

Thematic Dossier

Diving Deeper into the Waters of “se” as a Clitic in Brazilian Portuguese

Mergulhando mais fundo nas águas do clítico “se” no português brasileiro

Magali Sanches Duran*

*Universidade de São Paulo (USP)
São Paulo | SP | BR
magali.duran@gmail.com
<https://orcid.org/0000-0002-3843-4600>

Adriana Silvina Pagano**

**Universidade Federal de Minas Gerais
(UFMG) | Belo Horizonte | MG | BR
apagano@ufmg.br
<https://orcid.org/0000-0002-3150-3503>

**Thiago Alexandre
Salgueiro Pardo***

*Universidade de São Paulo (USP)
São Paulo | SP | BR
tasparado@icmc.usp.br
<https://orcid.org/0000-0003-2111-1319>

Abstract: This article reports on a corpus-based study of the pronoun “se” in Brazilian Portuguese aimed at developing a proposal for annotating the different uses of the pronoun in corpora to be leveraged in Natural Language Processing (NLP). We analyzed 732 occurrences of “se” retrieved from Porttinari-base, a treebank annotated following the Universal Dependency guidelines. The results allowed us to develop a proposal for syntactic and morphosyntactic annotation, which is expected to make human annotation simpler and more consistent as well as aid supplementary annotation, as in thematic role annotation, and enhance the performance of NLP tasks, as in automatic extraction of information.

Keywords: clitic pronouns; Portuguese language; syntactic analysis; corpus annotation; impersonalization; pronominal verbs; inchoative alternation.

Resumo: Este artigo apresenta um estudo baseado em *corpus* do pronome “se” no português brasileiro, com o objetivo de desenvolver uma proposta de anotação dos diferentes usos deste pronome para fins de Processamento de Linguagem Natural (PLN). Analisamos 732 ocorrências de “se” extraídas do *corpus* Porttinari-base, um *treebank* anotado de acordo com as diretrizes das Universal Dependencies. Os resultados permitiram desenvolver uma proposta de anotação sintática e morfosintática que poderá tornar a anotação humana mais simples e consistente, auxiliar a anotação de papéis semânticos, e melhorar o desempenho de tarefas de PLN, como a extração automática de informações.

Palavras-chave: pronomes clíticos; língua portuguesa; análise sintática; anotação de *corpus*; impessoalização; verbos pronominais; alternância incoativa.

1 Introduction

The term Artificial Intelligence (AI) has become a major focus of attention in the past couple of years. Although different definitions can be found in the literature and in the mainstream media, practically all of them agree that AI refers to machines programmed to perform tasks that require human intelligence, aiming at supporting humans in a variety of ways. To do that, AI draws on the fact that machines have a distinct and significantly superior processing capacity compared to that of a human being, which may be put to the service of human society to deal with the enormous amount of data produced since the advent of the Web.

In the field of language studies and applications, AI techniques have been employed to create software to convert spoken language into written language (speech transcription) and written language into spoken language (speech synthesis), to translate input from one language into another (machine translation), to answer questions and summarize texts, among many others. In the last decade, machines have evolved with training for such tasks, and today their ability to process linguistic input surprises even staunch skeptics, especially when the latest advances in generative AI and large language models are considered.

It is important to remember that Language AI evolution has reached its present status after decades of computational modeling and processing of human languages, mainly in the area known as Natural Language Processing (NLP), experimenting with several different methods of knowledge acquisition. The input for such an endeavor is stored in textual data collections that constitute the corpus/corpora on which systems are developed, trained, and evaluated. With vast amounts of data and substantial processing power, the most modern machine learning methods can find patterns in the data and achieve good results without (or with limited) human intervention. However, there are scenarios in which the amount of data is small, the processing capacity is more modest, and/or the linguistic phenomena of interest are very sophisticated and not sufficiently understood and therefore not properly computationally modeled, resulting in AI systems with limited performance. In such cases, the strategy to develop a better AI system remains providing annotated data to the machine (covering the linguistic phenomenon of interest), and the annotated data is usually manually or semi-automatically produced by humans.

The practice of corpus annotation consists in enriching the linguistic input with labels that add a human interpretation to a given language segment (token, phrase, sentence, paragraph, text, etc.) or to the relationships between segments. There are several types of annotation that require linguistic analysis, with part-of-speech, syntactic functions and semantic roles being some of the most widespread. Nevertheless, there are annotations that are not linguistic in nature, such as classifying customer service comments into

the classes “praise”, “criticism” and “suggestion”, an annotation that is often made by the customers themselves.

At present, even when large language models trained with huge amounts of data achieve good performance in many tasks, corpora and databases curated by humans are still necessary. As AI models reflect the patterns existing in the data with which they were trained, the annotation of corpora with the greatest granularity possible and based on solid guidelines that adhere to the principles of responsible AI has become a fundamental task. This may optimize the results obtained based on these corpora and lead NLP to have a qualitative impact on machine learning models.

From the perspective of language studies, improving corpus annotation also constitutes a relevant task, especially due to the recognized potential of corpora to help develop and validate linguistic theories and improve our descriptions with an impact on several fields of study, such as historical investigations, characterization of language typologies, translation studies, and language teaching and learning, among many other possibilities.

Corpus linguistics has evolved throughout the past decades and is now an established theory and methodology for language studies and related fields. While the concept of corpus remains essentially the same – collection of texts in electronic format –, the characteristics of a corpus can vary in terms of the potential to leverage information from it. To the linguist, raw text corpora can offer, with minimal processing, valuable data about frequency and collocation of words. However, corpora enriched with linguistic information, as is the case of morpho-syntactically and semantically annotated corpora, allow more linguistically-informed queries, improving the quality and depth of the results.¹ For NLP purposes, annotated corpora have significant advantages over raw text corpus as well, as annotated corpora allow for training and development of models that will enable many of the tasks in natural language understanding and generation.

Several relevant annotated corpora of Brazilian Portuguese can be readily accessed and queried (cf. Mello *et al.*, 2022). When it comes to syntactically annotated corpora (called “treebanks” in NLP), there are nowadays a few of them which can be openly queried thanks to several available tools. This is the case of treebanks for Brazilian Portuguese annotated following the Universal Dependencies (UD) guidelines (Nivre *et al.*, 2020; Marneffe *et al.*, 2021). While the quality and the consistency of the annotations have greatly improved thanks to a project like UD, there is still room for enhancing treebanks with annotations of particular language constructions that are relevant to natural language understanding and generation in Portuguese.

In this paper, we tackle one such construction and attempt to develop guidelines to annotate the different functions of the pronoun “se” in Brazilian Portuguese with the greatest granularity possible. For NLP, distinguishing the functions of “se” is still a challenge and has an impact on several downstream computational tasks. For example, detecting whether the pronoun “se” is used reflexively or reciprocally allows extracting more accurate information, such as tracking *who* did *what to whom* in tasks of automatic information extraction from news

¹ Unannotated corpora do not allow more linguistically-informed queries, and many open source concordancing software and platforms do not provide linguistic pre-processing functionalities which could support, for instance, querying for part-of-speech and other linguistic information.

reports, clinical narratives, business texts, etc. Detecting when “se” is used in so-called synthetic passive and impersonal constructions is fundamental for tasks involving language generation (translation, summarization, etc.) and to create accurate knowledge graphs for fact-checking.

