

CLITICS IN BRAZILIAN PORTUGUESE: RECOVERY THROUGH SCHOOLING

Vilma Reche-Corrêa

Unicamp/UnB, Brazil

Abstract: The structural differences between written and spoken languages in Brazil are so striking that some grammatical constructions must be learned at school. Such is the case of the third-person clitics, lost in spoken language in the last century. This variacionist analysis will show that the substitution of the vernacular forms, the lexical or the null pronouns, by this particle, in a variable context, is due not only to careful speech and a higher degree of learning but also to grammatical factors.

Keywords: sociolinguistics, dialectology, grammar, standard language, clitics.

1. DIRECT OBJECT STRATEGIES IN BRAZILIAN PORTUGUESE

The structural differences between written and spoken language in Brazil are so deep that some grammatical constructions must be learned at school. Such is the case of the clitic, as in (1), lost in the spoken language in the last century.

- (1) Eu vi o menino, mas Maria não o viu.
 I saw the boy but Mary not *him* saw
 'I saw the boy, but Mary didn't see him.'

Brazilians do not make use of the third person clitic, especially illiterate people and children. It is shown to be absent before schooling and to increase through reading and teaching. It is, however, the only acceptable anaphoric direct object in written language.

This study, based on Corrêa (1991)'s data, will present the results of a quantitative analysis of the recovery of the third-person clitic through school. In BP (Brazilian Portuguese) usual forms for the anaphoric direct object are lexical pronouns, noun phrases and ellipsis, as in (2):

- (2) a. Eu vi o menino, mas Maria não viu *ele*.
 I saw the boy but Mary not saw *he*

- b. Eu vi o menino, mas Maria não viu o menino.
 I saw the boy but Mary not saw *the boy*
- c. Eu vi o menino, mas Maria não viu [ø].
 I saw the boy but Mary not saw [ø]
 'I saw the boy, but Mary didn't see him.'

The clitic was shown to be the usual form of anaphoric direct object in the past but now the situation is quite different. So, we can hypothesize that formal education is the agent which preserves the archaic system.

Given this, what we want to know is to what extent the educational system successfully recovers the third person clitic, making the learner avoid both the strong pronoun and the null clitic.

The analysis of students' data will show that the substitution of other direct object forms for clitics in a variable context, in the way this takes place, is due to social and linguistic factors.

2. LEARNING THE CLITICS AT SCHOOL

With these goals in mind, the data were collected through recordings and written narratives produced by students of the elementary school and their teachers. Illiterate informants also had their narratives recorded to complete the sample. Play acting, simulating a thief in action in a lunchroom, was used as an eliciting strategy. The basic task was to determine what forms the informants would use in object position as anaphoric items: clitics, strong pronouns, null objects or full NPs (repetitions, epithets, quantifiers).

The results on Tables 1 and 2, adapted from Corrêa (1991), were obtained in the oral and written tasks.

Table 1. Distribution (%) of anaphoric direct objects for level of education in oral narratives

Type	Illiterate	1st/2nd	3rd/4th	5th/6th	7th/8th	Educated
null	68	73	76	72	72	65
lexical pronoun	26	24	9	18	21	7
NP	5	3	15	8	7	14
clitic	-	-	-	2	1	14

Table 2. Distribution (%) of anaphoric direct objects for level of education in written texts

Type	1st/2nd	3rd/4th	5th/6th	7th/8th	Educated
null	59	64	48	52	10
lexical pronoun	8	6	14	11	-
NP	33	18	13	5	5
clitic	-	12	25	32	86

In these tables, two points are striking: before schooling, the strong pronoun, the definite descriptions and chiefly the null object are the anaphoric resources available while in the educated speakers' written texts, the clitics are the almost exclusive anaphoric direct objects. The clitic, however, appears only and rarely in the last grades in the students' speech.

