



THIRD-PERSON ANAPHORIC DIRECT OBJECTS: THE ACCUSATIVE CLITIC IN THE SPOKEN LANGUAGE OF CAXIAS, MARANHÃO

OBJETO DIRETO ANAFÓRICO DE TERCEIRA PESSOA: O CLÍTICO ACUSATIVO NA LÍNGUA FALADA DE CAXIAS – MA

DOI: 10.5281/zenodo.23050613



*João Gabriel Dias Sousa*¹

*Catarina de Sena Siqueira Mendes da Costa*²

ABSTRACT

This article, part of an ongoing doctoral dissertation, analyzes the use of the accusative clitic as a strategy for resuming third-person anaphoric direct objects with nominal antecedents in the speech of Caxias, Maranhão, considering its frequency, social distribution, and linguistic-discursive contexts. The study is grounded in Variationist Sociolinguistics, in Weinreich, Labov, and Herzog's (2006 [1968]) concept of orderly heterogeneity and in Labov's (2008 [1972]) assumption that linguistic variation is systematic and conditioned by internal and external factors. As empirical background, it draws on Pinto (2015), Lima (2016), Lauer (2022), and Santana (2025), whose findings indicate the low productivity or absence of the accusative clitic in Brazilian speech communities. Methodologically, this is a basic, documentary, quantitative, and descriptive-explanatory study combined with qualitative analysis. The corpus comprises 72 sociolinguistic interviews from the *Atitudes Linguísticas dos Falantes do Maranhão (ALFMA)* database, yielding 1,065 occurrences of third-person anaphoric direct objects with nominal antecedents. Only six involved accusative clitics (0.6%), which precluded multivariate analysis. Occurrences concentrated among four speakers with secondary or higher education, within the two younger age groups, and predominantly among men. All antecedents resumed were human; morphosyntactically, three enclitic forms occurred with infinitives and three proclitic forms with finite verbs. The accusative clitic is residual, socially restricted, and discursively circumscribed in the sample, not a productive vernacular strategy, though it remains available for occasional use.

1 Master's Degree in Linguistics from Universidade Estadual do Piauí. PhD candidate in Linguistics at Universidade Federal do Piauí, Brazil. Email: joaogabrielvg@gmail.com.

2 PhD in Linguistics from Universidade Estadual de Campinas. Professor at Universidade Federal do Piauí, Brazil. Email: costacatarina@uol.com.br.





Keywords: variationist sociolinguistics; third-person anaphoric direct object; accusative clitic; speech of Caxias, Maranhão; Brazilian Portuguese.

RESUMO

Este artigo, recorte de tese de doutorado em andamento, analisa o emprego do clítico acusativo como estratégia de retomada do objeto direto anafórico de terceira pessoa com antecedente nominal na fala de Caxias – MA, considerando frequência, distribuição social e contextos linguístico-discursivos de ocorrência. A investigação fundamenta-se na Sociolinguística Variacionista, na concepção de heterogeneidade ordenada de Weinreich, Labov e Herzog (2006 [1968]) e nos pressupostos de Labov (2008 [1972]). Segundo esses autores, a variação é sistemática e condicionada por fatores internos e externos à língua. Como antecedentes empíricos, mobilizam-se os estudos de Pinto (2015), Lima (2016), Lauar (2022) e Santana (2025), que evidenciam a baixa produtividade ou a ausência do clítico em comunidades de fala brasileiras. Metodologicamente, trata-se de pesquisa básica, documental, quantitativa e descritivo-explicativa, articulada à análise qualitativa. O corpus compreende 72 entrevistas sociolinguísticas do banco *Atitudes Linguísticas dos Falantes do Maranhão* (ALFMA), das quais se extraíram 1.065 ocorrências dessa variável. Entre elas, apenas seis corresponderam ao clítico acusativo, totalizando 0,6%, o que inviabilizou a análise multivariada. As ocorrências concentraram-se em quatro informantes com Ensino Médio e Superior, nas duas faixas etárias mais jovens e, predominantemente, entre homens. Todos os antecedentes retomados eram humanos. Morfossintaticamente, registraram-se três formas enclíticas com infinitivo e três proclíticas com verbos finitos. Conclui-se que o clítico é residual, raro, social e discursivamente circunscrito na amostra analisada, sem configurar estratégia produtiva do vernáculo. Permanece disponível, contudo, em usos pontuais.

Palavras-chave: sociolinguística variacionista; objeto direto anafórico de terceira pessoa; clítico acusativo; fala de Caxias – MA; português brasileiro.

