

RIBEIRO DA SILVA, Lara. The acquisition of the indeterminate and passive *se* clitic in Brazilian Portuguese: outcomes of an acquisition through writing. *ReVEL*, vol. 16, n. 31, 2018. [www.revel.inf.br/eng]

## THE ACQUISITION OF THE INDETERMINATE AND PASSIVE *SE* CLITIC IN BRAZILIAN PORTUGUESE: OUTCOMES OF AN ACQUISITION THROUGH WRITING

Lara Ribeiro da Silva<sup>1</sup>

lararsilva22@gmail.com

**ABSTRACT:** Kato, Cyrino & Correa (2009), discussing the distance between written and spoken Brazilian Portuguese (BP), show that the school system tends to recover certain stylistic features of 19th Century Portuguese grammar. Those features are treated as grammatical fossils by the authors. Amongst them, we can find the indeterminate and passive clitic *se*. This research examines the acquisition of such clitic, mainly by highly educated speakers of BP. One of its goals is to see to what extent the *se* clitic is acquired as a second language (L2) grammar, as Kato (2005a) has proposed. To achieve that, we examine different proposals for L2 acquisition and also the diachronic changes involving the clitic, focusing on its present syntactic status. Our results support her proposal. We have analyzed five different corpora, one for each stage in the acquisition of written BP. Our results show that the indeterminate interpretation overruled the passive one, becoming the preferred form. They also show that the acquisition of the *se* clitic is dependent on schooling.

**KEYWORDS:** L2 Acquisition, Generative Grammar; clitic *se*.

**RESUMO:** A fim de explorar o tema da distância entre o que se fala e o que se escreve no Português Brasileiro (PB), Kato, Cyrino e Correa (2009) propõem que a instituição escolar recupera determinados traços estilísticos da gramática portuguesa do século XIX e os postula como gramaticais, sendo um desses fósseis gramaticais o clítico *se* nas funções de indeterminador e de apassivador. Diante disso, o presente trabalho tem como objetivo investigar como se dá a aquisição desse clítico, principalmente por falantes letrados do PB. Pretende-se averiguar a hipótese de que a aquisição desse clítico para o PB se comporta tal como um item de uma língua estrangeira (L2), conforme proposto por Kato (2005a). Com efeito, após a observação e a análise de um *corpora* composto por cinco bancos de dados, cada um representativo de uma fase da aquisição da escrita, viu-se que a função indeterminadora de *se* tomou espaço na língua e parece ser preferida em relação à forma apassivadora independente do material em que a sentença é produzida. Do mesmo modo, confirmou-se a hipótese de que a aquisição de *se* indeterminador e de *se* apassivador dependem da escolarização.

**PALAVRAS-CHAVE:** Aquisição de L2; Gramática Gerativa; clítico *se*.

---

<sup>1</sup> Master in Linguistics at the State University of Campinas – Unicamp.

## INTRODUCTION

According to Kato (2005), linguistic innovations focus on speech, and on writing for those who are still learning this feature. On the other hand, the writing of literate adults or of individuals already proficient in writing restricts their innovations to the lexical scope and, in this context, their syntax remains conservative since the writing itself resists linguistic changes. With this, it is observed that at the end of the schooling process, literate adults also begin to show essentially writing marks in their speech. This happens because, according to Kato, Cyrino & Correa (2009), school uses certain fossils of language, which are certain stylistic marks that are no longer part of the internalized grammar of Brazilian Portuguese (BP) and postulate them as grammatical. Kato (2017) observes that preschool children only present forms determined of learning by selection, whereas, after the beginning of schooling, it is possible to note variation in their speech between forms acquired by selection and forms learned through instruction.

Therefore, by being inserted into the world of letters successfully and acquiring the rules of the writing grammar, their I-language is altered, and it is possible to question how these marks of writing coexist with the speech naturally acquired by the child. In view of this questioning, Kato, Cyrino and Correa (2009) argue that the linguistic knowledge of a literate speaker of BP is therefore constituted by a nuclear grammar, reflected in its orality, and a periphery verified by the presence of such writing in the reproduction of stylistic rules acquired during the schooling process. In another way, the BP literate speaker, according to these authors, behaves just like a bilingual speaker who alternates between forms present in his nuclear grammar and his peripheral grammar by means of code-switching.

This fact is related to Kato's (1999, 2005a) proposal that the written language taught in school is so distant from the one spoken by PB speakers that it resembles the result of learning an L2. According to Kato (1999):

O problema principal no ensino de português é a distância entre a língua falada, vernacular, pré-letramento, e a língua escrita institucionalizada, que a escola visa como meta da aprendizagem. Em todas as línguas se observam diferenças entre a fala e a escrita, mas no português do Brasil (PB) existe um fosso de tal ordem que para a criança aprender a ler e a escrever, ela enfrenta a tarefa da aprendizagem da língua estrangeira. (KATO 1999: 1)<sup>2</sup>

---

<sup>2</sup> “The main problem in Portuguese teaching is the distance between the spoken, vernacular, pre-literacy language, and the institutionalized written language, which the school aims at as a learning

Therefore, this work aims to investigate the influence of the learning of writing on speech, taking as hypothesis that the indeterminate and passive clitic *se*, as in sentences (1) and (2), behaves as a linguistic loss retrieved by the school and that its natural acquisition by BP speaking children is hampered since other forms of indetermination of the thematic role that subjects have incorporated into the internal grammar of these speakers. The acquisition of this mark, then, would only be possible through the schooling process, when it would become part of the peripheral grammar of the speaker. Passive *se* structures would be even more difficult to be acquired since, even in the school environment, its appearance has been restricted by internal processes of the language.

- (1) Vive-se mal neste país.  
'Live *se* badly in this country.'  
One lives badly in this country.
- (2) Por meio do rádio difundiram-se as principais ideias sobre o assunto.  
'Through the radio spread *se* the main ideas on the subject.'  
Through the radio the main ideas on the subject were spread.

In order to do so, we collected data from five representative databases of different moments of BP acquisition and writing learning. The methodology of data collection will be described in the third section of this paper.

This work is also based on the study of Nunes (1990), who considers that there was a process of syntax reanalysis in which the expletive subject of indeterminate sentences with *se* was reinterpreted as a referential null pronoun and the indeterminate constructions with *se* became more frequent than the constructions with the passive form. This generated other language changes involving the clitic *se* in Brazilian Portuguese, as its deletion in finite sentences and insertion in infinite sentences.

This paper is organized as follows: the first section will briefly focus on presenting the theoretical framework that supported the research; the second section will describe the data collection and analysis processes performed; and, finally, the

---

goal. In all languages there are differences between speech and writing, but in Brazilian Portuguese (BP) there is a gap of such order that for the child to learn to read and to write, it faces the task of learning a foreign language." (KATO 1999: 1, my translation).

third and will highlight the discussions about the results found as well as the final considerations of this research.

## **1. THEORETICAL FRAMEWORK**

This first section will be dedicated to the theoretical review of the framework that supports this research: first, theories on second language acquisition will be presented and second some theories on code-switching. Finally, we will see studies on the status of the indeterminate and passive *se* clitic in the theoretical framework of Generative Grammar.

### **1.1 SECOND LANGUAGE ACQUISITION**

As aforementioned, Kato (1999, 2005a) states that the literate speaker behaves like a bilingual, due the code-switching process between his nuclear grammar and his grammar acquired at school, being variation a result of this process. Therefore, the linguistic knowledge of this BP speaker is different from that of a speaker who was not so affected by schooling and school literacy (KATO, CYRINO & CORREA, 2009). This knowledge resembles, thus, to a second language (L2) acquisition.

In order to investigate the processes that are behind this acquisition according to Generative Grammar, we would like to initially elucidate the concepts of Universal Grammar (UG), E-Language and I-Language of Chomsky (1981, 1986), which are pillars of the Generative Theory.

Considering that, at the end of the acquisition of the first language (L1), the child makes use of linguistic resources that were not available in the given stimulus, the so-called input, the theory of Universal Grammar arises in response stating that there is an intern initial acquisition device, UG itself, which contains all the possible properties of natural languages and allows any language to be learned by the child.