Several authors have already addressed the problem of the multiplicity of meanings constructed by the pronoun “se” (among the most recent, Cançado; Amaral, 2016; Martins; Nunes, 2016; Godoy; Pinheiro, 2023), some of which are specifically concerned with how to assign morphosyntactic and syntactic annotation tags to facilitate the automatic processing of “se” (Duran; Aluísio, 2011; Duran *et al.*, 2013; Souza; Freitas, 2023, among others). If part of speech tags provide good accuracy in sorting out the conjunction “se” from the pronoun “se” in a corpus, the same is not the case with regard to the different uses of the pronoun “se”. In fact, the first challenge is human annotation, which has to deal with the diversity of views of grammarians and linguists, as it is the case with the distinction between impersonalizing “se” and “se” in the so-called “synthetic passive”.

In seeking theoretical grounds for choosing a set of labels to annotate the pronoun “se”, in order to facilitate the automation of syntactic annotation and the annotation of semantic roles, we came across different proposals that agree on some points and diverge on others. By using a syntactically annotated corpus, we tested the adherence of the taxonomies to the occurrences of “se” in order to observe whether all the occurrences in the corpus fit into one of the classes or whether there were cases of difficult or dubious classification. The corpus used as the basis for the study was Porttinari-base (Duran *et al.*, 2023), a corpus of 8,418 sentences extracted from the Folha-Kaggle corpus and syntactically annotated according to the UD guidelines. The objective of this article is to report this experience and present a proposal for annotating “se” that is simpler and more suitable for NLP.

In Section 2, we briefly review the theoretical works on the functions of “se”, seeking support for the possible annotation labels for this pronoun. In Section 3, we outline the methodology used in the searches. In Section 4, we present our analysis, illustrating it with examples extracted from the corpus, and, in Section 5, we outline our conclusions and directions for future work.

2 State-of-the-Art on Linguistic Analysis of “se”

The study on “se” in Brazilian Portuguese dates back to some of the first grammatical accounts of the language, with several grammarians and linguists having attempted different taxonomies to classify this pronoun. More recent studies have reviewed early approaches in the light of changes in language use from those times to present-day Brazilian Portuguese. Among recent studies, Cançado and Amaral (2016) draw on Nunes (1990, 1995) and Godoy (2012) to propose a taxonomy of nine different uses of “se”, which the authors categorize based on particular criteria and explain through a series of distinctive probes. These uses and the examples given by the authors are:

- (a) Reflexive use – instances in which “se” is a verb argument and has a thematic role. As such, it cannot be omitted and can be probed by adding “ela/ele mesma/o” (“herself”, “himself”, “itself”) or “elas/eles mesmas/os” (“themselves”). In sentences with reflexive “se”, the entity in the subject may be affected in part or as a whole. Hence, “A menina se

lavou” (“The girl washed herself”) may be interpreted as the girl having washed all her body or part of it.

Example: “A menina se lavou.” (“The girl washed herself”).²

- (b) Reciprocal use – instances in which a noun phrase realizes a plural subject and the action can be interpreted as having reciprocal impact among the individuals involved. A typical probe is adding “um(a)s a(o)s outra(o)s” (“each other”, “one another”).
Example: “As meninas se cumprimentaram” (“The girls greeted each other”).
- (c) Middle “se” – instances in which “se” might retain reflexivity but is not a verb argument, nor does it fulfill a thematic role. “Se” is omitted for some verbs in some dialects and is typically associated with verbs of movement. Unlike sentences with reflexive “se”, entities in subjects are affected as a whole entity and cannot be affected in part. Also, unlike reflexive “se”, they cannot be probed by substituting “ela/ele mesma/o” (“herself”, “himself”, “itself”) or “elas/eles mesmas/os” (“themselves”).
Example: “A menina se sentou” (“The girl sat down”).
- (d) Inchoative “se” – instances in which “se” operates an alternation from a causative-transitive to an inchoative-intransitive construction. In these instances, there is alternation from object to subject position. A probe to distinguish these instances from reflexive and middle ones is to add an adjunct for purpose, which cancels out an inchoative reading.
Example: “A menina (se) machucou”. (“The girl got hurt”).
- (e) Medial “se” – instances in which “se” occurs with present tense verbs and construes a generic meaning whereby the subject is assigned an abstract property. A probe for these instances is also to add an adjunct for purpose, which cancels out a medial reading.
Example: “Meninas se machucam facilmente” (“Girls get hurt easily”).
- (f) Synthetic passive “se” – instances in which “se” occurs in passive constructions of the synthetic type, with direct transitive verbs, whereby “se” is an argument of the verb and agent of the passive construction. “Se” occurs in a postverbal position and construes a meaning of indeterminacy, which is reinforced by the fact that number agreement between the verb and the postverbal subject is not usually observed outside of the standard language. Unlike inchoative uses, synthetic passive “se” admits an adjunct of purpose; however, no agent introduced by the preposition “por” (“by”) can be added.
Example: “Machucou-se a menina” (“The girl was hurt”).
- (g) Indeterminate “se” – instances in which “se” construes a meaning of indeterminacy in constructions with indirect transitive verbs. “Se” is a verb argument and has a thematic role, since it cannot be omitted and can be paraphrased by a sentence having an indefinite pronoun as subject.
Example: “Fala-se da menina por aí” (“People talk about the girl”).
- (h) Inherent “se” – instances in which “se” is not a verb argument and does not fulfill any thematic role. Verbs in these instances are referred to as pronominal verbs and the

² All examples in this article were translated by the authors.

clitic “se” is in some cases optional. Sentences with inherent “se” are not derived from transitive sentences, which distinguishes them from medial ones.

Example: “Amenina (se) arrependeu” (“The girl felt apologetic”; “The girl felt remorseful”).

- (i) Emphatic “se” – instances in which “se” has no syntactic function, though its omission construes a different meaning.

Example: “Lá (se) vai a menina” (“There goes the girl”).

Unlike Cançado and Amaral (2016), Martins and Nunes (2016) argue that criteria for distinguishing synthetic passive “se” from impersonal “se”, such as subject-verb agreement and subject position, have practically become useless in present Brazilian Portuguese, with restrictions on agreement and position being observed only in more formal, written language.

Drawing on a different theoretical framework, namely Usage-based Construction Grammar, Godoy and Pinheiro (2023) propose a taxonomy of six main constructions using “se” in Brazilian Portuguese, some of them overlapping with instances pointed out by Cançado and Amaral (2016) as well as making use of similar probes. These are:

- a) Reflexive “se” constructions have roles such as agent/patient or experiencer/theme fulfilled by the same referent. These roles implicate animacy and volition. Adding “ela/ele mesma/o” (“herself”, “himself”, “itself”) or “elas/eles mesmas/os” (“themselves”) is a useful probe in this case. Examples of verbs that are typically used in reflexive “se” constructions are “aceitar-se”, “cuidar-se”, “julgar-se”, “matar-se”, “perceber-se”, “prejudicar-se”, “reconhecer-se” (“to accept oneself”, “to take care of oneself”, “to consider oneself”, “to kill oneself”, “to perceive oneself”, “to harm oneself”, “to recognise oneself”);
- b) Reciprocal “se” constructions can be probed by adding “um a/com/de o outro” (“each other”, “one another”). Examples of verbs that are typically used in reciprocal “se” constructions are “abraçar-se”, “amar-se”, “casar-se”, “conhecer-se”, “juntar-se”, “parecer-se”, “reunir-se”, “separar-se”, “unir-se” (“to hug each other”, “to love each other”, “to marry each other”, “to know each other”, “to meet”, “to resemble each other”, “to gather”, “to split up”, “to join”);
- c) Dynamic middle constructions are named after Câmara Jr. (1972) and are characterized as having the same referent for agent and affected. There is an internally-driven movement within the agent, which is also affected by it. Examples of verbs that are typically used in these constructions are “alimentar-se”, “aproximar-se”, “casar-se”, “deitar-se”, “omitir-se”, “pronunciar-se”, “manifestar-se” (“to eat”, “to get closer”, “to get married”, “to lie down”, “to refrain”, “to speak out”, “to express oneself”);
- d) Cognitive middle constructions implicate mental verbs construing perception and emotion and have a subject that is both experiencer and affected. Animacy is a required feature, even though, unlike dynamic middle constructions, there is no agentivity. Examples of verbs that are typically used in these constructions are “acalmar-se”, “concentrar-se”, “decepcionar-se”, “esquecer-se”, “inspirar-se”, “sentir-se” (“to calm down”, “to concentrate”, “to be disappointed”, “to forget”, “to be inspired”, “to feel”);