1. INTRODUCTION

The third-person anaphoric direct object constitutes a domain of Brazilian Portuguese grammar in which different linguistic forms may serve to resume a referent previously introduced in discourse. In this relation, the anaphoric element refers back to an antecedent recoverable from the discourse chain. In Brazilian speech, such resumption may be realized through four variants: null object, anaphoric noun phrase, lexical pronoun, and accusative clitic. Taking the noun phrase “o livro” as the antecedent, for example, constructions such as “comprei o livro, mas ainda não li Ø”, with a null object; “ainda não li o livro”, with an anaphoric noun phrase; “ainda não li ele”, with a lexical pronoun; and “ainda não o li”, with





an accusative clitic, are possible. Although they differ in phonetic realization and morphosyntactic configuration, these forms may belong to the same domain of variation when they preserve the basic referent recoverable from context.

Among these strategies, the third-person accusative clitic occupies a singular position in spoken Brazilian Portuguese. The prescriptive tradition reserves the forms *o*, *a*, *os*, *as* and their allomorphs for pronominal expression of the third-person direct object (Bechara, 2019). Actual language use, however, reveals a historical reorganization of this system. Tarallo (1996) identifies the reduction of the accusative clitic as one of the features that came to distinguish Brazilian from European Portuguese grammar by the end of the nineteenth century, while Cyrino (1994), from a diachronic perspective, documents the expansion of the null object and the retreat of overt forms. In variationist research on speech, Duarte (1986) had already recorded the clitic as a rare strategy in comparison with alternatives such as the lexical pronoun and the empty category. Nunes (1996), in turn, relates this pattern to a grammar in which third-person accusative clitics ceased to be regular vernacular forms, whereas Pagotto (1996) attributes to formal education a relevant role in their introduction into or maintenance within speakers' repertoires. This relationship with schooling, however, does not entail an automatic correspondence between higher levels of education and use of the variant, nor does it justify interpreting every clitic occurrence as unequivocal evidence of formal speech. The degree of attention speakers pay to their own speech may also affect the distribution of socially evaluated variants (Labov, 2008 [1972]).

The recent studies selected as benchmarks for this investigation make the peripheral position of the clitic empirically visible, although they draw on different communities and methodological designs. In Florianópolis, Pinto (2015) recorded only one occurrence among the 666 tokens with nominal antecedents considered in the comparison, approximately 0.2%. In the popular speech of Fortaleza, Lima (2016) found 35 clitics in 8,255 occurrences, corresponding to 0.4%, leaving the variant in a residual position. In rural communities, the retreat was even more pronounced: Lauar (2022) recorded no clitics among 2,604 occurrences from the rural area of Santa Leopoldina, just as Santana (2025) found none in the 855 tokens





from the *quilombola* community of Maracujá. These findings converge in showing the low productivity of the form, but they do not permit its presence or absence to be attributed to a single cause. Education level, literacy practices, community characteristics, corpus design, and degree of speech monitoring may participate in distinct sociolinguistic configurations; accordingly, residuality and absence should not be treated as equivalent phenomena.

It is against this background that the present article is situated, as a subset of a broader investigation into variation in the third-person anaphoric direct object with a nominal antecedent in the speech of Caxias. If previous studies document residual occurrence of the clitic in two urban samples and absence in the two rural communities considered, it remains pertinent to determine the status this variant assumes in an urban community in the interior of Maranhão, a setting in which the phenomenon has not yet been described. The aim is to establish its frequency and examine its distribution among speakers, as well as the linguistic properties and discourse contexts associated with the recorded occurrences. The question guiding this article is therefore: “How can the use of the accusative clitic as a resumption strategy for third-person anaphoric direct objects with nominal antecedents in the speech of Caxias be characterized, considering its frequency, social distribution, and the linguistic contexts in which it occurs?”

The hypothesis is that the accusative clitic has low productivity in the spontaneous speech of Caxias and does not constitute an ordinary strategy for resuming third-person anaphoric direct objects with nominal antecedents. When it is recorded, its use is expected to display restricted social and discursive distribution, consistent with a form whose maintenance may be related to formal schooling and situations involving greater speech monitoring.

In line with this question, the general objective of this study is to analyze the use of the accusative clitic as a resumption strategy for third-person anaphoric direct objects with nominal antecedents in the speech of Caxias, considering its frequency, social distribution, and the linguistic-discursive contexts in which its occurrences are realized. Three specific objectives follow from this general objective: to quantify the absolute and relative frequency





of the accusative clitic within the set of occurrences of third-person anaphoric direct objects with nominal antecedents; to describe the distribution of occurrences according to education level, age group, and sex, identifying the social profiles of the speakers in whose production the variant was recorded; and to characterize the data linguistically, examining, among other relevant aspects, the semantic-referential properties of the antecedents, the form and position of the clitics, verb form, and the discourse contexts in which the variant occurs.

In addition to this introduction, this article comprises sections on the theoretical framework, methodology, results and discussion, and conclusions, which present, respectively, the foundations of the analysis, the methodological procedures, the interpretation of the data, and a synthesis of the study's main findings.