E-Language can be defined as language as an idiom, a social, political and geographical object. The Chomskian focus is, therefore, the I-Language: individual, intensional and internal, it represents the linguistic knowledge that a speaker has of his language. Thus, it contains the so-called Principles – a set of characteristics that do not vary from one language to another – and Parameters – marked properties with binary (+) or (-) values that define the variation between languages.

When all the values of the Principles are defined from UG, Chomsky (1986) states that the speaker's nuclear grammar is defined. This nuclear grammar can be accompanied by a marked periphery (CHOMSKY 1981), in which the phenomena of change and innovation are inserted. Therefore, the child would arrive at school with their nuclear grammar established, but only those that are massively exposed to the grammar of writing will have in their periphery the so-called linguistic fossils (KATO 2005a).

If, for first language acquisition, the initial stage of the acquisition is UG, for second language acquisition, this initial stage still generates controversies among the researchers. White (2003) presents these discussions and divides hypotheses into two large groups: those who claim that the initial state of L2 acquisition is L1 and those that support the idea that L2 is acquired from Universal Grammar.

One theory inserted in the first group is the Null Access Hypothesis: The Full Transfer Full Access Hypothesis by Schwartz & Sprouse (1994, 1996 apud WHITE 2003). These authors state that a speaker in the L2 acquisition process starts from the complete grammar of his L1. In this way, with the exception of the phonetic matrix of the morpholexical items, the initial state of the acquisition of L2 is the final state of L1 and, therefore, the development of L2 is determined by the conjugation of the following factors: initial state, input, UG apparatus and learning conditions.

Similarly, Clahsen (1988 apud WHITE 2003) also states that the parametric options available in UG for the L1 acquisition phase does no longer exist for L2 speakers. Bley-Vroman (1989 apud WHITE 2003), in turn, parts from this proposal to postulate the Fundamental Difference Hypothesis, according to which the acquisition of L2 depends on affective factors, motivation and attitudes arising from formal instruction in L1.

Kato (2005b) rejects this hypothesis because it does not take into account that the parametric values of L2 are different than those of the speakers' L1. To exemplify this problem, she assumes that

Um americano aprendendo o português poderá começar preenchendo sempre o objeto, como em inglês, mas se ele aprende efetivamente a gramática do português, deverá começar a usar nulos em contextos relevantes. Poderíamos pensar que isso é relativamente fácil pois se trata de uma propriedade que pode ser o resultado apenas de um apagamento fonético. Lembremos-, porém, que o objeto nulo seria apenas uma das manifestações de uma propriedade mais abstrata, que licencia também outras construções. De fato, línguas de objeto nulo não licenciam apenas objetos nulos, mas exibem outras características (cf Raposo 1998) (...). Só

poderíamos afirmar que o americano remarcou apropriadamente o parâmetro do objeto nulo se apresentar o conjunto de propriedades que se aprende através da definição de um valor de parâmetro. Além disso, algumas dessas propriedades seriam dedutivamente aprendidas, não necessitando de exposição direta a elas. (KATO 2005b: 196)<sup>3</sup>

In contrast to the total Full Access Hypothesis, some researchers, such as Selinker (1972 apud WHITE 2003) rely on interlanguage systematicity, on the comparison between L1 speakers and L2 speakers and on the poverty of stimulus. Since interlanguage grammars often have features that do not come from L1 or L2, but are in line with universal constraints, it is seen that there is no other possibility but the access to UG. In addition, Birdsong (1992), another researcher in this perspective, reaffirms that L2 learners acquire parametric values that are different from their L1 and that, therefore, it is a knowledge that goes beyond the input of their L1. Platzack (1996 apud WHITE 2003), author of The Initial Hypothesis of Syntax, based on this evidence, proposes that the initial stage for the acquisition of L2 is identical to the initial stage of the acquisition of L1.

Finally, we assume the hypothesis of Partial Access, according to which access to the UG is only allowed through L1, as defended by Kato (2003, 2005b) for the acquisition of school grammar as L2 and supported by Meisel (2000), who argues that the parametric values set by the speaker at the end of the acquisition of L1 cannot be altered in the acquisition of L2. This theory can take two other forms: a) parametric values can be changed, since only the principles activated during the acquisition of L1 are accessed by L2; and b) only principles which are not parametrized can be accessed for L2 to be acquired. In addition, in accordance with the Null Access Hypothesis, Kato (2003) also points out that the intermediate grammar of the L2 learner (interlanguage) presents UG constraints that are not found in the input provided to the speaker.

---

<sup>3</sup> “An American person learning Portuguese can start by always filling the object, as in English, but if they really learn Portuguese grammar, they should start using nulls in relevant contexts. We might think that this is relatively easy because it is a property that may be the result of only a phonetic deletion. Let us remember, however, that the null object would be only one of the manifestations of a more abstract property, which also permits other constructions. In fact, null object languages do not only license null objects, but display other characteristics (see Raposo 1998) (...). We could only affirm that the American (speaker) has properly remarked the parameter of the null object if they present the set of properties that is learned through the definition of a parameter value. In addition, some of these properties would be deductively learned, not requiring direct exposure to them.” (Kato 2005b: 196, my translation).

## 1.2 CODE-SWITCHING

As seen previously, Kato, Correa and Cyrino (2009) propose that the alternation between the peripheral grammar and the nuclear grammar of the literate subject occurs as codeswitching (CS). This phenomenon can be conceptualized as Poplack (2001) does:

Code-switching (CS) refers to the mixing, by bilinguals (or multilinguals), of two or more languages in discourse, often with no change of interlocutor or topic. Such mixing may take place at any level of linguistic structure, but its occurrence within the confines of a single sentence, constituent or even word, has attracted most linguistic attention (POPLACK, 2001, p. 1).

Kato (1994), looking at the CS process between two very different languages, has already argued that the more distant the languages involved, the more restricted CS will be to isolated items, as is the case of Portuguese and Japanese speakers. It is also important to highlight the difference between CS and borrowing (REYES 1982 apud KATO 1994). Both are codemixing processes, but the latter involves only the insertion of a single word, while the first requires the interaction of two grammars and their distinct syntaxes.

The permeability of grammars, that is, what is transferred from one language to another and to what extent, becomes, thus, a point of discussion between authors who mainly investigate bilingualism. Kato (2001) takes up this question and analyzes two diametrically opposed views: Weinreich (1953 apud KATO 2001) and Bickerton (1981 apud KATO 2001) who argue that elements of the L2 grammar permeate a bilingual discourse but are not incorporated into language as a code; and Silva-Corvalan (1993 apud KATO 2001) who states that any linguistic trait can be transferred from one language to another provided they are compatible in structure.

Kato (2001) adopts the proposal of Silva-Corvalan (1993 apud KATO 2001) and states that CS finds favorable conditions to emerge in the following situations:

- a) Formal learning of a foreign language by a child after his/her literacy in the mother tongue has been completed;
- b) Late acquisition of a L2 by an adult immigrant;
- c) Acquisition of a second language at the beginning of schooling;
- d) Simultaneous acquisition of more than one language.

All situations, except the first one, involve an immersion context in L2. This indicates that CS would be, in some way, culturally motivated. On the other hand, it is seen that schooling is a determining factor, since in situations like (c), it is noticed that the first grammar becomes fossilized to the detriment of the one acquired later by the speaker (KATO 1996). This situation would be the one that would most closely resemble the situation studied here: the child arrives at school with their nuclear grammar and begins to live with another grammar at school, in which they will be literate. However, instead of becoming more fluent in school grammar, the speaker alternates between the two codes.

### 1.3 THE INDETERMINATE AND PASSIVE *SE* CLITIC IN GENERATIVE GRAMMAR

Nunes (1990), in his famous work "*O Famigerado Se*", synchronically and diachronically analyzes constructions with indeterminate and passive *se* clitics in order to define a theoretical status for these constructions and verify the processes of change by which these constructions passed. Sentences (3) and (4) respectively exemplify these constructions:

(3) Alugam-se casas. (NUNES 1990)

‘Rent(pl)-se houses.’

One rents houses.

(4) Aluga-se casas. (NUNES 1990)

‘Rent(sing)-se houses.’

One rents houses.

(5) A casa foi alugada. (NUNES 1990)

‘The house was rented.’

The house was rented.

In opposition to so-called periphrastic passives sentences, as (5), the author groups (3) and (4) under the designation of pronominal passives sentences. Before beginning his analysis, he considers what was previously said by Galves (1987) and Everett (1985), indicating his points of disagreement.