- e) Inchoative middle constructions, also called anti-causative, resultative, ergative or simply inchoative constructions, implicate verbs that construe meanings of change and impact. The subject is fulfilled by the affected, but no agent is expressed. Examples of verbs that are typically used in these constructions are “acidentar-se”, “machucar-se”, “salvar-se”, “tornar-se”, “transformar-se” (“to have an accident”, “to injure oneself”, “to save oneself”, “to become”, “to transform oneself”);
- f) Impersonal constructions are constructions where the subject is neither agent nor experiencer nor affected, that is, the subject does not act, experience, suffer or is affected by any event. Most constructions of this type implicate indefinite and inanimate pronouns in subject position. Examples of verbs that are typically used in these constructions are “dever-se”; “explicar-se”, “encontrar-se”, “traduzir-se”, “refletir-se” (“to be due to”, “to explain oneself”, “to find oneself”, “to translate into”, “to be reflected in”).

Upon examining the taxonomies above, we can notice some common criteria typically used to distinguish between the different uses of “se”. These criteria include subject-verb agreement, identification of verb arguments and thematic roles, clitic position with respect to the verb, and alternations with other constructions.

Authors also point out multiple interpretations for a single sentence, as is the case of “João se jogou na piscina” (“João threw himself into the pool”), which, according to Godoy and Pinheiro (2023, p. 138), gives rise to two possible readings. In a reflexive reading, “João” would activate an external mechanism to propel himself into the pool. Hence, Agent and Patient would be temporally apart roles. However, in a more typical reading, which would be a dynamic middle reading, “João” jumped into the pool, the movement being triggered internally by “João”, who is at the same time Agent and Patient.

Hence, interpretation has a significant impact on how “se” can be classified in a given sentence. However, the criteria for distinguishing middle voice as suggested by Cançado and Amaral (2016) includes the possibility of clitic dropping, which does not seem to be the case in the middle voice example provided by Godoy and Pinheiro (2023), since “João jogar na piscina” would be missing information (“Jogou quem?”, “Jogou o quê?”) (“Who was thrown?”, “What was thrown?”).

Therefore, aiming at a more consistent corpus annotation for prospective use in NLP tasks, we decided to explore an annotated treebank in order to gather insights as to the different constructions found and to what extent they could be accounted for by the taxonomies in the literature.

3 Methodology

We queried the corpus Porttinari-base³ (Duran *et al.*, 2023), which comprises 8,418 sentences from Folha-Kagle, a dataset of news articles from the newspaper *Folha de S. Paulo*. Porttinari-base is annotated with PoS tags, morphosyntactic categories and dependency relations. All

³ POETISA. *Porttinari 2.0*. Brazil: POeTiSA, [20--]. Available at: <https://sites.google.com/icmc.usp.br/poetisa/porttinari>. Accessed on: 23 Sept. 2025.

annotations follow the guidelines of the project Universal Dependencies (Marneffe *et al.*, 2021), a framework for cross-linguistically consistent treebank annotation aiming to facilitate multilingual natural language processing and enable comparative linguistic studies. This framework has so far been applied to 179 languages.

We uploaded the corpus to Arborator-Grew⁴ (Guibon *et al.*, 2020), a free-access dedicated tool to edit and query annotated treebanks following UD guidelines.

Our queries for the form “se” yielded 1,008 occurrences in Porttinari-base, which were filtered out according to their PoS tags. We thus obtained the following results in Table 1.

Table 1 – Frequency of occurrence of PoS tags for the form “se” in the corpus

PoS Tag	Number of occurrences	Percentage
PRON	732	72.6
SCONJ	274	27.2
PROPON	002	0.2
Total	1008	100

Source: the authors.

As we can see, the large majority of occurrences are “se” as pronouns (PRON), whereas the subordinating conjunction (SCONJ) “se” (“if”) is far less frequent. Very few instances of “se” are tagged PROPON (proper noun) as “se” is spelled with an initial capital letter and is part of a book title.

We further queried the occurrences of “se” as a pronoun to examine its position in the sentence in relation to the verb. Table 2 shows the results we obtained.

Table 2 – Frequency of position of “se” with regard to the verb

Clitic placement	Number of occurrences	Percentage
Proclisis (preverbal)	589	80.5
Enclisis (post-verbal)	143	19.5
Total	732	100

Source: the authors.

Drawing on Cançado and Amaral (2016)’s criteria for distinguishing uses of “se”, we developed heuristics to query the corpus using regular expressions which combined syntactic and morphosyntactic functions. As our corpus is a treebank annotated for dependency syntax, we explored dependency relations which can be mapped to particular functions of “se” in the corpus sentences.

- ♦ To find occurrences of “se” as a direct object, we used the regular expression: pattern {Y [upos = PRON, lemma=“se”]; X -[obj]-> Y}, which retrieves instances of verbs having the pronoun “se” in a dependency relation tagged as obj. This query yielded 23 occurrences.

⁴ ARBORATOR-GREW. France: [s. n.], c2024. Available at: <https://arborator.grew.fr>. Accessed on: 23 Sept. 2025.

- ◆ To find occurrences of “se” as an indirect object, we used the regular expression: pattern {Y [upos = PRON, lemma=“se”]; X -[iobj]-> Y}, which retrieves instances of verbs having the pronoun “se” in a dependency relation tagged as iobj. This query yielded 5 occurrences.
- ◆ To find occurrences of “se” as an indeterminate subject, we used the regular expression: pattern {Y [upos = PRON, lemma=“se”]; X[Person=3, Number=Sing]; X -[expl]-> Y} without {X -[nsubj]|nsubj:pass|csubj]-> S} without {X -[obj]-> O}, which retrieves third-person singular verbs having the pronoun “se” in a dependency relation tagged as expl (expletives) and no dependency relation tagged as nsubj (subject) or obj (object). This query yielded 175 occurrences.
- ◆ To find occurrences of “se” in synthetic passives, we used the regular expression: pattern {Y [upos = PRON, lemma=“se”]; X[Person=3]; X -[expl]-> Y; X -[nsubj]|nsubj:pass|csubj|obj]-> S; S>>X}, which retrieves instances of third-person verbs having the pronoun “se” in a dependency relation tagged as expl (expletives) and a dependency relation tagged as nsubj (subject) or obj (object) to the right of the verb. This query yielded 63 occurrences, out of which 8 had a plural verb form and 55 a singular verb form.
- ◆ To find occurrences of “se” in reciprocal constructions, we used the regular expression: pattern {Y [upos = PRON, lemma=“se”]; X[Person=3, Number=Plur]; X -[expl|obj|iobj]-> Y}, which retrieves plural verb forms having the pronoun “se” in a dependency relation tagged as obj (direct object), or iobj (indirect object), or expl (expletive). This query yielded 106 occurrences.
- ◆ To find occurrences of “se” having no argument function, we used the regular expression: pattern {Y [upos = PRON, lemma=“se”]; X[Person=3]; X -[expl]-> Y; X -[nsubj]|csubj]-> S; S<<X}, which retrieved verbs having the pronoun “se” in a dependency relation tagged as expl (expletive) and a nsubj (subject) dependency relation to the left. This query yielded 316 occurrences.