Comparing, now in more detail, the results of both Tables, we find reverse results. Null objects in educated informants' speech reach 65% and also constitute the major anaphoric resource for all the schooling stages. But it is drastically reduced in the written texts at this same school level (10%). The clitic, on the other hand, although highly incident in the educated speakers' written narratives (86%), is practically absent in students' speech (1% at the end of the elementary school) and only gradually acquired in their written production. This means that it is only at the pre-university level that it starts to develop, as a necessary weapon to face the entrance exams. Schooling strongly affects the use of the lexical pronoun in oral production, falling from 24% in the 1st grade to 7% in the university level. It is entirely banned from the written production of university students and clearly avoided even by less literate subjects (see percentages under 14%).

Observing the distribution of anaphoric direct objects through time, in Kato, *et al.* (1995), in Table 2.b., we can notice, in percentages, that the inverse situation took place.

Table 2.b. Distribution of anaphoric direct objects through time

	16th Cent.	17th Cent.	18th Cent.	19th Cent.	20th Cent.
Null object	10,6	12,6	18,4	45,9	82,8
Full pronoun	0,3	0	0	4,1	5,6
3rd p. clitics	89,1	87,4	81,6	50	11,2
NP	0	0	0	0	0,4

(Adapted from Table 4, in Kato, Cyrino & Corrêa (1995))

The high percentages (above 81%) in the 16th, 17th and 18th centuries show that during these centuries PB was a language with third person clitics and, practically speaking, the only variant form was the null object with a low frequency of use. Full pronouns were also used but in a vanishingly small percentage. In the 20th century, the third person clitic is almost lost, while the use of the null object increases from 10,6% to 82,8%.

The comparison of this diachronic table with the students' data suggest that higher education recovers the loss of the third person clitic, almost lost in the last century, and bans the strong pronoun. The use of the null object, which is constant through the schooling stages, is much more used now, in the 20th century diachronic table, as well as in the younger students' data.

As for the position of the clitic, Tables 3 and 4 show the occurrences of proclisis and enclisis according to degree of schooling in written and oral texts respectively.

Table 3. Occurrences of 3rd person accusative clitics according to position and schooling in written texts

Position	1st/2nd	3rd/4th	5th/6th	7th/8th	university	Total: 57
proclisis	-	4	1	4	6	15
enclisis	-	-	13	14	12	39
two clitics (proc. & encl.)			3			3

Table 3, only with the clitics, shows that, although in BP standard written texts it is the sanctioned direct object, the clitic is absent in the first grades and is extremely rare in the speech of all of the speakers. In this table we see that the first four occurrences of the third person are examples of proclisis:

- (3) a. elas nem acoqueciam (= nem a conheciam) (3rd grade)
 they didn't even her(cl)-know (= didn't even her(cl) know)
 'They didn't even know her'
- b. a tal moça a tinha pegado (3rd grade)
 such a girl it(cl) had taken
 'That girl had taken it'
- c. Ele a pegou (graduate)
 he it(cl) took
 'He took it'
- d. e (ele) lhe revistou(graduate)
 and (he) her/him (dat. cl) inspected
 'And he inspected him'

Enclisis comes just afterwards, in oral and written production, as the preferred position, and the change of position stimulated by formal instruction brought with it the occurrences of double clitic or one clitic and one pronoun, one on each side of the verb (three occurrences in the 5th/6th grade). These are examples like the following:

- (4) a. ...para o identificá-lo (5th)
 to him(cl) identify-him(cl)
 'to identify him'

Similarly, in Kato (1994), example (25) has the same underlying structure as (26). Thus, the difference between a sentence containing only the strong pronoun in object position, [26], and a sentence with a clitic and a strong pronoun, as [25], is that the clitic is null in the latter.

[25] Eu te_i amo você_i
 I you(cl) love you
 'I love you'

[26] *Eu ø-vi ele*
 I ø(cl)-saw he
 'I saw him'

(Extracted from Kato, et al., 1995)

Table 4, taken from oral data, shows that the clitic is absent in the elementary grades and extremely rare in the speech of all speakers.