2. THEORETICAL FRAMEWORK

The analysis of linguistic variation adopted in this study is situated within Variationist Sociolinguistics, whose formulation is based on the principle that heterogeneity does not represent disorganization within the system. For Weinreich, Labov, and Herzog (2006 [1968]), linguistic structure accommodates orderly differentiation across speakers, styles, and contexts, such that alternation among forms is part of the competence of members of a speech community. From this perspective, regularity is not opposed to variation; rather, it is manifested precisely in the constraints that make certain realizations more likely in some environments than in others. The analytical task, therefore, is to identify the factors that regulate this distribution rather than to classify one form as correct and the others as deviations.

This principle is operationalized through the notion of the linguistic variable, which, in Labov (2008 [1972]), corresponds to a domain in which two or more variants compete in the realization of a comparable function. The choice among them is therefore conditioned by factors internal and external to language. Linguistic constraints encompass structural, semantic-referential, morphosyntactic, and discursive properties; extralinguistic factors





concern, among other dimensions, education level, age group, sex, and style. These factors do not operate mechanically or in isolation. Variationist analysis seeks to determine how they combine probabilistically, thereby explaining why the same community has competing forms without using them with the same frequency in every context.

Extending this model to syntax requires careful delimitation of the envelope of variation. The debate revisited by Labov (1978) shows that syntactically distinct constructions can be treated as variants when they preserve a comparable referential function within rigorously circumscribed contexts. For this reason, syntactic investigation cannot be reduced to counting forms: quantification makes it possible to measure frequencies and conditioning, while qualitative interpretation of contexts helps account for discursive, stylistic, and social values that cannot be exhausted by statistical indices. In the domain of the anaphoric direct object, this articulation is necessary because the competing strategies differ formally but resume referents previously introduced in discourse.

In Brazilian Portuguese, the third-person anaphoric direct object constitutes a domain in which four strategies may occupy the same functional space: the accusative clitic, the lexical pronoun, the anaphoric noun phrase, and the null object. The prescriptive tradition privileges the clitics *o*, *a*, *os*, *as* and their allomorphs as forms for third-person direct-object resumption (Bechara, 2019). Brazilian speech, however, has historically reorganized this system. Tarallo (1996) identifies the reduction of the accusative clitic as one of the features distinguishing Brazilian from European Portuguese grammar by the end of the nineteenth century, while Cyrino (1994), from a diachronic perspective, documents the expansion of the null object and the retreat of overt forms. In variationist descriptions of speech, Duarte (1986) had already recorded the clitic as a rare strategy competing with the lexical pronoun and the empty category.

The low productivity of the clitic in Brazilian speech does not, however, mean that the form has disappeared entirely from the social repertoire. Nunes (1996) argues that generations of Brazilian speakers came to acquire a grammar in which third-person accusative clitics ceased to be regular vernacular forms. Pagotto (1996), in turn, attributes an important role to



formal education in introducing or maintaining these forms in speakers' repertoires. In this sense, schooling matters not as a factor that automatically determines clitic use, but as a setting for sustained contact with writing, the standard norm, and practices of linguistic evaluation. Thus, a form with low productivity in the vernacular may remain available to different degrees among speakers exposed to literate practices.

With regard to stylistic monitoring, which is also implicated in this relationship, the attention speakers devote to their own speech may alter the distribution of socially evaluated variants, particularly when one of them is associated with prescriptive correctness or prestige (Labov, 2008 [1972]). Spontaneous speech and more closely monitored speech are not discrete compartments, but points along a continuum of attention to speech. Accordingly, the occasional emergence of a form learned through schooling does not require the entire utterance to conform fully to the standard norm. The analytical interest lies in the conditions under which the variant becomes available and in the type of discourse context in which it is activated. This caution prevents each clitic occurrence from being taken, by itself, as proof of formality or mastery of the written norm.

The four studies that constitute the recent overview of the issue in this article make this retreat empirically visible, although they are based on different communities and methodological designs. In Florianópolis, Pinto (2015) analyzed urban speech at two points in time and found only one accusative clitic. Considering the subset with nominal antecedents adopted for comparison, the occurrence represents 1 of 666 tokens, approximately 0.2%. Such rarity precludes any specific modeling of the variant and does not justify attributing its use to a particular social constraint. The study nevertheless shows that the clitic may survive residually in an urban sample alongside much more productive strategies. With respect to animacy, Pinto (2015) found that [-animate] antecedents slightly favored the null object, a result that indirectly points to greater compatibility between animate referents and overt strategies.

Also in an urban context, Lima (2016) provides a particularly relevant Northeastern benchmark. In the NORPOFOR corpus, which focuses on the popular speech of Fortaleza, 35





clitic occurrences were recorded among 8,255 tokens, or 0.4%, alongside 38.2% anaphoric noun phrases, 37.7% null objects, and 23.6% lexical pronouns. The clitic therefore remained residual even in a large sample. The author excluded the formal elocution modality because she considered it more susceptible to speech monitoring, a decision that underscores the methodological importance of the degree of attention paid to speech. In addition, the sample included speakers with up to eleven years of schooling, but no higher education. In the multivariate analysis, education level conditioned the lexical pronoun rather than the clitic in isolation: the group with 9 to 11 years of schooling disfavored the lexical pronoun and favored the set of the remaining forms. The finding cautions against inferring that higher levels of schooling necessarily entail greater clitic use.