As in any query with regular expressions, not all the sentences matching the expressions are expected to fully meet the occurrences of “se” targeted in our searches. This required manual inspection of each sentence and verifying if the targeted pattern was actually matched.

Building upon the results of our queries, we performed a preliminary clustering of the retrieved sentences leveraging their syntactic characteristics, to be explained in Section 4.

4 Results

We will first report results from our analysis of *argumental* “se” (when “se” plays the role of an argument of the verb) (4.1). Subsequently, we will focus on *non-argumental* “se”: used in impersonal constructions, which are discussed in grammars under the headings of “se” as “index of subject indeterminacy” (for intransitive and indirect transitive verbs) and “passive particle” (for direct transitive verbs) (4.2); and used in pronominal verbs (4.3). Finally, we will present an annotation proposal (4.4) and discuss insights from our analysis and implications for annotating corpora (4.5). A free translation into English for each example in Brazilian Portuguese can be found in the Appendix.

4.1 Argumental “se”

When analyzing a construction having a clitic “se”, an annotator first needs to decide whether “se” is a verb argument or not. If a cliticized verb has a subject that does not trigger the event denoted by the verb, i.e., a subject that has no volition over the verb’s action, this subject cannot fulfill the role of Agent; hence, “se” cannot be a verb argument. For “se” to be argumental, there has to be an Agent.

In examples (1), (2) and (3), the subjects “política” (“politics”), “mulher” (“woman”) and “oportunidade” (“opportunity”) do not fulfill the role of Agents and hence “se” cannot be argumental. It is worth noting that the feature of animacy is not necessarily related to agentivity: in (2), for instance, the noun “mulher” is animate, but is neither Agent nor Cause; it is an Undergoer of the verb “tornar-se” (“to become”).

- (1) A política agora **se parece** mais com o que era antes da crise.
- (2) A mulher **se torna** humana ao mesmo tempo em que é objetificada.
- (3) Quando começa a mexer em três setores, essa oportunidade **se perde** em função da falta de estrutura.

When the subject triggers the event denoted by the cliticized verb, it fulfills the role of Agent⁵ and, in that case, “se” may or may not be argumental. “Se” is argumental when the entity in subject position occupies also the object position, that is, it is at the same time Agent and Patient of the event. The replacement of the entity that occupies the object position is used as a probe in (4a), (5a) and (6a), which attest to the verbs allowing for the argument.

- (4) Em queda na pesquisa, Doria pode **se atropelar** na ambição e fracassar.
 - (4a) Em queda na pesquisa, Doria pode **atropelar** João com sua ambição.
- (5) As garotinhas podem **se perguntar** agora: eu quero ser cabeleireira ou chanceler?
 - (5a) As garotinhas podem **perguntar** a Maria agora: você quer ser cabeleireira ou chanceler?

⁵ As far as we can see, Causes are not subjects in cliticized verbs.

- (6) Rocque diz que ainda **se considera** neoliberal, mas **se classifica** como “ex-banqueira e ex-economista”.
- (6a) Rocque diz que ainda **considera** João neoliberal, mas **o classifica** como “ex-banqueiro e ex-economista”.

However, if the person affected by the action is the Agent of that action, but “se” cannot be substituted by a pronoun in another person, “se” is non-argumental. This can be seen in (7) and in its ungrammatical rewriting (7a). However, in instances as in (8) and (8a), as the verb “reunir” has a transitive meaning (“to gather”) and a pronominal meaning (“to meet”) in Portuguese, a different meaning is construed in (8a), but the sentence is not ungrammatical. Besides being non-argumental, these constructions remain reflexive: a shift in person for the Agent entails a shift in person for the reflexive pronoun, as (7b) and (8b) show.

- (7) O mercado **se apeg**a a isso e vê que está muito complicado operar, afirma.
- (7a) *O mercado **os apeg**a a isso e vemos que está muito complicado operar, afirma.
- (7b) Nós **nos apeg**amos a isso e vemos que está muito complicado operar, afirma.
- (8) Sem JK, Frente Ampla **se reún**e no Rio para planejar estratégia.
- (8a) Sem JK, Frente Ampla **reúne-os**⁶ no Rio para planejar estratégia.
- (8b) Sem JK, nós **nos reun**imos no Rio para planejar estratégia.

Nonetheless, some instances may have multiple interpretations:

- (9) A narrativa **se concentra** em alguns de seus filmes e despreza outros.
- (10) O Brasil é um país que **se define** pela copulação incessante dos contrários.

In (9), “narrativa” (“narrative”) is the subject of “concentrar-se” (“to focus”) as well as “desprezar” (“to dismiss”), the two verbs being coordinated. Hence, “narrativa” (“narrative”) can be said to have volition, as “desprezar” (“to dismiss”) is an action that involves an Agent. Strictly speaking, only animate nouns can be Agents, even though animate nouns are not always necessarily Agents. However, in instances like (9), inanimate nouns are used in constructions which assign animacy to them.

In (10), “Brasil” (“Brazil”) may be interpreted as Agent (the one who “define” (“defines”)) and “se” as a verb argument (direct object). However, “Brasil” (“Brazil”) may also be interpreted as Patient, implicating a non-argumental “se”. In this particular case, the preposition “por” (“by”), typically used to introduce Agents and Causes, allows us to interpret (11):

- (11) O Brasil é um país **definido** pela copulação incessante dos contrários.

However, if a preposition other than “por” (“by”) were used, “Brasil” (“Brazil”) could be more plausibly interpreted as Agent, as shown in (12) with the preposition “como” (“as”):

⁶ “Reunir” may occur as a direct transitive verb (“to gather”, “to recruit”) and as a pronominal indirect transitive verb (“to meet somebody”). (8b) construes a different meaning from (8).

- (12) O Brasil é um país que **se define** como uma copulação incessante de contrários.

These different readings of the same sentence show how slippery the interpretation of “se” is.

4.1.1 Reciprocity in Focus

Argumental “se” is always reflexive and, when the verb is in plural form, it may be reciprocal or not. Reciprocity is more readily interpreted in the case of verbs such as “abraçar” (“to hug”), “beijar” (“to kiss”), “enfrentar” (“to face”), “cumprimentar” (“to greet”), whereas in other instances a reciprocal reading will be prompted by adding “um ao outro” (“each other”, “one another”) or “mutuamente” (“mutually”), for instance in “acusar” (“to accuse”) and “culpar” (“to blame”).

“Se” as a reciprocal direct object:

- (13) Os terceiros colocados **se enfrentam** por uma chance na repescagem.

“Se” as a reciprocal indirect object:

- (14) Na época, quatro cooperativas **se deram** as mãos para criar a Fecoagro.

In Section 4.3 we will see examples of reciprocal meanings in non-argumental “se”.

4.1.2 Further Complements in “se” Constructions

Argumental “se” may or may not be the sole internal argument of a verb. Further complements may be present, such as a direct object realized by a noun phrase, an indirect object realized by a prepositional noun phrase, a direct objective clause, an indirect objective clause or a predicative. These are illustrated by the following examples:

- (15) Ao todo, 32 petroleiras **se inscreveram** *para a disputa*.
- (16) Está **se perguntando** *por que o tratamento fiscal desse produto é diferente dos demais?*
- (17) É quase a mesma proposta do A Bola que Pariu, feito por torcedoras que **se declaram** *fanáticas* e escrevem sobre seus próprios clubes.
- (18) Ainda que seu canto fosse as dores da matriarca, Sangaré conta ter percebido que as mulheres que a ouviam também **se viam** *decifradas* nele.