The problem surrounding Everett's (1985) ideas involves his assumption that there is only one lexical entry for the various uses of the clitic *se* (reflective, ergative, indeterminate, passive, and inherent) in all Romanic languages. To support this

proposal, the author states that it can be generated anywhere and moved to anywhere. Indeterminate *se* would then be generated under INFL, where it receives its thematic role of the external argument and nominative case, besides being bound by the expletive that occupies subject position. Its cliticization would take place in Phonetic Form and its traces [+ Case] and [+ Argument] would be selected. On the other hand, passive *se* since it is associated with derived subjects, would cliticize in the syntax and lose its anaphoric character. The clitic would then receive external argument thematic role in INFL and accusative case, leaving the nominative case for the internal argument that agrees with the verb. Galves (1986) already judged the theoretical costs of this analysis to be very high, mainly due to the dissociation between Case and Argument. Nunes (1990), like Galves (1986), disagrees with this analysis, but neither does he reinforce Galves's proposal, although it seems more reasonable to him.

Based on Borer's free indexing hypothesis (BORER 1985 apud NUNES 1990), Galves (1986) considers that *se*, as it is a lexical clitic that, like the others, is associated with argumentative positions, can never be an argument or play the role of identifying empty categories in these positions. Thus, in passive sentences, the clitic would perform the thematic role of the external argument and the NP following to the verb the thematic role of internal argument. However, only the NP would receive Case (nominative) since it is coindexed with AGR; *se*, as the realization of the thematic function attributed by the verb to the external argument, would not need to receive Case. Sentences with indeterminates, in contrast, would present a chain receiving the thematic role of the internal argument and nominative case, and *se* would be coindexed with AGR, making PRO subject of a pronominal anaphora. For Nunes (1990), this proposal is questionable since sentences with passive *se* would be generated without a subject and Galves would, like Everett (1985), disregard the anaphoric relation of the clitic in these sentences as well.

The author then reaffirms the need for an analysis that solves these theoretical questions. His proposal originates, like those of the other authors, from the work of Chomsky (1981) in which he postulates that if the accusative case is absorbed, the NP postposed to the verb receives a nominative case as it constitutes a link of a CHAIN and the position of the subject is checked.

Despite the formal differences, Nunes clarifies that the interpretation of the indeterminate reference is the same for sentences with indeterminate *se* and passive *se*. This is why sentences like (6) are often seen as ambiguous:

(6) *Tinha-se alugado uma bela casa.* (NUNES 1990)

‘Had-se rented a beautiful house.’

One had rented a beautiful house.

A beautiful house was rented.

Then he analyzes the clitic *se* in the following way: in passive sentences *se* absorbs the thematic role and the accusative case of the external argument and is connected by a thematic chain link that involves the internal argument, indeterminating the reference of the external argument. On the other hand, for indeterminate *se*, the author proposes that this clitic only needs to satisfy his traits [+an, -pro] and, therefore, it is linked by a null referential pronoun, bearer of the thematic role of the external argument and of the nominative case, making an anaphoric-pronominal relation of indeterminate reference.

In addition to considerations regarding the status of *se*, Nunes (1990) also presents a diachronic study on the changes undergone by these structures in BP. He observed that, from the nineteenth century, when constructions with indeterminate *se* became majoritarian, this clitic came to be inserted in sentences with infinitives and gerunds as (7) and (8), respectively:

(7) *É impossível se achar lugar aqui.* (NUNES 1990)

‘Is impossible *se* find place here.’

It is impossible to find a place here.

(8) *Está se sujando muito a rua.* (NUNES 1990)

‘Is *se* dirtying much the street.’

One is messing up the street a lot.

Specifically remarking type (7) sentences, Nunes (1990) explains:

Até o século XVIII estruturas infinitivas, como as finitas, privilegiavam as construções com *se* apassivador (com concordância). Entretanto, quando no século XIX, as passivas pronominais com discordância *se* tornam majoritárias, as construções infinitivas dão um salto considerável,

ultrapassando as finitas na preferência pela construção com *se* indeterminador. Isso pode constituir um indício de que é nesse momento a se dar o fenômeno da inserção de *se* junto a infinitivos, pois a inserção deve ter fornecido combustível adicional para a mudança que substituiu construções infinitivas com concordância por construções sem concordância. Dito de outro modo, no século XIX o português brasileiro, ao invés de PRO, começa a preferir a categoria *pro* para ocupar a posição de sujeito de infinitivas. (NUNES 1990: 115)<sup>4</sup>

The author therefore calls this clitic “*Se Sarney*”, in reference to the publication of Josué Machado in the section “Língua” of the Brazilian magazine “Imprensa” from 1989 (nº 25), of which the following is part:

‘Como se evitar a decomposição do país.’ Ganha um prêmio quem descobrir a função do ‘se’ desse título publicado no Caderno de Sábado do Jornal da Tarde de algumas semanas atrás. O prêmio será uma página inteira da obra *Marimbondos de Fogo*, do dr. José Sarney, em francês. Por que esse prêmio? Porque esse ‘se’ pode ser batizado com o nome de Sarney. Crueldade, talvez, mas é um *se* muito em evidência, sem função nem utilidade. Para não dizer ridículo e errado. Mas isso não fica bem dizer. O ‘se’ Sarney é aquele que vem grudado à maioria dos verbos no infinitivo, sem que se saiba para que serve.’<sup>5</sup>

## 2. DATA COLLECTION AND ANALYSIS

This section will describe the process of data collection and analysis that seeks to reaffirm the hypothesis that the clitic *se* in Brazilian Portuguese is part of the peripheral grammar of the literate subject of BP since its acquisition occurs only through schooling.

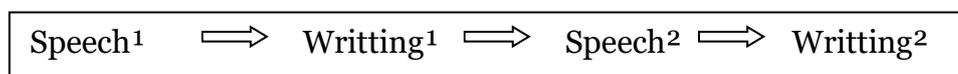
---

<sup>4</sup> Until the eighteenth-century infinitive structures, like the finite ones, privileged constructions with passive *se* (with agreement). However, when, in the nineteenth century, passive pronominal sentences with disagreement become majority, infinitive constructions take a considerable leap, surpassing the finite ones in preference for constructions with indeterminate *se*. This may be an indication that it is at this moment that the phenomenon of insertion of the infinite constructions occurs, since the insertion must have provided additional fuel for the change that replaced infinitive constructions with agreement by constructions without agreement. Put another way, in the nineteenth century Brazilian Portuguese, instead of PRO, begins to prefer the category *pro* to occupy the position of subject of infinitives. (NUNES 1990: 115, my translation).

<sup>5</sup> “How *SE* avoid the decomposition of the country.’ Win a prize to discover the function of the ‘se’ of that title published in the Saturday Book of ‘Jornal da Tarde’ a few weeks ago. The prize will be a full page of the work *Marimbondos de Fogo*, by dr. José Sarney, in French. Why this prize? Because this ‘se’ can be called Sarney. Cruelty, perhaps, but it is a *se* very much in evidence, without function or utility. Not to mention ridiculous and wrong. But that’s not good to say. The ‘se’ Sarney is one that is attached to most of the verbs in the infinitive, without knowing what it is for.” (my translation).

## 2.1 METHODOLOGY

The choice of databases was based on Kato's (1986) scheme, which illustrates the distance between speech and writing and its influence, as represented below:



**Picture 1<sup>6</sup>:** Scheme of the sequence of the learning process with its influence from writing to speech (taken from Kato, 1986, p. 11).

In this scheme, what is called Speech<sup>1</sup> corresponds to the pre-written period, characterized as unconscious and innovative. Writing<sup>1</sup>, conscious and innovative, that comes soon after this period, would then be very close to a transcription of speech. The conservatism of writing would appear only in Writing<sup>2</sup>, which, already under the effect of school instruction, is neutral in relation to speech. Coming to the point of our interests in this work, the scheme then presents Speech<sup>2</sup>, which is conservative and essentially filtered by the technology of writing, this modality of speech can be conscious or not, leaving the grammatical innovations to Speech<sup>1</sup> and Writing<sup>1</sup>. Thus, we see only the lexical innovations in Writing<sup>2</sup> and Speech<sup>2</sup>, because the writing technology represses innovations and makes the speaker return to forms already eliminated, or at the threshold of disappearance (KATO 1999). To verify the behavior of the clitic in more formal writing situations, a Writing<sup>3</sup> was added to this research.