In studies of rural communities, this pattern of low productivity becomes more pronounced, beginning with the rural area of Santa Leopoldina, where Lauar (2022) recorded no clitic among 2,604 occurrences. The author interpreted this absence as evidence that the form was not part of the community's vernacular, relating it to schooling, writing, and formal contexts. Even so, her own analysis shows that education level did not “restore” the clitic: its statistical effect fell on the anaphoric noun phrase. In another sociolinguistic configuration, Santana (2025), investigating the rural *quilombola* community of Maracujá, likewise found no clitics among the 855 occurrences analyzed. Because the low educational diversity of the sample prevented education level from being controlled, the author related the absence of the form to sociohistorical conditions, limited access to schooling, and lower exposure to prestigious norms.

The contrast between the two urban samples, in which the clitic appears residually, and the two rural communities, in which the variant is not recorded, is suggestive but does not allow an “urban setting” to be converted into a direct cause. Community type, educational composition, literacy practices, corpus design, and degree of monitoring may combine in different ways. Rather, the comparison reveals that the clitic occupies a peripheral position in all the studies and that its observability depends on social and methodological conditions that



must be examined jointly. For this reason, very low frequencies require inferential caution: absence, residuality, and productivity represent distinct statuses.

Finally, there is a relevant linguistic convergence for understanding overt resumption forms. Lima (2016), Lauar (2022), and Santana (2025) show that animate, human, and/or specific antecedents favor the lexical pronoun above all, whereas less animate or less individuated referents show greater compatibility with the null object or nominal resumption. In Lauar (2022), [+animate +human] antecedents strongly favored the full pronoun; in Santana (2025), [+animate] and [+specific] likewise favored this strategy. These findings cannot be transferred automatically to the clitic, whose scarcity precluded analyses of the variant itself in the studies reviewed. They nevertheless establish an important referentiality axis: overt pronominal realization tends to concentrate among referents with greater semantic-discursive salience.

From this theoretical and empirical body of evidence emerges a picture in which the accusative clitic must be analyzed simultaneously as a variant within a variable syntactic system and as a socially marked form because of its historical relationship with the standard norm, writing, and schooling. At the same time, the studies caution against monocausal explanations: schooling, monitoring, referential properties of the antecedent, and community profile do not operate independently and do not produce universal effects. It is at this intersection of structural regularity, social distribution, and discursive conditions of use that the presence or absence of the clitic acquires sociolinguistic significance.

3. METHODOLOGY

3.1 Characterization of the Study

This study, conceived as a subset of a broader investigation³ into variation in the third-person anaphoric direct object with a nominal antecedent in the speech of Caxias, is situated within the field of Variationist Sociolinguistics. It therefore assumes that distinct forms may

3 Doctoral dissertation.





compete within the same functional domain and display distributions systematically conditioned by linguistic and social factors, in accordance with Weinreich, Labov, and Herzog's (2006 [1968]) concept of orderly heterogeneity and the developments proposed by Labov (2008 [1972]). In terms of purpose, the research is basic; with regard to technical procedures, it is documentary, since it uses a previously constituted oral corpus; in terms of approach, it is quantitative, following Prodanov and Freitas (2013); and, with respect to its objectives, it is descriptive-explanatory, with an exploratory dimension arising from the absence of a previous description of the phenomenon in the community investigated (Gil, 2017; Severino, 2016). In this subset, however, quantification was combined with qualitative analysis of the contexts of occurrence, a procedure required by the residual frequency of the accusative clitic. Frequencies and distributions were therefore taken as a descriptive basis without being dissociated from the sociolinguistic interpretation of the data (Guy; Zilles, 2007).

3.2 Constitution and Data Collection

The data come from the *Atitudes Linguísticas dos Falantes do Maranhão* (ALFMA) database, developed at the Centro de Estudos Superiores de Caxias, Universidade Estadual do Maranhão (CESC/UEMA), under the coordination of Antônio Luiz Alencar Miranda. The database contains recorded sociolinguistic interviews conducted face to face and organized around a semistructured interview guide, with open-ended questions that elicit narratives, evaluations, and accounts of personal experience (Miranda, 2016). For this study, 72 interviews were used, corresponding to 72 speakers from the urban area of Caxias.

In stratifying the sample by sex, age group, and education level, two sexes were considered, male and female; three age groups, 18 to 30 years, 31 to 49 years, and 50 years or older; and four education levels, lower elementary education, upper elementary education, secondary education, and higher education. Crossing these factors yielded 24 social cells,





each comprising three speakers, for a total of 72 participants. The distribution resulted in 36 speakers of each sex, 24 per age group, and 18 per education level.