4.1.3 Summing Up Argumental “se”

As stated above, the first decision to be made by human annotators is distinguishing between argumental “se” and non-argumental “se”. As we have shown, argumental “se” may have two syntactic functions: direct object and indirect object, each of which may be reflexive or reciprocal. In terms of arguments, argumental “se” constructions may have a single argument or more than one. An additional argument may be a phrasal or clausal direct object, a phrasal or clausal indirect object or a predicative. Table 3 displays a summary of possible configurations of argumental “se”, that is, the syntactic contexts in which it occurs. An “x”

indicates co-occurring arguments and the word “se” fills in the spaces corresponding to its function in the respective example.

Table 3 – Configurations of argumental “se”

ARGUMENTAL "SE"									
Subject	Direct Object	Indirect Object	Predicative	Clausal Direct Object	Clausal Indirect Object	Non Reciprocal Example	Translation	Reciprocal Example	Translation
X	se					Ele se puniu.	He punished himself.	Eles se abraçaram.	They hugged each other.
X	se		X			Ele se considera fraco. Ele se viu desprezado.	He considers himself weak. He felt despised.		
X	se	X				Ele se responsabilizou pela falta.	He took responsibility for the mistake.		
X	se				X	Ele se forçou a comer.	He forced himself to eat.		
X	X	se				Ele se reservou o direito de faltar.	He reserved the right to be absent.	Eles se deram as mãos.	They held hands.
X		se		X		Ele se permitiu fazer uma pausa.	He allowed himself to take a break.		

Source: the authors.

Note: X stands for realized arguments besides that of “se”. Examples illustrate configurations.

4.2 Non-Argumental “se” as an Impersonalization Strategy

There are different strategies for impersonalizing a construction in Portuguese, using the clitic “se” or not. Constructions with “se” impersonalization are not reflexive, as they allow “se” only and do not admit other pronouns such as first-person singular “me” (“me”) and first-person plural “nos” (“us”).

Cançado and Amaral (2016) argue for impersonalizing “se” (index of subject indeterminacy and passive particle) to be considered an argument, as it fills in the slot for an Agent. We, however, consider impersonalizing “se” as implicating the absence of Agent, since “se” cannot answer the question “who?” or “what?” to probe an argument. The impersonalizing “se” obliterates the role of Agent, thus reducing the number of expressed arguments of the verb.

4.2.1 Impersonalizing Intransitive Verbs and Indirect Transitive Verbs

Intransitive verbs and indirect transitive verbs have a single noun phrase as an argument. When “se” suppresses this noun phrase, no other noun phrase can be promoted to subject. When this is the case, the subject is said to be indeterminate:

- (19) Em resumo: **chora-se** de tanto e por qualquer coisa que o choro justo se perde na choradeira.

- (20) Segundo a PF, **trata-se** da maior apreensão da droga já realizada pela delegacia do aeroporto.
- (21) Em momento algum, **fala-se** em privatização do ensino superior do RJ, resumiu Meirelles.

Pronominal intransitive verbs having an object unpredicted in their argument structure are considered instances of subject indeterminacy, rather than synthetic passives, their active voice counterparts being precluded:

- (22) **Sobe-se** uma pequena escada e encontra-se uma antiga casa com quartos separados do corredor por cortinas. (*uma escada é subida)
- (23) **Corre-se** o risco de judicializar e desfazer. (*o risco é corrido)
- (24) **Vive-se** hoje, ao contrário, uma saudável correção de rumos. (*uma correção é vivida)

4.2.2 Impersonalization of Direct Transitive Verbs and Ditransitive Verbs

From a syntactic perspective, instances of direct transitive verbs and ditransitive verbs impersonalized with “se” are typically classified (although not unanimously, cf. Martins; Nunes, 2016) as synthetic passives, as only one of the noun phrases is suppressed (that realizing the role of Agent) and the remaining noun phrase (realizing the role of Patient) takes on the subject function (as a Patient allows for an analytical passive voice paraphrase, it is treated as a passive subject rather than a direct object in an indeterminate subject construction).

Results from our queries in the Portinari-base corpus, however, point to a more productive analysis, particularly for the purpose of annotating corpora for NLP tasks, which demands high consistency in annotations for maximal generalization in machine learning models.

As pointed out in our review of the literature, Cançado and Amaral (2016) use clitic position as a criterion for characterizing synthetic passives: “se” occurs in a postverbal position. The noun phrase realizing the Patient occupies a postverbal position as well, which is not a typical position for subjects in Portuguese, and there may be number agreement between subject and verb, although this is not necessarily so.

Examples (25), (26), (27) and (28) from our corpus show instances of “se” constructions and a paraphrase with an analytical passive voice in brackets. Examples (25) and (26), however, do not meet Cançado and Amaral’s (2016) criterion, since the clitic is in preverbal position.

- (25) Como não desejar uma transformação em um país onde **se abriga** [é abrigado] um permanente e vergonhoso abismo social?
- (26) A isso **se aliou** [foi aliada] a privatização de aeroportos.
- (27) Já no Norte e Sudeste, **ultrapassou-se** [foi ultrapassada] consideravelmente a marca de 100%.
- (28) Por aqui, **tributa-se** [é/são tributado(s)] menos renda e propriedade e mais bens e serviços.

When the role of Patient is in preverbal position, the construction is generally not classified as synthetic passive, even though paraphrases with an analytical passive voice are admissible, as shown in (29), (30), (31) and (32) in brackets.

- (29) A outa metade **se elegeria** [seria eleita] por outro sistema, mais provavelmente o de voto em lista.
- (30) A euforia **se baseia** [é baseada] na valorização de 276% da divisa no ano.
- (31) Isso **se repete** [é repetido] até que ela entra, seguida por três homens sem rosto.
- (32) Isso **se aplica** [é aplicado] a inclinações políticas, gostos artísticos, preferências sexuais.

Examples like those above point to inconsistencies in the treatment of verbs with “se” in constructions in which there is only a noun phrase realizing a Patient. Postverbal position of the noun phrase realizing a Patient is for many authors the main criterion for classifying a synthetic passive voice construction, and preverbal position the main criterion for classifying an active voice construction. Nevertheless, in a construction having a focalized Patient and a suppressed Agent, the Patient is analyzed as subject of an active construction, instead of a passive one; therefore, it would be productive to annotate the postverbal noun phrase as a direct object. This would offer numerous advantages for NLP, as it would allow for analyzing those constructions as following a canonical SVO order in Portuguese: preverbal noun phrases realizing subjects (33) and postverbal ones, objects (34).

- (33) “Nossa corretora recomendou contratar apenas as coberturas *que* realmente **se encaixem** no nosso perfil e conforme a utilização do veículo”, diz a advogada.
- (34) Nesse sentido, **encaixa-se o trabalho** apresentado pelo grupo Cão no evento de lançamento.

In the case of verbs that admit inchoative alternation, the Theme occurring in preverbal position is annotated as subject (35), and it would be productive to annotate the Theme occurring in postverbal position as object of an indeterminate subject construction (36).

- (35) Quanto mais o *setor* **se torna** competitivo, mais o recurso estará à disposição.
- (36) O Trench demitiu Miller em julho, quando **se tornou** pública a *suspeita* sobre o ex-procurador.

Further examples problematizing transitivity as a criterion to distinguish two impersonalizing strategies (synthetic passives and indeterminate subject constructions) can be found in (37) and (38), and a weird paraphrase into analytical passive voice is shown, between brackets.

- (37) Por que não **se discute** [é discutido] melhor isso?
- (38) Em um dos cartazes, **lia-se** [era lido]: “professores desmotivados, alunos prejudicados”.