We analyzed Five databases, each corresponding to one step of the scheme. Finally, in order to represent the process of write learning more consistently, data from a fifth database were also collected. They are the following:

- Representing Speech<sup>1</sup>: approximately 200 transcribed recordings of two children between 2 and 5 years old belonging to the “*Projeto Aquisição da Linguagem Oral*” collection provided by *Centro de Documentação Alexandre Eulalio* (CEDAE - IEL/UNICAMP);
- Representing Writing<sup>1</sup>: Approximately 100 essays from the 1st to 4th grade children's writing database on the project entitled “*A relevância teórica dos dados singulares no processo de aquisição da linguagem*”

---

<sup>6</sup> The translation of the scheme was made by us for this paper.

*escrita*” developed by Prof. Dr. Maria Bernadete Marques Abaurre, Prof. Dr. Raquel Salek Fiad and Prof. Dr. Maria Laura Trindade Mayrink Sabinson in IEL/UNICAMP;

- Representing Writing<sup>2</sup>: 120 essays considered as above the average of the university entrance exam provided by the *Comissão Permanente para Vestibulares* (COMVEST);
- Representing Speech<sup>2</sup>: 60 transcribed recordings belonging to the NURC/SP Project (Urban Authentic Standard, São Paulo), of which 40 correspond to the conversation between two informants and 20 to the formal statements made individually by the informants.
- Representing "Writing<sup>3</sup>", in which writing would be consistently consolidated: 80 news articles from the newspaper *Folha de São Paulo* taken from its policy book in the following intervals: from 1980 to 1990 and from 2000 to 2010.

In all, 1897 occurrences of *se* were collected, which are divided among: indeterminate *se*, exemplified in (9); passive *se*, exemplified in (10); pronominal *se*, exemplified in (11); reflexive *se*, exemplified in (12); *se* Sarney, exemplified in (13); and reciprocal *se*, exemplified in (14).

(9) Espera-se que as futuras gerações saibam dar o devido valor a elas. (COMVEST)

‘Hope *se* that the future generations know give the owing value to them.’

It is hoped that future generations will appreciate their value.

(10) Só assim, através da conscientização, que se podem alcançar mudanças importantes. (COMVEST)

‘Only this way, through awareness, that *se* can achieve changes important.’

Only this way, through awareness, one can achieve important changes.

(11) A dengue que se prolifera em um determinado local. (Children’s essays)

‘The dengue that *se* proliferates in a certain place.’

The dengue that proliferates in a certain place.

(12) ...pa se enfeitá... (CEDAE)

‘...To *se* adorn.’

To adorn yourself.

(13) Está se usando esse critério. (NURC)

‘Is *se* using this criterion.’

It is been used this criterion.

(14) Dois amigos que se amavam muito. (Children’s essays)

‘Two friends that *se* love much.’

Two friends who loved each other very much.

To perform the data analysis, they were systematized according to external and internal variables. The following external variables were considered: database, type of recording (for NURC data) and date of performance.

The internal variables were selected based on the characteristics considered to be determinant for the use of each type of *se* by the authors that make up the theoretical reference of this research. we controlled:

- The type of *se*: indeterminate, passive, Sarney, pronominal, reflexive and reciprocal;
- The type of verb: direct transitive, indirect transitive, ditransitive, unaccusative, unergative and state verbs;
- Verbal tense: present, past tense, tense imperfect, future tense, infinitive, gerund, subjunctive present, subjunctive future;
- The agreement between the verb and the explicit internal argument;
- The animacy feature of the explicit internal argument;
- The position of the explicit internal argument;
- The clitic position in relation to the verb.

The descriptive statistics presented below were obtained with the Microsoft Excel 2013 program. In the face of the distributions found, a multivariate logistic regression analysis was performed to verify if the interaction between the variables is statistically relevant.

As a clarification, since the objective of the work is to observe how the acquisition of the indeterminate *se* and passive *se* occurs, data in which the clitic was suppressed were not collected.

## 2.2 RESULTS

The first thing that can be observed in relation to the data was their distribution among the databases (Table 1). As expected, both indeterminate and passive *se* have an inexpressive presence in the data obtained from the CEDAE and from the children's essays. For this reason, in these databases, the presence of the reflexive *se*, the pronominal *se* and the reciprocal *se* was also controlled to emphasize that the acquisition of these other types of *se* seems to behave more naturally and more prematurely, whereas the acquisition of the indeterminate and passive clitic seems to depend exclusively on the process of schooling – an affirmation allowed by the significant increase of these marks in vestibular essays, NURC transcriptions and *Folha de São Paulo* articles. We verified that, although the particles are homophones, only those whose clues are present in the input provided to the child in the acquisition phase are acquired. In this way, children who have in their input typical syntactical marks of writing can acquire these elements more easily.

Type of <i>se</i>		CEDAE	Children's essays	COMVEST	NURC	Folha de São Paulo	Overall	%
Indeterminate	N	2	16	114	533	205	870	45.86%
	%	0.23%	2%	13%	61%	24%		
Passive	N	0	4	52	44	60	160	8.43%
	%	0.00%	2.50%	32.50%	27.50%	37.50%		
Pronominal	N	15	33	0	0	0	48	2.53%
	%	31.25%	68.75%	0.00%	0.00%	0.00%		
Reflexive	N	52	68	0	0	0	120	6.33%
	%	43.33%	56.67%	0.00%	0.00%	0.00%		
Sarney	N	0	0	100	425	150	675	35.58%
	%	0.00%	0.00%	14.81%	62.96%	22.22%		
Reciprocal	N	12	12	0	0	0	24	1.27%
	%	50.00%	50.00%	0.00%	0.00%	0.00%		
Overall	N	81	133	266	1002	415	1897	100.00%
	%	4.27%	7.01%	14.02%	52.82%	21.88%		

**Table 1:** Distribution of *se* among the databases.

Although there is a pattern in the data distribution, the logistic regression results showed that there is no interaction between the two variables, type of database and database (p. value = 0.9977). However, when analyzing the interaction between type of record (oral or written) and type of *se*, we verified that the differences are statistically relevant for sentences with *se* Sarney, passive *se* and indeterminate *se* (p.value <0.01), that is, the number of occurrences is sensitive to the material through which speech is transmitted.

There was also a significant interaction between age group (child, adolescent and adult) and type of *se* (p. value < 0.01). This is due to the fact that, as the clitic acquisition depends on schooling, its appearance depends on the age of the speaker.

In addition, it should be noted that the two occurrences of indeterminate *se* in the data collected from CEDAE do not corroborate the idea that this acquisition is natural. Since all the children who participated in the project come from families with a high level of education, the language that served as input is much closer to the norm, which seems to have contributed to the production of this data. Furthermore, these two occurrences, presented in (15) and (16), were observed in the speech of the same child. The first, at 3 years and 9 months; the second, at 4 years and 2 months:

(15) Cabou-se a história. (CEDAE)

‘Ended-*se* the story.’

The story ended.

(16) É aqui que se guarda. (CEDAE)

‘Is here that *se* save.’

Here one keeps.

Sentence (15) also helps to recognize that exposure to written data from children's books and cartoons, for example, is another form of contact with a language that is closer to the norm. This type of sentence is typical of children's stories and, for this reason, may be crystallized by the child. The vernacular language in BP can, therefore, present this type of structure.

This statement is also valid for data collected from children's essays, in which 16 sentences were recorded with indeterminates. Also, being the daughter of highly educated parents, the child who produced most of the data probably also had input

from a parental speech that was more filtered by writing, which motivates the presence of data such as (16), (17) and (18) in their notebooks of the first cycle of Basic Education:

(16) Aquilo que se fez anteriormente. (Children's essays)

'What that *se* done previously.'

What one did previously.

(17) Quando ele terminou de dizer se ouviu uma voz esquisita dizendo...

(Children's essays)

'When he finished of say *se* heard an odd voice saying...'

When he finished saying, there was an odd voice saying...

(18) Ai a gente recorta a ponta e se faz a mesma coisa com outro papel colorido. (Children's essays)

'Then we cut the tip and *se* do the same thing with another colored paper.'