Table 4. Occurrences of 3rd person accusative clitics according to position and schooling in oral texts

Position	illiterate	1st/2nd	3rd/4th	5th/6th	7th/8th	university	Total: 7
proclisis	-	-	-	-	-	1	1
enclisis	-	-	-	2	1	3	6

An important finding in these data is that enclisis was categorical with the third person clitic and was used with non-finite verbs. There are 17 occurrences of infinitive (with an auxiliary, a preposition or in a compound verb), as in (5-6):

-oral:

- (5) a. *O guarda foi buscá-lo.* (5th)
 the policeman went to fetch-him
 'The policeman went to fetch him'
- b. *ela olhava pra carteira, tentava agarrá-la* (6th)
 she looked at the wallet, tried to grab-it
 'She looked at the wallet, tried to grab it'
- c. *ela conseguiu pegá-lo* (graduate)
 she managed to catch-him
 'She managed to catch him'

-written:

- (6) a. *até que (ø) conseguiu pegá-la* (5th)
 so that (he) managed to get-it
 '...so that he managed to get it'
- b. *o garçom foi servi-las.* (8th)
 the waiter went to serve-them
 'The waiter went to serve them'
- c. *ela continuou a observá-lo* (graduate)
 she continued to observe-him
 'She continued to observe him'

We see here that in written production enclisis appeared in the 5th grade and proclisis almost disappears. Other studies on the clitic (Nunes, 1993; Cyrino, 1993; Kato, *et al.*, 1995) show that, according to the situation, the clitic could be either proclitic or enclitic. So, in this sense, the movement of the clitic seems to be lost now.

The general view of data shows that clitics have a [+ specific] feature. Table 5 demonstrates how [+ specific] direct objects gradually change from other forms to the clitic: In the 1st/2nd grades, 23 anaphoric objects were used, 17 [+ specific] and 6 [- specific], but none was a clitic. From the 3rd/4th years on, its percentage always increases. This means that the other forms appear more and more as [- specific].

Table 5. Occurrences of clitics according the specificity of the antecedent in written texts

Grade	[+ specific]		[- specific]	
	occ.	%	occ.	%
1st/2nd	0/17	-	0/6	-
3rd/4th	4/14	29	0/11	-
5th/6th	17/35	49	0/14	-
7th/8th	18/32	56	0/15	-
University	17/17	100	1/2	50

During the acquisition of the clitics, they are, in fact, in complementary distribution: while the clitics have a [+ specific] feature, the null form and the others show a [- specific] feature. In 64 examples of 3rd person clitics only one was [- specific] and occurred in a written text produced by an educated informant.

The variable rule analysis (Rand and Sankoff, 1990) was used here to investigate the influence of linguistic factors checking the clitic against other forms of third person anaphoric object used at school. The factors Animacy and Place of Referent and the factor combining Degree of Schooling and Type of Text were checked in this way.

a) Animacy: as for the semantic feature [animacy], clitics and null direct objects are compared in Tables 6 and 7.

Table 6. Occurrence of clitics for animacy and school grade in written texts.

Grade	[- animate]		[+ animate]		Total
	occ.	%	occ.	%	
1st/2nd	0/10	0	0	0/8	0/18
3rd/4th	1/12	8	3/6	50	4/18
5th/6th	4/21	19	13/15	87	17
7th/8th	3/14	21	15/19	79	18
University	8/8	100	9/9	100	17

Table 6 shows the written texts of the students with the predominant [+ animate] clitic in more than 50% (see examples in (7.a-e)). For educated informants, however, the clitic can have a [- animate] or a [+ animate] antecedent (see examples in (7 f-g):

- (7) a. (Ela/Ele) Levou-o até as moças para... (5th)
 (s/he) Brought-him to the girls for...
 'S/he brought him to the girls'
- b. ... (ela/ele) agarrou-a pelo colarinho (6th)
 (s/he) grabbed-her by the collar
 ...'s/he grabbed her by the collar'
- c. A garçonete serviu-as. (7th)
 the waitress served-them
 'The waiter served them'
- d. ...e (ela/ele) prendeu-a. (8th)
 and (s/he) arrested-her
 "...and s/he arrested him'
- e. ...e (eles) agarraram-na. (8th)
 and (they) grabbed-it
 "...and they grabbed it (= the purse)'
- f. ... a moça descreveu-a... (graduate)
 ... the girl described-her...
 'The girl described her (= the thief)'
- g. ...escondendo-a no bolso. (graduate)
 hiding-it in the pocket
 '...hiding it in his pocket'

Table 7: Occurrences of null direct objects for animacy and school grade in written texts.