Another issue in the distinction between synthetic passives and indeterminate subject constructions is the criterion of agreement, since examples in which there is no number agreement between verbs and postverbal noun phrases are increasingly frequent in Brazilian Portuguese, as illustrated by examples (39), (40), (41), (42) and (43) retrieved from the corpus. In these examples, weird paraphrases into analytical passives (in brackets) suggest that an indeterminate subject reading would be more suitable.

- (39) Para percorrer todos os lugares, próximos uns dos outros, **leva-se** [é levado/são levados] em torno de 20 minutos.
- (40) Este seria um lado positivo de **se ter** [ser tido/serem tidos] representantes de partidos políticos tradicionais na disputa.
- (41) “Este é um momento crítico de definição a respeito do VOD, tem que **se falar** [ser falada/serem faladas] as coisas agora”, diz Klotzel.
- (42) A grande corrupção é definida como abuso para **se atingir** [ser atingido/serem atingidos] altos níveis de poder.
- (43) Um dos temas mais instigantes deste Setembro Amarelo – mês em que **se celebra** [é celebrada/são celebradas] a valorização da vida e a prevenção do suicídio – é o crescimento do número de jovens que se matam.

Moreover, instances in which the object is in preverbal position for the purpose of focalization seem to be most adequately accounted for by an indeterminate subject reading, rather than a passive one, as examples (44), (45), (46) and (47) illustrate.

- (44) E não é dízimo que **se fala** [é falado], é Edirzímio!
- (45) Aí eu aprendi que disco não **se faz** [é feito] reunindo dez, 12, 14 músicas.
- (46) Não é verdade o que **se fala** [é falado], que eu não tinha o direito de sair do país.
- (47) O atrito entre os componentes é a causa daquele apito chato que **se ouve** [é ouvido] ao frear.

Recurrent examples retrieved from our corpus also point to instances of pronominalized direct transitive verbs having a clausal complement to be most adequately accounted for by an indeterminate subject reading. In fact, their paraphrasing into synthetic passives does not yield natural constructions in Brazilian Portuguese, as the following examples show.

- (48) **Sabe-se** [é sabido] até onde pode ir o STF quando exposto a insinuações tão constrangedoras.
- (49) **Supunha-se** [era suposto] que Nuzman conhecesse o sistema bancário e o uso de cartões de crédito.
- (50) “Agora tem sol, nem **se pensa** [é pensado] que daqui a dois dias vai passar um tufão por aqui”, diz ela.

- (51) **Compreende-se** [é compreendido] que a sociedade veja com receio as concessões feitas a grupos que causaram tanto sofrimento ao país.
- (52) **Admita-se** [seja admitido] que o assoprador de apito não tenha visto a braçada de Jô.
- (53) **Acrescente-se**, [seja acrescentado] de todo modo, que a caracterização do crime de uso de informação privilegiada (“insider trading”, no jargão técnico) é tarefa complexa e relativamente nova no país.
- (54) **Recorde-se** [seja recordado] que os mercados de ações e câmbio viveram um terremoto em 18 de maio, dia seguinte à primeira notícia da famigerada conversa entre o presidente Michel Temer (PMDB) e Joesley Batista.
- (55) Digamos que **se consiga** [seja conseguido], com impostos, retirar metade da renda dos 10% mais ricos.
- (56) Com ele **busca-se** [é buscado] mitigar os custos reputacionais brutais que acumularam.
- (57) **Deve-se** [é devido] trabalhar com isso, avisa.
- (58) Já a pasta do Planejamento pretende avançar na desestatização, a começar por Congonhas, com o qual **se pretende** [é pretendido] obter R\$ 6 bilhões.

Finally, as already hinted at, the criterion of number agreement between verb and noun phrase for distinguishing synthetic passives and instances of indeterminate subject is not a productive one, as it is useful mainly for plural noun phrases, and even in such cases, agreement is increasingly not observed. Within our corpus, only two instances evidenced full number agreement between direct transitive verbs in plural form and their postverbal noun phrases:

- (59) São setores diferentes, não **se veem** [são vistas] *empresas de saneamento* fazendo essa migração.
- (60) Permite-se no Brasil que **se agreguem** [sejam agregados] *delatores* indefinidamente a um mesmo tronco de acusações.

4.2.3 Summing Up Non-Argumental “se” as an Impersonalization Strategy

Upon our analysis of retrieved occurrences of “se” that could be likely candidates to impersonalization instances, we filtered out 129 instances of subject indeterminacy and 32 instances of direct object that could be interpreted as subjects of synthetic passives. Nevertheless, we agree with Martins and Nunes (2016) on the little productivity in attempting to distinguish between instances of subject indeterminacy from synthetic passives. Table 4 displays a summary of possible configurations of impersonalizing “se”. An “x” indicates co-occurring arguments or context of occurrence (in the case of adjectival clause and nominal clause).

Table 4 – Configurations of impersonalizing “se”

NON-ARGUMENTAL “SE” AS AN IMPERSONALIZATION STRATEGY									
Subject	Direct Object	Indirect Object	Clausal Direct Object	Clausal Indirect Object	Adjectival Clause	Nominal Clause	Position	Example	Translation
		X					postverbal	Reclama-se de tudo.	People complain about everything.
		X					preverbal	De tudo se reclama.	People complain about everything.
			X				postverbal	Supõe-se que eles morreram.	They are believed to have died.
				X			postverbal	Preocupa-se mais em punir do que em educar.	People are more concerned with punishing than with educating.
						X	postverbal	Essa é a vantagem de se ter dinheiro. Essa é a vantagem de não se ter dinheiro.	That is the advantage of having money. That is the advantage of not having money.
					X		postverbal	Esse é o resultado que se esperava.	That is the expected result.
X							postverbal	Perdeu-se o prazo.	The deadline was missed.
X	X						postverbal	Exige-se dos alunos bom comportamento e assiduidade.	Good behavior and attendance are required of students.
X				X			postverbal	Acusou [acusaram]-se os fiscais de manipular os resultados.	The inspectors were accused of manipulating the results.
X							preverbal	Essa fruta se come com casca.	This fruit is eaten with its peel.
X							postverbal	Corre-se esse risco.	There is such a risk.
							postverbal	Vive-se bem aqui.	People live well here.

Source: the authors.

4.3 Non-Argumental “se” with No Syntactic Function

Apart from instances of argumental “se” and non-argumental “se” as an impersonalization strategy, there are constructions in which cliticized verbs have only one noun phrase argument which occurs in preverbal position and is annotated in the corpus as subject. Among these, we found differences pertaining to syntax (type of complement required by the cliticized verbs) and semantics (semantic role of subjects). When the subject is Agent or Experiencer, verbs can be inflected for person. On the other hand, when the subject is Theme or Patient, verbs can only take a third person and can only inflect for number (singular or plural).

We will report on each of the types of construction found in our corpus and provide a table summarizing them at the end.

4.3.1 Intransitive Verbs with Agent as Subject

There are cliticized verbs which have undergone grammaticalization, thereby suppressing the role of Patient (and consequently the function of “se” as a direct object). These are intransitive verbs having an Agent as subject, categorized as middle “se” by Cançado and Amaral (2016) and dynamic middle constructions by Godoy and Pinheiro (2023):

- (61) Afinal, sua federação fez de tudo para naturalizar o *garoto* que **se mudou** para o país aos 13, com um contrato de formação com o Real Madrid.
- (62) *Profissionais* de todos os tipos vão precisar de novas habilidades para **se orientar** com êxito no meio dessas transformações econômicas.