Then we cut the tip and one does the thing with another colored paper.

Kato's (2017) claim about the acquisition of third-person clitics in BP also fits in these situations:

a variação na criança começa, em geral, com o início da escolarização com formas adquiridas em sua gramática nuclear e variantes fornecidas pela escola. Há casos, porém, de crianças expostas a ambientes altamente letrados, com pais que regularmente leem para os filhos, sem haver instrução de regras, situação em que a criança já pode exibir variação. (KATO 2017: 8)<sup>7</sup>

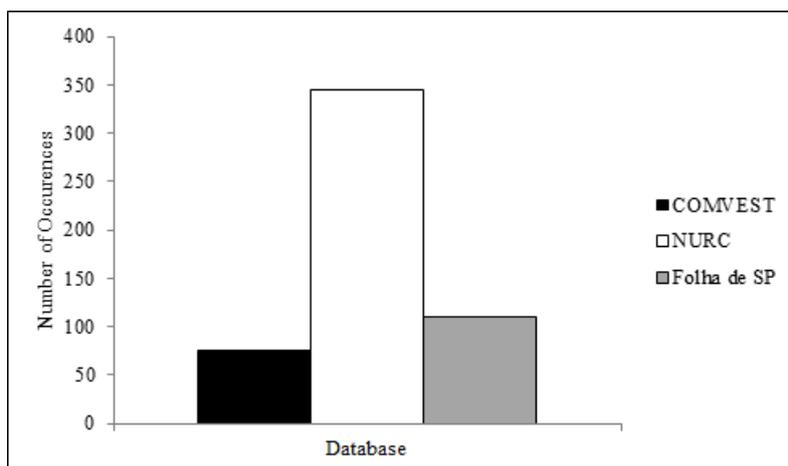
With this, it is possible to reaffirm that the acquisition of indeterminate and passive *se* is, in fact, linked to schooling. After 12 years of Elementary and Secondary Education, what is observed in the COMVEST essays is an increase to 13% of sentences with indeterminate *se*.

Another point to be highlighted in Table 1 is the considerable use of structures with *se* Sarney. The first data with this type of structure appears in COMVEST essays which have 63% of their occurrences registered in the NURC transcriptions, surpassing the indeterminate and passive forms of the clitic. This situation could be explained by the degree of formality and pressure of the use of the prestige modality

---

<sup>7</sup> "variation in the child usually begins with schooling with forms acquired in their nuclear grammar and variants provided by the school. There are cases, however, of children exposed to highly literate environments, with parents who regularly read to their children, without rule instruction, a situation in which the child may already exhibit variation." (KATO 2017: 8, our translation)

in each of the situations since, as Nunes (1990) states, it is the innovative form most accepted by the speakers but condemned by Traditional Grammar. In addition, one sees a much higher frequency of *se* with infinitives in speech data of the NURC in detriment of the writing data of the COMVEST and Folha de São Paulo. Chart 1 shows the distribution of these constructs between the databases in which they were found:



**Chart 1:** Amount of *se Sarney* per database.

Thus, it is seen that their acquisition depends on schooling, but the form does not necessarily depend on the written record to occur.

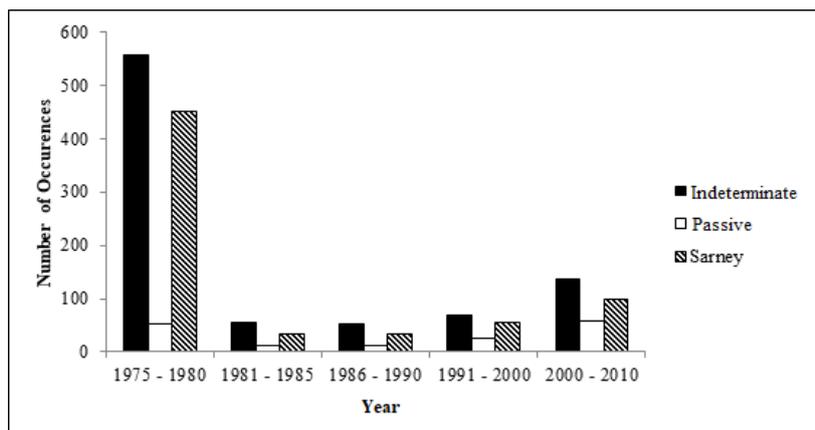
Likewise, the considerable presence of passive *se* in a competitive text, such as COMVEST, gives it a degree of adequacy to the written form, which is quite satisfactory. However, contrary to what was expected, the percentage of occurrences of passive *se* in speech data reaches close to 28%. Even though the highest percentages are in the written materials, it is seen that the form is not yet completely extinct from speech.

The discrepancy between the percentages of indeterminate *se* and passive *se* also draws attention. Considering what Nunes (1990) has argued in his work on the substitution of sentences with passive *se* by sentences with indeterminate *se*, it is seen that the percentage of use of the indeterminate form exceeds 77% in the total of the data. Even in written materials, there is a preference for the use of the indeterminate rather than the passive. This corroborates the idea that

No português brasileiro, essa mudança linguística praticamente já está em seus momentos finais: a construção com *se* indeterminador se tornou a forma canônica e a antiga construção com *se* apassivador ficou relegada à escrita. O dialeto brasileiro registra ainda duas mudanças mais recentes: o

apagamento de *se* indeterminador em construções finitas e sua inserção em construções infinitas. (NUNES 1990: iv)<sup>8</sup>

Thus, the process of syntactic reanalysis that has caused indeterminate *se* to substitute passive *se* in BP is confirmed. However, the passive disappearance and the expressive fall of the indeterminate is a linguistic change that is under way. In view of this, it is necessary to investigate the frequency of these data over the years in which they were collected. Chart 2 shows this frequency:



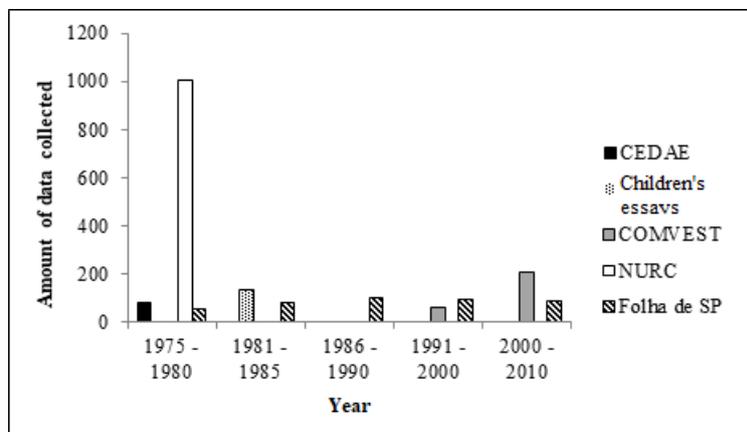
**Chart 2:** Frequency of each type of *se* found per year.

Chart 2 reaffirms the higher frequency of indeterminate *se* in relationship to the other two in question. Logistic regression tests, however, did not show any interaction between the type of variables per year (p. value = 0.9963).

Returning to chart 1, we can see that *se* Sarney is the second most used, leaving only less than 50 occurrences of passive *se* in all analyzed periods. On the other hand, it reveals that the greatest amount of indeterminate *se* and *se* Sarney occurred in the period between the years 1975-1980. Between 1981 and 1990, the frequencies found were very similar, not exceeding 100 occurrences of each type of *se*. Since 1991, the frequency of use of these types has grown again, especially after the 2000s. This growth is probably due to the fact that most data from the 2000s are written and that the pressure of using traditional forms is more intense. This fact may also explain the increase in the number of passive *se* in those years, in order to reach the highest number of occurrences. Since the situation requires formality, the peripheral grammar of the speaker, which contains the forms of the school, is triggered and the

<sup>8</sup> “In Brazilian Portuguese, this linguistic change is practically already in its final moments: the construction with indeterminates became the canonical form and the old construction with passive *se* was relegated to writing. The Brazilian dialect also records two more recent changes: the erasing of indeterminate *se* in finite constructions and its insertion in infinite constructions.” (NUNES 1990: iv, my translation).

use of passive sentences emerges as the hallmark of this grammar. For this reason, Chart 2, to be analyzed, needs to be compared to the data in Chart 3 below. In it, the question of the irregularity of the amount of data per year is presented:



**Chart 3:** Amount of data collected in each database per year.

It is observed that the amount of data is not regular between the databases. The highest number of sentences collected is in the NURC, which presents data only from the period between 1975 and 1980, as well as the CEDAE. Likewise, the children's essays are all from the period 1981-1985. COMVEST data, in turn, represent the 1900s and 2000s, while only data from the Folha de SP could be collected in all periods analyzed.