The envelope of variation comprised occurrences of third-person anaphoric direct objects with nominal antecedents, realized through four variants: null object, anaphoric noun phrase, lexical pronoun, and accusative clitic. Cases were included when a clear anaphoric relation was established between an overt nominal antecedent—SN/DP or a proper name—and a third-person direct-object position, with a referent recoverable from discourse and the possibility of covariation among the variants. The resumption had to be produced by the speaker, although the antecedent could have been introduced by the interviewer. Sentential, clausal, or propositional antecedents were excluded, as were exophoric or purely situational references; referents that were merely inferable, ambiguous, or not reliably recoverable; structures that did not constitute third-person anaphoric direct objects; crystallized expressions without an identifiable anaphoric referent; and stretches in which interruption, overlap, or transcription uncertainty compromised classification.

3.3 Data Analysis

Corpus processing began with a complete reading of the 72 transcribed interviews. During this procedure, occurrences belonging to the envelope of variation were identified and extracted in their discourse contexts, with the resumption form, nominal antecedent, clause in which the anaphor occurred, and relevant linguistic and social factors recorded. After the established criteria had been applied, the overall empirical dataset comprised 1,065 occurrences. Each token was classified and then coded using a string of characters corresponding to the factors controlled for processing in GoldVarb X, developed by Sankoff, Tagliamonte, and Smith (2005).

Although the overall corpus contained all four variants, only six occurrences corresponded to the accusative clitic. This extremely low frequency made it impossible to conduct a specific multivariate analysis taking the clitic as the application value of the rule.





With only six tokens distributed across different social and linguistic contexts, the estimation of factor weights would be subject to empty or excessively sparse cells, parameter instability, and excessive dependence on individual occurrences, thus failing to provide a sufficient empirical basis for reliable probabilistic inferences.

Because of this limitation, the six occurrences were examined in terms of absolute and relative frequency; distribution by education level, age group, and sex; semantic-referential characterization of the antecedents; morphosyntactic characterization of the forms; position of the clitic in relation to the verb; verb form; and analysis of the discourse contexts in which the variant appeared. The set was also compared qualitatively with the results of Pinto (2015), Lima (2016), Lauar (2022), and Santana (2025). This procedure made it possible to identify concentrations, patterns, restrictions, and associations in the data, but it does not permit probabilistic favoring to be attributed to any factor. Thus, differences observed according to education level, sex, age group, humanness of the referent, or any other property were treated as descriptive tendencies and qualitative associations rather than as statistically demonstrated effects on the use of the accusative clitic.

4. RESULTS AND DISCUSSION

Only six valid occurrences were recorded, a number insufficient to support a reliable multivariate analysis. Chart 1 synthesizes the social, linguistic, and discursive profile of these tokens, which are examined individually in the sequence that follows.

Chart 1 – Social, linguistic, and discursive profile of the six accusative clitic occurrences

Token	Speaker	Social profile	Nominal antecedent	Clitic form	Position and verb form
1	INF22	Male; 31–49; secondary education	<i>uma pessoa da minha família</i> [+human, fem. sg.]	<i>ajudá-la</i>	Enclisis; infinitive
2	INF28	Male; 18–30; secondary	<i>tua mãe</i> [+human, fem.]	<i>respeitá-la</i>	Enclisis; infinitive





Token	Speaker	Social profile	Nominal antecedent	Clitic form	Position and verb form
		education	sg.]		
3	INF28	Male; 18–30; secondary education	<i>os políticos</i> [+human, masc. pl.]	<i>os coloca</i>	Proclisis; finite verb, present
4	INF28	Male; 18–30; secondary education	<i>os políticos</i> [+human, masc. pl.]	<i>os colocou</i>	Proclisis; finite verb, past
5	INF32	Male; 31–49; higher education	<i>os vizinhos</i> [+human, masc. pl.]	<i>os vejo</i>	Proclisis; finite verb, present
6	INF65	Female; 18–30; higher education	<i>a pessoa</i> [+human, fem. sg.]	<i>constrangê-la</i>	Enclisis; infinitive

Source: prepared by the authors, based on data from the ALFMA database.

When the social distribution of the six occurrences is considered, they are found to have been produced by only four speakers, corresponding to Interviews 22, 28, 32, and 65. This already imposes an initial interpretive caution, since the variant was not widespread among speakers in the sample but instead appeared as a form localized in a small number of individuals. This caution is compounded by the fact that no occurrence was recorded among speakers with elementary education. The four occurrences produced by speakers with secondary education and the two produced by speakers with higher education point to an association between residual clitic use and higher levels of formal schooling. Although the small number of tokens does not permit the statistical claim that education level favored the variant, the profile of the occurrences suggests that school-mediated contact with the written norm plays an important role in the occasional maintenance of this form.