Among pronominal intransitive verbs having an Agent as subject, there is a group of verbs which allow for indirect object suppression and transformation into reciprocal intransitive ones, when a plural or collective subject fulfills the role of co-Agent and co-affected by the verb's action. These verbs denote union or separation. Examples are: “divorciar-se” (“to get divorced”), “separar-se” (“to get separated”), “casar-se” (“to get married”), “encontrar-se” (“to meet”), “juntar-se” (“to join”), “reunir-se” (“to get together”), “unir-se” (“to unite”), “aliar-se” (“to ally”), “aproximar-se” (“to get closer”), etc. Examples (63) and (64) illustrate reciprocal intransitive verbs.

- (63) O grupo de trabalho criado pelo Conselho **se reúne** pela primeira vez na próxima semana, na quarta-feira (4/10).
- (64) “É impressionante o número de pessoas que estão **se aproximando**”, disse.

4.3.2 Intransitive Verbs with Theme as Subject

There are pronominal verbs which have undergone grammaticalization, thereby suppressing the role of Cause and promoting the Theme to subject. These verbs are classified as inchoative “se” by Cançado and Amaral (2016) and as middle inchoative construction by Godoy and Pinheiro (2023).

- (65) Não é só a identidade que **se embaralha**, mas a própria percepção do tempo.
- (66) A cada ano, principalmente com a reforma da Previdência, o rombo **se estende** muito.

Among the group of intransitive verbs having a Theme as subject, in instances of ditransitive verb constructions, a direct object and indirect object were simultaneously promoted to subject in an inchoative construction, thereby yielding a kind of reciprocal meaning as is the case of constructions with argumental “se”. In the corpus, for instance, we found the verb “misturar” (“to mix”) in a ditransitive verb construction: “alguém mistura X com Y” (“somebody mixes X and Y”); and in an intransitive inchoative construction: “X e Y se misturam” (“X and Y mix”). This kind of alternation can occur with other verbs such as “confundir” (“to mix up”), “separar” (“to separate”), “somar” (“to sum up”), and “unir” (“to join”).

- (67) São universos que não **se misturam**.
- (68) Como uma novela de TV brasileira, são vários personagens, em núcleos que **se misturam**.
- (69) Os destinos de ambos **se separaram** no Equador.
- (70) Desde o início das investigações, de fato, as arestas na colaboração premiada **se somavam**.

Another group of intransitive verbs having a Theme as subject are verbs synonymous to “acontecer” (“to happen”), such as “passar-se” (“to take place”) and “dar-se” (“to occur”):

- (71) No filme, que não situa a época em que **se passa**, os alvos políticos são resumidos num dos personagens, um líder da luta por terras.
- (72) A novela **se passa**, principalmente, no Tocantins, incluindo sua capital, Palmas.
- (73) Isso pode **se dar** de várias maneiras, que incluem a participação em projetos de pesquisa e a colaboração com cientistas.

4.3.3 Intransitive Verbs with Experiencer as Subject

Verbs having an Experiencer as subject can inflect for different persons. Constructions of this type involve verbs which seem to be mostly derived from transitive constructions which have a Cause or a Causer as subject. The subject position of the thematic role of Cause is taken by the affected, which has the role of Experiencer. The Cause, however, may be realized as an adverbial modifier. Godoy and Pinheiro (2023) consider some of these instances cognitive middle constructions. Unlike inchoative constructions, which have a Theme as subject, constructions having an Experiencer as subject can inflect for person in verbs and pronouns. This type of construction involves mostly verbs of change in psychological state.

- (74) Há três anos, **abalou-se** com a morte do pai, Patrocino, após um AVC. (“abalar-se” = “ficar abalado” (“to be shaken”))
- (75) Ela **se emocionava** extraordinariamente com qualquer coisa. (“emocionar-se” = “ficar emocionado” (“get emotional”))
- (76) Os ânimos **se acalmaram** dias depois. (“acalmar-se” = “ficar calmo” (“became calm”))
- (77) Quando a reportagem perguntou por que ele precisava da informação, ele **se atrapalhou**. (“atrapalhar-se” = “ficar atrapalhado” (“to fumble”))
- (78) O São Paulo **se enervou** em campo, mas quase venceu nos acréscimos com cabeça de Militão. (“enervar-se” = “ficar nervoso” (“get nervous”))
- (79) Juan não parece **se intimidar** – viu de perto a guerra civil de Angola, em 1985, e ajudou o Haiti durante a epidemia de cólera e o terremoto de 2010. (“intimidar-se” = “ficar intimidado” (“to be intimidated”))
- (80) Alckmin **se irritou** e disse que o “novo” na política é “falar a verdade”. (“irritar-se” = “ficar irritado” (“to get angry”))

- (81) Portanto, não **se surpreenda** se aparecerem cada vez mais atores chineses ou sino-americanos em blockbusters hollywoodianos. (“surpreender-se” = “ficar surpreso” (“to be surprised”))

4.3.4 Indirect Transitive Verbs with Agent as Subject

Indirect transitive verbs having an Agent as subject can inflect for different persons. This is the case for the group of verbs characterized in this subsection. Some of these verbs are monosemic, and the pronominal meaning is the only one they have. Examples are “queixar-se” (“to complain”), “esbaldar-se” (“to strive”), “esforçar-se” (“to make an effort”), “esquivar-se” (“to dodge”), “refugiar-se” (“to take refuge”), “vangloriar-se” (“to boast”). Other verbs are polysemic and have a different meaning when they are not pronominal, having different syntactic complements. For instance, a verb may demand a direct object when it is not pronominal and an indirect object when it is pronominal – for instance, “prestar” (“to lend”) and “prestar se” (“to lend oneself to”) –, and a direct object when it is not pronominal and an indirect clausal object when it is pronominal – “recusar” (“to reject”) and “recusar-se” (“to refuse”). There are verbs which are polysemous even when they are pronominal. For example, “basear-se” (“to base”) means “tomar por base” (“to take as a basis”) when there is an Agent in subject position, and “ser baseado” (“to be based”) when there is a Theme in subject position.

- (82) Seria um baque nas contas públicas, **queixaram-se** os economistas.
- (83) A verdadeira questão é: por que um periódico supostamente científico **se presta** a isso?
- (84) Sobre a relação com a PF, sua assessoria diz que a PGR vai **se basear** em “diálogo, respeito mútuo, valorização do trabalho de investigação da polícia e cooperação”.

Some of these instances are classified as inherent “se” by Cançado and Amaral (2016), a category which according to the authors allows clitic dropping. But this seems to apply to monosemic pronominal verbs; for polysemic verbs, the clitic is essential to distinguish different meanings. This is the case with the following pronominal verbs, which, requiring an indirect object, construe a meaning that is different from the meaning they construe as direct transitive verbs.

- (85) Mas deu espaços à equipe de Cuca, escalada para **se aproveitar** disso. (“to take advantage”) (cf. “aproveitar” (“to enjoy”))
- (86) **Desculpou-se** pela confusão. (“to apologize”) (cf. “desculpar” (“to forgive”))
- (87) Sua esperança é que o eleitor gaste tempo para **se informar** e “ir além do discurso”. (“to find out”) (cf. “informar” (“to inform”))
- (88) O senado vai **se posicionar** sobre essa decisão. (“to take a position”) (cf. “posicionar” (“to position”, “to place”))

- (89) Quando um dos guerrilheiros **se rendeu**, ele gritou: “Sou Che Guevara, e valho mais a vocês vivo do que morto”. (“to surrender”) (cf. “render alguém” (“to subdue somebody”))

The pronominal verbs denoting union or separation may occur as reciprocal intransitive verbs (already illustrated in Section 4.3.1) or as an indirect transitive verb, as illustrated in this section. Even with an *indirect object* complement, this group of verbs meets the characteristics of middle “se” according to Cançado and Amaral (2016) and dynamic middle construction according to Godoy and Pinheiro (2023).

