrastic construction, possibly

constituted a case of change in progress. As noted earlier in the present discussion, that previous work focused specifically on the contraction of the auxiliary.

In the present paper, I have also noted that the frequency of the contracted form (78%) is even higher now than in the earlier work (69%). It's worth recalling that the data discussed in this paper were obtained from sociolinguistic interviews recently recorded, whereas Mendes 1999 analyzed data recorded in the 1970's. This may indicate that the contracted form is becoming even more frequent.

In order to refine the analysis, I proceeded to a crosstabulation of the linguistic factor group analyzed above and the social factor group "informants' age". Figure 1 below shows the results of the crosstabulation for the tokens where there is an intervening constituent between *estar* and the gerund, without computing those where the intervening constituent is a clitic pronoun. In other words, this graph illustrates the correlation with the speakers' age for the tokens on the third line of table 2 (a total of 70 tokens).

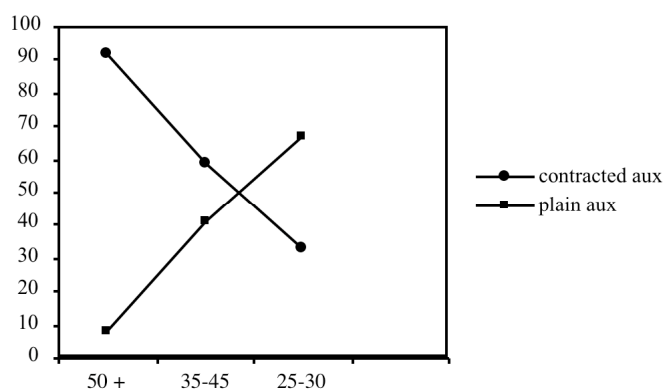


Figure 1 – Correlation between age groups and the insertion of other constituents between the auxiliary and the gerund

This graph clearly indicates a change in progress, in apparent time (Labov 2001). The insertion of other syntactic constituents between the auxiliary and the gerund is becoming less likely, in those cases where the auxiliary is phonetically contracted. In other words, these results show that it's more likely to find occurrences of *estar* (plain) + *X* + *gerund* than of *tá* (contracted) + *X* + *gerund*, in the speech of the youngest speakers. The results are inverted for the oldest speakers, among whom the occurrence of intervening syntactic constituents between *tá* and the gerund is more frequent, in spite of the auxiliary contraction.



### 3. Conclusion

The analysis presented here confirms the hypothesis on the correlation between the phonetic reduction of *estar* and the placement of other syntactic constituents between the auxiliary and the gerund. More specifically, it presents clear evidence supporting a process of change through which the contracted verb *estar*>*tá* may be reanalyzed as an affix to the gerund.

In future work, another interesting question could be addressed: are certain syntactic structures avoided *because* of phonological processes? Such a question is theoretical in nature, and the answer to it would constitute an important contribution to the understanding of grammaticalization. Moreover, by paraphrasing that question in methodological terms, it becomes of great interest to variationist sociolinguistics: is the relationship between variable and factor groups unidirectional? Such a question could be addressed by the comparison between the analysis developed in this paper and another, in which the intervening constituents are analyzed as a variable, whereas the different phonetic forms of the auxiliary would constitute a factor group.

### References

- Castilho, A. T. de. 1995. A Gramaticalização. *Estudos Lingüísticos e Literários* 19: 25-64.
- Heine, B., Ulrike Claudi and Friederike Hünemeyer. 1991. *Grammaticalization – A Conceptual Framework*. Chicago: The University of Chicago Press.
- Heine, Bernd. 1993. *Auxiliaries: Cognitive Forces and Grammaticalization*. Oxford: Oxford University Press.
- Hopper, P. and E. Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.
- Ilari, R. 2000. Notas para uma semântica do passado composto em português. *Actas do Congresso Internacional* organizado por motivo dos 20 anos do português no ensino superior. Departamento de Língua e Literatura Portuguesa da Faculdade de letras da Universidade Eötvös Lóránd, Budapeste. pp.224-247.
- Ilari, R. and I. Mantoanelli. 1983. As formas progressivas do português. *Cadernos de Estudos Lingüísticos* 5, pp.27-60. UNICAMP.
- Labov, W. 2000. *Principles of Linguistic Change. Vol II: Social Factors*. Cambridge: Blackwell.
- Mendes, R. B. 1999. A Gramaticalização de Estar + Gerúndio no Português Falado. MA. Thesis. Campinas: UNICAMP.
- Mendes, R. B. 2005. Estar + Gerúndio e Ter + Particípio – Aspecto Verbal e Variação no Português. Ph.D. Dissertation. Campinas: UNICAMP.
- Preti, D. e H. Urbano (orgs.) 1987. *A Linguagem Falada Culta na Cidade de São Paulo. Vol. III Entrevista*. São Paulo: TAQueiroz.