Considered qualitatively as well, the distribution by age group adds another relevant element to the interpretation. If the clitic were a conservative form remaining from an earlier stage of the language, a higher concentration among older speakers might plausibly be expected. The variant, however, appeared only in the two younger age groups, with four occurrences among speakers aged 18 to 30 and two among those aged 31 to 49. No





occurrences were recorded among speakers aged 50 or older. In the sample analyzed, this pattern suggests that the clitic is not maintained as a feature of generational conservatism among older speakers, since its occurrence appears to depend more on schooling and possible monitoring effects than on continuous vernacular transmission.

The analysis of the sex variable, although it likewise could not be subjected to statistical inference with respect to the clitic, revealed a concentration among male speakers, who produced five of the six occurrences, compared with only one produced by a female speaker. This result, however, was interpreted cautiously because three occurrences belonged to the same male speaker in Interview 28. Male predominance therefore could not be interpreted as social favoring. Rather, it indicated that the variant appeared idiosyncratically and was concentrated in particular individual profiles. Even so, the single female occurrence shows that the variant is not restricted to the male group.

Among the linguistic constraints observable qualitatively, the clearest restriction was semantic-referential in nature. All clitics resumed human antecedents: “*uma pessoa da minha família*”, “*tua mãe*”, “*os políticos*”, “*os políticos*”, “*os vizinhos*”, and “*a pessoa*”. No occurrence involved an animal or inanimate antecedent. From a prescriptive perspective, this pattern is significant, since the accusative clitic could resume referents with different degrees of animacy. In these Caxias data, however, its limited use was restricted to [+human] referents. Thus, the variant does not behave as a fully available form for direct-object resumption in general, appearing instead as a rare resource activated in contexts of high human referentiality.

The morphosyntactic profile of the occurrences likewise showed a highly restricted pattern. Three tokens occurred with infinitive verbs and feminine enclitics in *-la*: “*ajudá-la*”, “*respeitá-la*”, and “*constrangê-la*”. The other three occurrences contained the clitic *os* in proclitic position before finite verbs: “*os coloca*”, “*os colocou*”, and “*os vejo*”. No instances of the forms *o*, *a*, *lo*, *los*, *no*, *na*, *nos*, *nas* were recorded. This distribution shows that the variant is realized in only a few formal configurations, alternating between enclisis in infinitival contexts and proclisis with a masculine plural form. Even so, the occurrence of the





clitic did not entail full adherence to the prescriptive pattern, since the excerpts in which it appeared also retained features characteristic of orality and vernacular speech.

Considering the tokens individually, the first occurrence was recorded in the speech of a male speaker aged 31 to 49 with secondary education, in a context in which he described a situation involving helping a family member: “*nós fizemos eu e minha esposa fizemos o que nós podia fazer pra ajudá-la*” (INF22). In this token, the form *ajudá-la* resumed a human feminine referent in an infinitival structure, constituting an enclitic occurrence compatible with the prescriptive pattern. The same utterance, on the other hand, displayed features of orality, such as “*nós podia*”, indicating that the clitic may arise sporadically even in a formulation that is not fully monitored in accordance with the standard norm.

The second case came from a male speaker aged 18 to 30 with secondary education, in response to a question about his relationship with his mother: “*Com tua mãe você se dá bem? Como é a relação de vocês? Resposta- Sim. É fruto de respeito e eu tento procurar sempre respeitá-la*” (INF28). The presence of the clitic here appeared to be related to schooling and, at the same time, to the evaluative and relatively monitored nature of the response.

The third occurrence, also from Interview 28, appeared in a comment about politics: “*quando, principalmente os políticos, é o alvo mais fácil de, de falar a respeito. Porque se a população os coloca lá é porque elas querem que ele, que eles cuidem com responsabilidade lá dos direitos deles*” (INF28). In this excerpt, the clitic *os* resumed the noun phrase “*os políticos*” and occurred in proclitic position before the verb *coloca*. The construction formed part of public and argumentative discourse in which the speaker evaluated political responsibility and the actions of the population.

The fourth token, still in the speech of the Interview 28 participant, retained the same antecedent and thematic field: “*E se eles tivessem essa preocupação acho que talvez não teria tanta, acho que normalmente não roubaria tanto assim porque se reocuparia mais com as pessoas e principalmente com quem os colocou lá*” (INF28). The form *os colocou* resumed “*os políticos*”, now with a past-tense verb.



Elsewhere in the sample, the fifth occurrence came from a male speaker aged 31 to 49 with higher education, while discussing his relationship with his neighbors: “*Gosto do local que mora? Os vizinhos são legais? Resposta- Sim, amo Caxias, quero casar aqui em Caxias poucos os vejo mais me trata bem tento ser cortez com todo mundo.*” (INF32). In this token, the form *os vejo* resumed a plural human referent and occurred in proclitic position. The construction appeared in a response marked by a positive evaluation of the city and by the speaker’s courteous self-presentation, as evidenced in “*tento ser cortez com todo mundo*”.