This methodological question indicates that, although Chart 2 shows a decrease in the use of the three types of *se*, this decrease can also be explained by the smaller amount of data, mainly speech, from the 1990s.

Moving on to the exposition of the internal variables, we will first present those that are related to the internal argument since it is a much-discussed element in the literature on *se* constructions. 126 sentences had a null internal argument; therefore, we discarded these sentences to verify the variables studied below

A curious finding in relation to internal arguments is the preference for inanimate arguments. Table 2 shows this distribution taking as a set the sentences in which the internal argument was realized by the speaker:

Type of se		- Animated	+ Animated	Overall	%
Indeterminate	N	777	12	789	81.85%
	%	98.48%	1.52%		
Passive	N	68	3	71	7.37%
	%	95.77%	4.23%		
Sarney	N	103	1	104	10.79%
	%	99.04%	0.96%		
Overall	N	948	16	964	100.00%
	%	98.34%	1.66%		

**Table 2:** Interaction between internal argument and type of interaction.

The interaction between these variables was also confirmed by logistic regression (p. value <0.04607).

Although themes are not always animated, only 1.67% of the total sentences have an internal [+animated] argument. Thus, sentences like (19) are much more frequent than (20):

(19) *Se cortava o bambu, tirava fatias de bambu.* (NURC)

‘*Se cut(past) the bamboo, took slices of bamboo.*’

One cut the bamboo, one took slices of bamboo.

(20) *Se contrata mulheres.* (NURC)

‘*Se hire women.*’

One hires women.

Still in relation to the internal argument, there was a recurrent preference in expressing it postposed to the verb, as shown in Table 3:

Type of <i>se</i>		Before the verb	After the verb	Overall	%
Indeterminate	N	245	538	783	81.13%
	%	31.49%	68.51%		
Passive	N	30	41	71	7.40%
	%	42.25%	57.75%		
Sarney	N	42	68	110	11.47%
	%	38.18%	61.82%		
Overall	N	317	647	964	100.00%
	%	33.06%	66.94%		

**Table 3:** Position of the internal argument in relation to the verb.

For the three types of *se*, the percentage of sentences like (21), where the argument is postposed to the verb exceeds 50%, while sentences in which the internal argument is postposed to the verb, as in (22), are approximately 40% of the total. This analysis was not applied to sentences in which the internal argument was not explicit or to sentences with unergative verbs such as (23):

(21) Já nem se fala mais na simples falsificação de cheques preenchidos.  
(FOLHA DE S. PAULO)

‘Already not *se* talk anymore about simple forgery of checks filled.’

One does not talk anymore about simple forgery of filled checks.

(22) Piada de Presidente da República não se faz. (FOLHA DE S. PAULO)

‘Joke about President of Republic not *se* make’

One does not make jokes about the President of the Republic.

(23) Do jeito que se vive, a gente vive correndo. (NURC)

‘In way that *se* live, people lives running.’

‘The way we live, we are always running.’

The interaction between variables was also confirmed by logistic regression (p. value < 0.05).

These data are very close to what was found by Nunes (1990). In his set of data referring to the twentieth century, the author verified an index of 84% of postposition of the internal argument in relation to the verb. Such percentage was useful to him to

understand the contexts in which there is lack of agreement between verb and internal argument since in 92% of the cases of disagreement he verified in the magazine *Veja* the sentences exhibited the internal argument postposed to the verb. To investigate whether the position of the internal argument interferes with agreement, Table 4 crosses these two factors for sentences in which the internal argument was made explicit by the speaker:

Internal argument		+Agreement	-Agreement	Overall	%
Before the verb	N	315	5	320	33.16%
	%	98.43%	1.57%		
After the verb	N	635	9	644	66.84%
	%	98.60%	1.40%		
Overall	N	950	14	964	100.00%
	%	98.54%	1.46%		

**Table 4:** Agreement with respect to the position of the internal argument.

There are isolated cases without agreement and therefore it is not possible to state safely whether the position of the internal argument is a determining factor, although the number of occurrences without agreement in sentences in which the argument is postposed to the verb is twice the number of sentences in which the argument is before the verb. Logistic regression results, on the other hand, showed interaction between variables ( $p$ . value < 0.01). One hypothesis to be considered to explain this interaction is the fact that the considerable preference for the agreement between verb and internal argument may indicate that the speakers interpret the internal argument as a subject.

It is interesting to observe, then, in which databases this disagreement appears more frequently, as revealed in table 5:

Agreement		CEDAE	Children's essays	COMVEST	NURC	Folha de São Paulo	Total	%
[+Agreement]	N	0	14	156	525	255	950	98.54%
	%	0.00%	1.48%	16.46%	55.17%	26.90%		
[-Agreement]	N	1	2	4	7	0	14	1.46%
	%	7.14%	14.29%	28.57%	50.00%	0.00%		
Total	N	1	16	160	530	255	964	100.00%
	%	0.10%	1.66%	16.63%	55.09%	26.51%		

**Table 5:** Agreement between verb and intern argument by database.

Although the amount of data in which there is a lack of agreement is very small, the increase in the percentage of sentences without agreement in the NURC data and their complete absence in the Folha de São Paulo data may indicate that sentences like (24), (25) and (26) tend to be less accepted by the speakers. If these forms were excluded from the written material that shows the prestige forms and even appear punctually in the schooling speech, it can be seen that this use does not have a good evaluation by the speakers. In addition, the results of the logistic regression indicated interaction among the variables (p. value < 0.05).

(24) Passou-se uns minutos e foi quando eu falei pra minha prima freiar.  
(Children's essays).

'Passed-se few minutes and was when I spoke to my cousin brake.'

A few minutes passed and that was when I told my cousin to brake.

(25) E os aniversários de casamento se comemora, não? (NURC)

'And the anniversaries of wedding se celebrate, don't?'

And wedding anniversaries are celebrated, are they not?

(26) só se faziam espetáculo dentro daquela linha. (NURC)

'only se make spectacles within that line.'

One only made spectacles within that line.

Considering all what has been discussed regarding the information about the internal arguments of the sentences with *se* and their characteristics, we will now see the particularities of the verbs observed in the data. When examining table 8 below,

which describes the behavior of verbs in relation to their transitivity with the sentences with each type of *se*, we can see that the categories of verbs that allow the use of *se* as a resource for the indeterminacy of the thematic role of the subject, have been expanded. The existence of indeterminate sentences accompanying direct transitive verbs, indirect transitive verbs, ditransitive, unergative, and state verbs goes against the prescriptions of Traditional Grammar, which generally prescribes that the use of indeterminate *se* is only restricted to the intransitive verb or the unaccompanied direct object, while the use of passive *se* must only be limited to direct transitive verbs.

Type of SE		Direct Transitive	Indirect Transitive	Ditransitive	Unaccusative	Unergative	State	Overall	%
Indeterminate	N	732	65	21	0	32	19	869	51.06%
	%	84.23%	7.48%	2.42%	0.00%	3.68%	2.19%		
Passive	N	138	2	6	0	6	6	158	9.28%
	%	87.34%	1.27%	3.80%	0.00%	3.80%	3.80%		
Sarney	N	510	50	20	0	65	30	675	39.66%
	%	75.56%	7.41%	2.96%	0.00%	9.63%	4.44%		
Overall	N	1380	117	47	0	103	55	1702	100.00%
	%	81.08%	6.87%	2.76%	0.00%	6.05%	3.23%		

**Table 8:** Transitivity of the verb in relation to the type of *se*.

Table 8 shows that the highest percentage of occurrences, for all types of *se*, appears in sentences with direct transitive verbs. The interaction between the variables was also confirmed by the logistic regression results (p.value < 0.01). The use of passive *se*, on the other hand, seems to follow the prescriptions of the norm since there are only two occurrences of these constructions with indirect transitive verbs:

(27) Então servem-se da água do rio. (NURC)

‘Then get(pl)-*se* from water of river.’