Grade	[- animate]		[+ animate]		Total
	occ.	%	occ.	%	
1st/2nd	10	55	8	45	18
3rd/4th	11	78	3	22	14
5th/6th	17	90	2	10	19
7th/8th	11	74	4	26	15
University	0	0	0	0	0

Conversely, Table 7 shows the higher percentages of [- animate] null direct objects in elementary school students' texts (from 55 to 74%) while [+ animates] have low percentages.

When the program was performed, the probabilities for using the clitic confirms the percentages for animacy: 0.77 for the feature [+ animate] and 0.31 for [- animate]. This means that the clitic is preferred with a [+ animate] semantic feature while other forms, being the null object the most used, appear for the [- animate].

b) Place of referent: the results for the semantic referent of the anaphoric direct objects are in Table 8.

Table 8: The use of the clitic against other forms of anaphoric direct object.

Place of the reference	Occurrences	%	Probability
adverbial clause	1/3	33	0.10
discourse	39/359	11	0.45
coordinate clause	15/154	10	0.55
main clause	8/34	24	0.78

Looking for the semantic referent of the anaphoric direct objects, a new aspect of the question appeared. The speakers preferentially use the clitic if the syntactic ties between the anaphoric object and its referent are stronger, for instance, the referent being in the main clause and the object in a subordinate. In these circumstances, the clitic will be chosen with 0.78 of probability. Differences between percentage and probability are expected when data are submitted to a variable rule analysis, where all the factors are crossed with each other, as is the case here. Reference in the adverbial clause has 33% of occurrence, but only three times the referent was in this place. On the other hand, discursive reference is very common for the object, but the speakers prefer other forms of anaphoric object with it. The clitic was used only 11% of the times, or so, there is 0.45 of probability for the clitic to occur in this context.

The coordination of both clauses also influences the use of the clitic with 0.55 of probability. If we consider that what we call a coordinate clause may have a high degree of semantic dependency, as is the case of an event and its explanation, conclusion or the opposition of two events, the link between the clitic and its referent can be understood in terms of semantic dependency. If the referent is far from the anaphoric object, in the discourse, the informants prefer to use other forms, especially the null object. This leads to the analysis of the null object as a pronoun, discourse bound, while the clitic is more syntactically tied. This is interesting if we consider that they all belong to the same paradigm, the paradigm of the direct object.

c) Degree of Schooling and Type of Text: this third factor resulted as shown in Table 9:

Table 9: Probability of occurrence of the clitic with degree of schooling and type of text.

Schooling/Type of Text	Occurrences	%	Probability
7th/8th - oral	1/101	1	0.08
3rd/4th - written	4/33	12	0.27
University - oral	4/29	14	0.65
5th/6th - written	17/69	25	0.79
7th/8th - written	18/56	32	0.81
University - written	18/21	86	0.99

Table 9 shows data in increasing order: in general the written texts have a higher probability of presenting the clitic: the higher the degree, the higher the probability, but educated informants tend to use it speaking or writing. What is strong in this table is that 1st/2nd grades students don't use the clitic at all, and 3rd/4th grades students use it only in written texts. In the 7th/8th grades students practically don't use the clitics to speak. What do they use so? Preferentially the null form, considered as vernacular like the full pronoun and the referential noun. Only the clitic is "learned" in BP.

Because of the circumstances in which the clitic appears in BP, in the highly monitored style acquired at school, some interesting cases of hypercorrection were found, like (5):

- (8) a. (ø) Revistaram-*lo*. (5th)
 (they) inspected-him
 'They inspected him'
- b. que á (=há) muito tempo (ø) não *havia* (=a via) (7th)
 that there is much time (s/he) didn't *exist (=it-saw)
 'that s/he didn't see for a long time'

In (8.a) we have a problem of spell-out at the phonetic output and, in (8.b), a total misanalysis of the clitic, seen as a syllable of a larger word.