The sixth and final occurrence was produced by a female speaker aged 18 to 30 with higher education, while commenting on situations involving linguistic correction or embarrassment in response to forms considered amusing: “*a palavra fica muito engraçada aí não tem como, a gente tenta segurar o sorriso mais geralmente eu nem corrijo a pessoa não porque eu fico com medo de constrange-la*” (INF65). The form *constrange-la* resumed the human feminine antecedent “*a pessoa*” and occurred in a metalinguistic context in which the interviewee herself reflected on correction, error, laughter, and embarrassment. This token is significant because it associated the clitic with a situation in which language was explicitly thematized; in other words, with a context more sensitive to normative evaluation.

Taken together, the six excerpts show that the occurrence of the accusative clitic in this study is not randomly distributed. The variant appeared in the speech of more highly educated speakers, in contexts of greater discursive elaboration, often linked to themes of respect, responsibility, courtesy, correction, or social evaluation. This combination of factors supports the interpretation that, in the sample, the clitic functions as a residue of a form learned or reinforced through schooling, without constituting an ordinary strategy in the vernacular of Caxias.

The six tokens found in this study do not contradict the general trend described in the literature; rather, they qualify it. The variant was present, unlike what was observed in the two rural communities investigated by Lauar (2022) and Santana (2025), but it remained at such a low level that it could not be treated as a statistically productive strategy, more closely approximating the results observed by Pinto (2015) and Lima (2016).



At the linguistic level, all antecedents resumed by the clitics bore the [+human] feature, an aspect that likewise establishes an indirect dialogue with the studies reviewed. In Lima (2016), Lauar (2022), and Santana (2025), high referentiality, associated with features such as [+animate], [+human], and [+specific], was frequently invoked to explain the favoring of overt forms, particularly the lexical pronoun. In the speech of Caxias, although the small number of clitics precluded any probabilistic inference, it is noteworthy that the variant accorded prestige by prescriptive grammar appeared only with human and salient referents.

5. CONCLUSIONS

The present article aimed to analyze the use of the accusative clitic as a resumption strategy for third-person anaphoric direct objects with nominal antecedents in the speech of Caxias, considering its frequency, social distribution, and the linguistic-discursive contexts in which its occurrences are realized. Examination of the data made it possible to delimit the status of this variant in the sample and showed that its presence, although effectively attested, is extremely restricted, both quantitatively and in its distribution across speakers and the contexts in which it appears.

In response to the research question, the results show that the accusative clitic has residual frequency in the data examined: only six occurrences were identified, corresponding to 0.6% of the corpus and produced by four speakers. No tokens were recorded among speakers with elementary education, whereas four were found among speakers with secondary education and two among those with higher education. With respect to age group, the occurrences were limited to the two younger groups, with none among participants aged 50 or older. Distribution by sex yielded five male occurrences and one female occurrence; this difference, however, must be qualified because three of the five male tokens belong to the same speaker. Linguistically, all resumed antecedents were [+human]. Morphosyntactically, three enclitic forms associated with infinitive verbs and three proclitic forms preceding finite verbs were observed. From a discursive perspective, the occurrences appeared in passages





involving, to varying degrees, elaboration of the response, social evaluation, respect, responsibility, courtesy, and metalinguistic reflection.

The hypothesis formulated in the introduction was therefore confirmed in its central aspect: the accusative clitic has low productivity and does not constitute, in the sample analyzed, an ordinary strategy for resuming the anaphoric direct object. The concentration of tokens among speakers with secondary and higher education is compatible with the prediction of a possible relationship with formal schooling, but the data do not permit the claim that education level determines or statistically favors use of the variant. Likewise, the contexts in which the clitics occurred provide qualitative indications of greater attention to speech in certain passages, particularly when the discourse content involves evaluation, responsibility, courtesy, or language itself. The small number of occurrences, however, prevents this interpretation from being converted into statistical evidence of greater speech monitoring.

In comparison with the studies adopted as references, the Caxias results occupy an intermediate position between absence and residuality. Unlike Lauar (2022) and Santana (2025), who recorded no clitics in their samples, the present investigation identified the variant in the corpus analyzed. At the same time, its extremely low frequency brings the observed pattern closer to those described by Pinto (2015) and Lima (2016), in which the clitic likewise remained residual. The study thus adds evidence from an urban community in the interior of Maranhão to the broader picture of anaphoric direct-object realization in Brazilian Portuguese. More precisely, it demonstrates that the accusative clitic is not entirely absent from the speech data examined, although neither does it constitute a productive or recurrent resource; within the limits of this sample, it is residual and socially and discursively circumscribed.

This interpretation must take into account a central limitation: six occurrences do not provide a sufficient basis for probabilistic generalizations about the factors associated with clitic use. The concentrations identified describe the sample but do not justify causal relationships or statistically demonstrated conditioning effects. Future studies using larger corpora, different speech styles, and data from other communities in Maranhão may





determine whether the observed association with particular social profiles, human referents, and contexts of greater discursive elaboration recurs in different configurations. Even so, the present study supports one conclusion: in the urban sample from Caxias, the accusative clitic survives as an actually realized possibility, yet one that is peripheral, rare, and restricted. This contributes, on the basis of local empirical evidence, to a more precise characterization of one of the least productive forms in the variable system of anaphoric direct-object resumption in Brazilian Portuguese.