Then one gets the water of the river.

(28) Porém utilizam-se de humor. (NURC)

‘But use(pl)-*se* of humor.’

But one uses humor.

Therefore, although the linguistic changes involving indeterminate *se* and passive *se* constructions have widened their contexts of uses, there is still a pattern in the choice in direct transitive verbs. It is also noted that there is no unaccusative verb among the data since it is necessary that the verb should have a subject so that the thematic role is indeterminate.

Another factor that presents regularity is the verbal tense of the verbs that accompany *se*. We have chosen to analyze this variable, first, to determine whether the use of *se* Sarney prevails in the gerund form or in the infinitive form. Then, the observation was extended to all types of *se* in order to verify patterns of occurrence, however, logistic regression results did not indicate interaction between these variables (p. value = 0.9). For indeterminate *se* and passive *se*, the present form is by far the preferred one. The *se* Sarney, in turn, appears more in sentences with verbs in the infinitive. This can be seen in Table 9:

Type of <i>se</i>		Pres.	Past Perf.	Past Imperf.	Fut. Pret.	Inf.	Ger.	Past Subj.	Pres. Subj.	Fut. Subj.	Overall	%
Indet.	N	616	67	137	14	0	0	9	26	1	870	51.03%
	%	70.80%	7.70%	15.75%	1.61%	0.00%	0.00%	1.03%	2.99%	0.11%		
Pass.	N	128	12	16	0	0	0	0	4		160	9.38%
	%	80.00%	7.50%	10.00%	0.00%	0.00%	0.00%	0.00%	2.50%	0.00%		
Sar.	N	0	0	0	0	530	145	0	0	0	675	39.59%
	%	0.00%	0.00%	0.00%	0.00%	78.52%	21.48%	0.00%	0.00%	0.00%		
Overall	N	744	79	153	14	530	145	9	30	1	1705	100.00%
	%	43.64%	4.63%	8.97%	0.82%	31.09%	8.50%	0.53%	1.76%	0.06%		

**Table 9:** Tense verb per type of *se*.

Nunes (1990) already defined *se* Sarney as the one that accompanied sentences in the infinitive, the gerund form being only an extension of these cases. In fact, we observed that only in 22% of the sentences with this type of *se* the verb was in the gerund form. These results were expected, since the infinitive form first appeared and is the one that Nunes (1990) studied. The insertion of *se* with verbs in the gerund would be, then, only a process deriving from the insertion of *se* with verbs in the infinitive form.

Since the position of the clitics is also an element that diverges between the speech BP and the written BP, the frequency of proclisis, enclisis and mesocclisis was verified in order to investigate if there is a relationship between the elements

recommended by the school normativity. Table 10 shows the distribution of the clitic position in relation to the verb for each type of *se*:

Type of <i>se</i>		Proclisis	Enclisis	Mesocclisis	Overall	%
Indeterminate	N	533	336	1	870	51.03%
	%	61.26%	38.62%	0.11%		
Passive	N	82	78	0	160	9.38%
	%	51.25%	48.75%	0.00%		
Sarney	N	470	200	5	675	39.59%
	%	69.63%	29.63%	0.74%		
Overall	N	1085	614	6	1705	100.00%
	%	63.64%	36.01%	0.35%		

**Table 10:** Position of *se* per type of *se*.

It can be observed that the preference for proclisis is generalized to all types of *se*. Therefore, there is no interaction between the variables, as seen by the logistic regression results (p. value = 0.9). This preference is expected since BP prefers the proclitic position for clitics (Pagotto, 1993). However, the use of enclisis and mesocclisis is unrelated to the case of passive sentences.

In addition, the intriguing use of mesocclisis, a rare structure that, more than passives sentences, has become an exclusivity of written materials and instigates to verify the frequency of sentences with *se* in each of the three positions per database. Table 11 represents these occurrences:

Type of <i>se</i>		Proclisis	Enclisis	Mesoclisiss	Overall	%
CEDAE	N	20	4	0	24	1.54%
	%	83.33%	16.67%	0.00%		
CHILDREN'S ESSAYS	N	36	24	0	60	3.85%
	%	60.00%	40.00%	0.00%		
COMVEST	N	58	202	1	261	16.76%
	%	22.22%	77.39%	0.38%		
NURC	N	525	230	0	755	16.76%
	%	201.15%	88.12%	0.00%		
FOLHA DE SP	N	72	384	1	457	16.76%
	%	27.59%	147.13%	0.38%		
Overall	N	711	844	2	1557	100.00%
	%	45.66%	54.21%	0.13%		

**Table 11:** Position of *se* per database.

In fact, the hypothesis above is confirmed for the mesoclisiss: the two occurrences, transcribed below, were observed in written environments:

(29) Então, no caso, far-se-ia uma farsa. (FOLHA DE S. PAULO)

‘Then, in that case, do-*se*-conditional suffix a charade.’

Then, in that case, one would do a charade.

(30) Poder-se-ia facilmente ficar impressionado. (COMVEST)

‘Can (conditional)-*se*- conditional suffix easily be impressed.’

One could easily be impressed.

The data also indicate that the enclitic form is linked to schooling. In the acquisition data, both uses, before and after the verb, are in sentences showed below by a childish song, or by children's literature:

(31) Dona chica ca ca admirou se se<sup>9</sup>. (CEDAE)

‘Mrs. Chica wondered *se*(*reflexive*).

Mrs. Chica wondered.

(32) Cabou-se a história. (CEDAE)

‘Ended-*se* the story.’

The story is ended.

<sup>9</sup> The repetition of the syllable “ca” and of the clitic “se” comes from the music to which the child refers.

It is only in the first years of school that this type of structure arises, even though the child has not learned to segment the words according to what the writing standard predicts:

(33) Lavou as mãos e foi se embora. (Children's essays)

'Washed the hands and went *se* out.'

He washed his hands and went out.

(34) Pitacuça sentouse<sup>10</sup> ao lado de um velho. (Children's essays)

'Pitacuça sitse next to an old man.'

Pitacuça sits next to an old man.'

(35) ...da seguinte maneira: dobrase<sup>11</sup> uma ponta assim. (Children's essays)

'...in the following way: foldse a tip like this.'

....as follows: on folds a tip like this.

However, at the end of schooling, the percentages referring to the use of enclisis increase and, in written materials, their occurrence practically doubles in relation to the use of proclisis.

### 3. DISCUSSION

Given the above, the data found seem to confirm what Nunes (1990) stated:

O surgimento da passiva pronominal sem concordância, apontado por Naro, desencadeou um processo de mudança linguística no português brasileiro em que a forma conservadora (com concordância) foi paulatinamente perdendo terreno para sua concorrente, culminando no momento sincrônico, em que deve sua sobrevivência na modalidade escrita à renitência da gramática normativa. (NUNES 1990: 77)<sup>12</sup>

Effectively, it was seen that the indeterminating form took space in the language and seems to be peripheral in relation to the passive form regardless the material of the sentence.

---

<sup>10</sup> The standard segmentation in BP would be "sentou-se", however, because it is written essays by children in the early stages of writing learning, this segmentation has not yet been acquired.

<sup>11</sup> The standard segmentation here would be "dobra-se".

<sup>12</sup> The emergence of the pronominal passive sentence without agreement, as pointed out by Naro, triggered a process of linguistic change in Brazilian Portuguese in which the conservative form (with agreement) gradually lost ground to its competitor, culminating in the synchronous moment, in which its survival in which it survives in the written modality, to the reluctance of normative grammar. (NUNES 1990: 77, my translation).

On the other hand, it is possible to make some caveats. Although the logistic regression tests did not indicate interaction between type of *se* and database (p. value = 0.9), passive *se* appeared in very similar frequencies when compared to speech and writing data. This seems to indicate that its use still goes beyond the limits of writing. Its frequency was not expected to be so high in 21st-century school essays. There are 33% of the sentences with the passive form in these essays, compared to 27% in the NURC, as shown in table 1. This data may indicate that the vernacular variety of the 1970s was closer to the norm than the current one, although there are still no studies on the behavior of these sentences so that a comparative study ought to be carried out to verify if there is in fact a change in progress.