3. CONCLUSIONS

As the literature used in this study has shown, the clitic was the usual form in the past. In this way, we can say that writing is preserving the archaic system. The results, in fact, suggest that higher education recovers the loss of the third person clitic. Additional comments must be made, however.

The situation now, in some respects, is shown to be different from the time in which one could say that BP was a language with third person clitics. The "recovered" clitic has banned the strong pronoun. The anaphoric nominal phrase, practically absent until the 19th century, has now a place among illiterate speakers and 1st grade students. The null object, which appeared in a low frequency of use in past centuries is now the preferred non-nominal anaphor, taking the position of the clitic in the speakers' preference.

As for the position, even though the clitic could be proclitic or enclitic in other phases of its history, now enclisis was showed to be the preferred position. So, in this sense, the movement of the clitic seems to be lost. We can explain this preference for enclisis as a result of the action of the prescriptivism of the formal education. Grammars are always inducing the students to use this position as the "right" one.

In another way, if we compare the occurrence of the animacy feature in different stages of the students' data, the results from university level subjects show the use of clitics either with a [+animate] or a [-animate] antecedent, but in the earlier stages they are sometimes [-animate]. As the clitic was acquired, an interchange took place between both forms in a progressive and complementary way: in the 1st/2nd grade, when the clitic was unknown to the students, the null object, being the elected object, was indifferently [\pm animate] until the 7th/8th grades when it became preferentially [-animate]; on the other hand, in the 3rd/4th grade, the clitic began to be used preferably as [+animate] but, further on, it became indifferently [\pm animate], being the only form of referential direct object for educated informants.

Following Kato (1994) we can say that certain fossils of language are institutionally chosen as "grammatical", but belong, in effect, to stylistics. That is the case of the prescription that enclisis is obligatory in sentence initial position as well as in other contexts, even though it violates the present directionality of cliticization, as BP speakers generally begin the sentences

with a first person clitic, although they don't do the same with third person clitics, maybe because they are reduced to a vowel and don't have phonological support to begin the sentence.

- (9) a. Me dê isto.
 me give this
 'Give me this.'
 b. *O vejo todos os dias.
 him (I) see everyday
 'I see him everyday.'

The use of the strong pronoun as a direct object, on the other hand, as in (10) is an innovation licensed by universal grammar, but, nevertheless, banned in formal language.

- (10) Vejo ele todos os dias.
 (I) see he everyday
 'I see him everyday.'

In sum, these data show that, in fact, educated people use clitics, while uneducated people and students do not, but these clitics are not exactly the same as those which existed in earlier stages of the language. This happened due to the fact that they were lost and had to be learned again by an external force: prescriptivism. Cases of hyper-correction are signs of the learning of this anaphoric object like a second language element. But the new paradigm is possible only because all of these variants are possible in universal grammar. BP standard grammar is broad enough to contain all of these forms, and the choice for one or the other variant is influenced by linguistic factors, as we saw here, but also for extra-linguistic factors like formality of situation and education.

REFERENCES:

- Corrêa, V.R. (1991). *O objeto direto nulo no português do Brasil*. Dissertação de mestrado, Unicamp, Campinas.
- Cyrino, S.M.L. (1990). *O objeto nulo no português do Brasil: uma mudança paramétrica?* Unicamp, Campinas.
- Kato, M.A., Cyrino, S.M.L. and Corrêa, V.R. (1995). *The recovery of diachronic losses through schooling*. Paper presented in the annual NWAVE Congress, Pennsylvania.
- Kato, M. A. (1994). *Português Brasileiro falado: Aquisição em Contexto de Mudança Lingüística*. Paper presented in *Congresso Internacional sobre o Português*. Lisboa.
- Nunes, J. (1993). *Direção de cliticização, objeto nulo e pronome tônico na posição de objeto em português brasileiro*. In: *Português Brasileiro: uma viagem diacrônica: homenagem a Fernando Tarallo* (I. Roberts and M. Kato (Orgs.)). Editora da Unicamp, Campinas.
- Rand, D. and Sankoff, D. (1990). *GoldVarb. A variable rule application for the Macintosh*. Centre de recherches mathématiques, Université de Montréal, Montréal.