REFERENCES

- BECHARA, Evanildo. *Moderna gramática portuguesa*. 39. ed. rev. e ampl. Rio de Janeiro: Nova Fronteira, 2019.
- CYRINO, Sônia Maria Lazzarini. *O objeto nulo no português do Brasil: um estudo sintático-diacrônico*. 1994. 227 f. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 1994.
- DUARTE, Maria Eugênia Lamoglia. *Variação e sintaxe: clítico acusativo, pronome lexical e categoria vazia no português do Brasil*. 1986. Dissertação (Mestrado) – Pontifícia Universidade Católica de São Paulo, São Paulo, 1986.
- GIL, Antonio Carlos. *Como elaborar projetos de pesquisa*. 6. ed. São Paulo: Atlas, 2017.
- GUY, Gregory R.; ZILLES, Ana. *Sociolinguística quantitativa: instrumental de análise*. São Paulo: Parábola Editorial, 2007.
- LABOV, William. Where does the linguistic variable stop? A response to Beatriz Lavandera. *Working Papers in Sociolinguistics*, Austin, n. 44, p. 1-17, abr. 1978.
- LABOV, William. *Padrões sociolinguísticos*. Tradução de Marcos Bagno, Maria Marta Pereira Scherre e Caroline Rodrigues Cardoso. São Paulo: Parábola Editorial, 2008 [1972].
- LAUAR, Aline Berbert Tomaz Fonseca. *Variação da expressão do objeto direto anafórico: uma análise da fala rural de Santa Leopoldina – ES*. 2022. 167 f. Tese (Doutorado em





Linguística) – Programa de Pós-Graduação em Linguística, Universidade Federal do Espírito Santo, Vitória, 2022.

LIMA, Tereza Maria de. *O emprego do objeto direto anafórico de terceira pessoa na língua falada de Fortaleza*. 2016. 157 f. Dissertação (Mestrado Acadêmico em Sociobiodiversidade e Tecnologias Sustentáveis) – Universidade da Integração Internacional da Lusofonia Afro-Brasileira, Redenção, 2016.

MIRANDA, Antônio Luiz Alencar. *Atitudes linguísticas dos falantes do Maranhão*. Projeto de pesquisa aprovado pela FAPEMA e UEMA. Caxias: UEMA, 2016.

NUNES, Jairo Morais. Direção da cliticização, objeto nulo e pronome tônico na posição de objeto em português brasileiro. In: ROBERTS, Ian; KATO, Mary Aizawa (org.). *Português brasileiro: uma viagem diacrônica*. 2. ed. Campinas: Editora da Unicamp, 1996.

PAGOTTO, Emílio Gozze. Clíticos, mudança e seleção natural. In: ROBERTS, Ian; KATO, Mary Aizawa (org.). *Português brasileiro: uma viagem diacrônica*. 2. ed. Campinas: Editora da Unicamp, 1996.

PINTO, Cecília Augusta Vieira. *Variação do objeto anafórico acusativo na fala de Florianópolis*. 2015. 164 f. Dissertação (Mestrado em Linguística) – Programa de Pós-Graduação em Linguística, Universidade Federal de Santa Catarina, Florianópolis, 2015.

PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar de. *Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico*. 2. ed. Novo Hamburgo: Feevale, 2013.

SANKOFF, David; TAGLIAMONTE, Sali A.; SMITH, Eric. *GoldVarb X: a variable rule application for Macintosh and Windows*. Toronto: Department of Linguistics, University of Toronto, 2005.

SANTANA, Cinara de Andrade Silva. *As estratégias de retomada anafórica do objeto direto: uma análise sociolinguística da fala da comunidade quilombola do Maracujá, Conceição do Coité – BA*. 2025. 124 f. Dissertação (Mestrado em Estudo de Linguagens) – Programa de Pós-Graduação em Estudo de Linguagens, Universidade do Estado da Bahia, Salvador, 2025.





SEVERINO, Antônio Joaquim. *Metodologia do trabalho científico*. 24. ed. São Paulo: Cortez, 2016.

TARALLO, Fernando. Diagnosticando uma gramática brasileira: o português d'aquém e d'além-mar ao final do século XIX. In: ROBERTS, Ian; KATO, Mary Aizawa (org.). *Português brasileiro: uma viagem diacrônica*. 2. ed. Campinas: Editora da Unicamp, 1996.

WEINREICH, Uriel; LABOV, William; HERZOG, Marvin I. *Fundamentos empíricos para uma teoria da mudança linguística*. Tradução de Marcos Bagno. São Paulo: Parábola Editorial, 2006 [1968].

Recebido em: 27/08/2026

Aprovado em: 14/09/2026

Publicado em: 30/09/2026