In addition, we verified that one of the changes that Nunes (1990) pointed out occurred massively: the insertion of *se* in sentences with infinitive and gerundive verbs or the increase of the use of *Se Sarney*, mainly in speech. This form was also inserted in written contexts and appears with a frequency close to the undetermined form of the clitic in both the COMVEST data and the Folha de São Paulo data. As for the suppression of the clitic in finite sentences verified by Nunes (1990), no data were collected singly, but this form appears competing with the presence of the clitic, as will be described below.

The clitic *se* behaved then in a regular way in the analyzed sentences. Structures with indeterminate *se* and passive *se* usually presented the same contexts: internal argument [-animated] postposed to the verb and with agreement, direct transitive verbs in the present and *se* in proclisis position.

Finally, the main hypothesis of our work was confirmed: the acquisition of indeterminate *se* and passive *se* depends on schooling. The data indicate that, in contrast to the reflexive, pronominal and reciprocal *se*, indeterminate and passive clitics are not naturally acquired by BP speakers. their appearance occurs first in written material and only near the end of the schooling their appearance in the speech is verified. In addition, the statistical tests showed interaction between the appearance of *se* and the variables age group and type of record. In representative databases of speech in the period of acquisition and the first years of school, their presence is inexpressive, taking about 12 years of schooling to become significant.

It is also possible to state that after acquisition, this knowledge resembles a foreign language, as Kato, Correa & Cyrino (2009) propose. Literate speakers alternate between naturally acquired forms of indetermination, such as the use of

generic expressions or the deletion of *se*, and the use of *se* as a code-switching process. Sentences like (19) and (23) prove this code-switching. While in (19), the speaker alternates between the use and suppression of indeterminate *se*, the speaker who enunciated (23) also uses the generic expression “*a gente*” to mark the indetermination of the subject.

(19) Se cortava o bambu, tirava fatias de bambu. (NURC)

‘*Se* cut(past) the bamboo, took slices of bamboo.’

One cuts bamboo, one took slices of bamboo.’

(23) Do jeito que se vive, a gente vive correndo. (NURC)

‘In way that *se* live, people lives running.’

The way we live, we are always running.

It is therefore argued that the internal knowledge of a Brazilian Portuguese schooled speaker involves two grammars, such as Kato (2005a) proposed: an internal naturally acquired grammar with innovations, and a peripheral acquired at school, conservative and ruled by the norms of Traditional Grammar. The first would contain other forms of indeterminacy and the second the use of indeterminate and passive *se*. This last grammar is triggered in situations of monitored speech or writing.

It is also possible to conclude with the considerations of Kato (1996) on the loss of the third person clitics in PB:

O processo de letramento, seja através de dados positivos encontrados na imersão espontânea do aprendiz em textos escritos, seja através de dados negativos (intervenção corretiva da escola), apresenta, nos dados externos dos aprendizes (sua língua-E), uma evidência quantitativa do sucesso escolar na recuperação dos fósseis, pelo menos na língua escrita. Contudo, a análise qualitativa revela indícios de que o processo difere muito da aquisição natural, pois esses dados apresentam inúmeros casos de hipercorreção e estratégias de esquiva, evidências de que há um comportamento consciente de monitoração do produto. Além disso, (...) há um conflito entre a Língua-I e o conhecimento meta-cognitivo do que é bem aceito socialmente (KATO 1996: 25).<sup>13</sup>

---

<sup>13</sup> The process of literacy, whether through positive data found in the spontaneous immersion of the learner in written texts, or through negative data (corrective intervention of the school), presents, in the external data of the learners (their E-language), quantitative evidence of the success in the recovery of fossils, at least in written language. However, the qualitative analysis reveals evidence that the process differs a lot from natural acquisition, since these data present numerous cases of hypercorrection and avoidance strategies, evidences showing that there is a conscious behavior of product monitoring. In addition, (...) there is a conflict between I-Language and meta-cognitive knowledge of what is socially well accepted (KATO 1996: 25, my translation).

#### 4. CONCLUSION

In the discussion above, as Galves (1987) stated, *se* will still run a lot of ink. It would be interesting, then, to complement this research, to collect more recent data of both speech and writing in order to verify if a change is in progress in BP. The area also lacks studies comparing clitic acquisition with other forms of indeterminacy to more accurately trace the behavior of competition between these forms. In addition, research into which elements make up the nuclear grammar and the peripheral grammar of the BP speaker has become a very fertile field of research as well as the ways in which the phenomenon of code-switching occurs in different situations, since every day the language learned in school and the language acquired at school are distancing themselves.

#### REFERENCES

- BIRDSONG, David. Ultimate attainment in second language acquisition. *Language*, v. 68. 1992.
- CHOMSKY, Noam. *Lectures on Government and Binding*. Dordrecht: Foris. 1981.
- CHOMSKY, Noam. *Knowledge of Language: It's nature, origin and use*. Praeger, New York. 1986.
- EVERETT, Daniel. On romance "se". Dissertação de Mestrado. Campinas. Unicamp. 1985.
- GALVES, Charlotte. A sintaxe do português brasileiro. *Ensaaios de Linguística*. Belo Horizonte, v. 13. 1987.
- KATO, Mary Aizawa. *No Mundo da Escrita*. São Paulo: Editora Ática. 1986.
- KATO, Mary Aizawa. Português Brasileiro falado: aquisição em contexto de mudança lingüística. In: DUARTE, Maria Eugenia; LEIRIA, Isabel. (eds). *Actas do Congresso Internacional sobre o Português*. Vol II, 1996.
- KATO, Mary Aizawa. Variação e mudança no Português Brasileiro: problemas de aquisição. *A Cor das Letras* (UEFS). Feira de Santana, n.3, 1999.
- KATO, Mary Aizawa. The permeability of grammars in bilingual language acquisition. In: *Bilingualism: Language and Cognition*, v. 4, n. 1, 2001.
- KATO, Mary Aizawa. Desvendando a gramática do português brasileiro. In: ALBANO, Eleonora; COUDRY, Maria; POSSENTI, Sirio; ALKMIN, Tânia (orgs.). *Saudades da Língua*. Campinas: Mercado de Letras, 2003.
- KATO, Mary Aizawa. A gramática do letrado: questões para a teoria gramatical. In: MARQUES, Mary Aizawa.; TEIXEIRA, José; LEMOS, Aida. (Orgs.). *Ciências da linguagem: trinta anos de investigação e ensino*. Braga: Universidade do Minho. 2005a.
- KATO, Mary Aizawa. A contribuição chomskiana para a compreensão da aprendizagem de L2. *Trabalhos em Lingüística Aplicada*. Campinas, v. 44, n.2, 2005b.

KATO, Mary Aizawa. A variação no domínio dos clíticos no português brasileiro. *Linguística*. Montevideo, v. 33, n. 1, 2017.

KATO, Mary Aizawa; CYRINO, Sonia Maria Lazzarini.; CORRÊA, Vilma Recha. Brazilian Portuguese and the recovery of lost clitics through schooling. In: PIRES, Acrísio; ROTHMAN, Jason. (eds.) *Minimalist Inquiries into Child and Adult Language Acquisition: Case Studies across Portuguese*. Berlin/New York: Mouton De Gruyter. 2009.

MEISEL, Jürgen. Revisiting universal grammar. *DELTA*, v. 16, n. especial. São Paulo: EDUC, 2000.

NUNES, Jairo. O Famigerado se: uma análise sincrônica e diacrônica das construções com se apassivador e indeterminador. Dissertação de Mestrado. Campinas. Unicamp. 1990.

PAGOTTO, Emílio. Clíticos, mudança e seleção natural. In: ROBERTS, Ian. & KATO, Mary Aizawa. *Português brasileiro: uma viagem diacrônica*. Campinas, Editora da Unicamp: 185-206. 1993.

POPLACK, Shana. Code-switching (linguistic). In: *International encyclopedia of the social and behavioral sciences*. Edição on-line. 2001. Disponível em: <<http://aix1.uottawa.ca/~sociolx/CS.pdf> > Acesso em 24-06-2015.

WHITE, Lydia. *Second Language Acquisition and Universal Grammar*. Cambridge: Cambridge University Press, 2003.

Submitted on June, 17, 2018.

Accepted on August, 28, 2018.