- (90) Existe expectativa de que o presidente **se reúna** ainda neste sábado *com Mouco e Moreira Franco*.
- (91) Assim como Gustavo Franco, **aproximou-se do Partido Novo**.
- (92) Em 2011, Twain **se casou com Frédéric Thiébaud**, o ex-marido da amiga com quem seu ex a traiu.
- (93) Maldonado havia **se juntado a um protesto de indígenas mapuche** que reivindicam um território na Província de Chubut, no sul do país, hoje sob posse da companhia italiana Benetton.

4.3.5 Indirect Transitive Verbs with Patient as Subject

As far as we know, constructions that present a topicalized Patient (to the left of the verb) are not approached in the literature as passive voice unless these constructions are analytical passives. Godoy and Pinheiro (2023) classify these instances as impersonal constructions. From our point of view, the clitic in these examples is used to convey a state from the patient’s perspective (the equivalents between parentheses reinforce this reading).

- (94) Os *capítulos* **se desdobram** em conflitos psicológicos da protagonista e as novas funções. (= “estão desdobrados”) (“to unfold”)
- (95) Esse sistema **baseia-se nas** entregas urbanas efetuadas por pessoas, caminhando, de bicicleta ou mesmo por outros meios de transporte disponíveis, a preços razoáveis e com redução de trânsito e poluição. (= “é/está baseado”) (“to be based”)
- (96) Isso **se aplica a** *inclinações políticas*, gostos artísticos, preferências sexuais. (= “é aplicável”) (“to apply”)
- (97) A classificação da equipe para a Copa do Mundo de 2018 **se apoia nisso**. (= “está apoiado”) (“to rely”)
- (98) Não se descarta que parte do aumento **se deva a um incremento** na notificação. (= “seja devido”) (“to be due”)

4.3.6 Cliticized Verbs Requiring a Predicative Complement

When a pronominal verb selects a predicative, several different configurations are possible. One of them is the pronominal verb linking a Theme as subject to a predicative, thus operating as a relational verb, either denoting permanence of state (“manter-se” (“to remain”)) or change of state (“tornar-se” (“to become”)), as illustrated by the following examples.

- (99) O resultado é que a sua participação no bolo total das despesas da União praticamente **se manteve** [permaneceu] *estável*.
- (100) **Tornou-se** [virou] *praxe* entre as seguradoras oferecer auxílio-residência, que inclui serviços como eletricista e encanador.
- (101) **Encontrava-se** [estava] *no país* em sua tentativa de exportar a revolução socialista ao mundo.

In other instances, there is an implicit linking verb in the “se” construction:

- (102) Muitos dos 60 pastores reunidos ali **se diziam** [diziam ser] *ex-militares* e exigiram do deputado que mostrasse como faria para criar um governo “firme”.
- (103) Uma porta-voz da Qiaobi, Xu Chunyan, não **se demonstrou** [demonstrou estar] *arrependida*.
- (104) Chamou também altos comandos e veteranos do Exército boliviano, que **se mostraram** [mostraram estar] *divididos* quanto à celebração.
- (105) Mas falta recheio para a mistura, que acaba **se mostrando** [mostrando ser] um *pastelão* de vento.
- (106) Ele vai se esforçar sozinho e, **sabendo-se** [sabendo ser] *aluno* da USP, ganhará orgulho misterioso e estudará sozinho.
- (107) Já o americano explosivo **se prova** [prova ser] na trama muito mais *tático* e focado do que seus acessos de raiva deixavam transparecer.

A passive auxiliary may also be implicit in the “se” construction:

- (108) Se alguém **se sentiu** [sentiu ter sido] *ofendido*, peço perdão, afirmou o jogador do Corinthians.

In factitive verbs, such as “fazer” (“to make”) and “deixar” (“to let”), a passive complement realized by an infinitive may be implicit in pronominalization:

- (109) Na visão de Mourão, eleições mostram que população **se deixou** *levar* [deixou-se ser levada] por promessas fáceis.
- (110) Em uma época em que se precisa diminuir e redistribuir o bolo, a voz dos mais vulneráveis não **se faz** *ouvir* [não se faz ser ouvida].

Finally, when pronominalized, a verb may require a nominal predicative (a noun or an adjective):

- (111) Estava na Califórnia **se formando** *mestre-cervejeira* quando soube da compra da Colorado pela Ambev.
- (112) Em 2005, a empresa passou a **se chamar** JBS (iniciais do nome de Zé Mineiro) e iniciou a internacionalização, com a compra da Swift Argentina.
- (113) Em artigo, escreveu que **se sentia** *perplexo* e amedrontado, submetido a humilhação e vexame sem precedentes.
- (114) Sob esse aspecto, o exemplo dos terraplanistas **se afigura** quase *benigno*.

The instances above point to the relevance of predicative complements as syntactic characteristics to distinguish between different meanings.

4.3.7 Summing Up Non-Argumental “se” with No Syntactic Function

While analyzing occurrences of non-argumental “se”, we found many constructions in which the Patient is subject and focus. These constructions are not acknowledged as passive constructions in Portuguese. These are constructions in which “se” is not reflexive, as is also the case in inchoative constructions, which have a Theme as subject and focus.

Clitic dropping did not prove a productive criterion to classify “se” constructions due to two main reasons: (i) this is mostly related to language variation, as it has higher incidence in some regions in Brazil than others; (ii) in polysemous verbs, where cliticization brings about multiple readings, clitic dropping may not happen, as the clitic is an essential element for meaning discrimination.

Verbs cliticized with non-argumental “se” may have an Agent, an Experiencer, a Theme or a Patient as subject. These constructions may select different types of complement and may be accounted for by different types of syntactic alternations. The main way to distinguish between each of the configurations is to assign semantic role labels, as the types of complements may be inferred from the text. Table 5 shows the main syntactic and semantic environments for non-argumental “se” with no syntactic function.

Table 5 – Configurations of non-argumental “se” with no syntactic function

NON-ARGUMENTAL “SE” WITH NO SYNTACTIC FUNCTION										
Subject	Thematic Role of Subject	Direct Object	Indirect Object	Clausal Direct Object	Clausal Indirect Object	Predicative	Reciprocal	Has Non-Pronominal Meaning	Example	Translation
X	AGENT							YES	Ele se mudou para cá no ano passado.	He moved here last year.
X	AGENT						X	YES	Eles se casaram sábado.	They got married on Saturday.
X	AGENT		X					NO	Ele se divorciou da esposa recentemente.	He recently divorced his wife.
X	AGENT				X			YES	Ele se recusou a cooperar.	He refused to cooperate.

X	THEME		YES	A montanha se formou há milhões de anos.	The mountain was formed millions of years ago.
X	THEME		X YES	As cores se misturaram.	The colors blended.
X	THEME	X	YES	A temperatura se manteve igual.	The temperature remained the same.
X	PATIENT	X	YES	A denúncia se apoia em evidências.	The complaint is based on evidence
X	PATIENT	X	YES	Ela se chama Nina.	Her name is Nina.
X	EXPERIENCER		YES	Ela se alegrou.	She cheered up.
X	EXPERIENCER	X	YES	Ela se sentiu feliz.	She felt happy.

Source: the authors.

4.4 Insights from the Syntactic and Semantic Analysis of “se”

Upon classifying all the occurrences of the pronoun “se” in a corpus of Brazilian Portuguese, we came to the conclusion that approaching “se” from two separate perspectives – a syntactic and a semantic one – proves more effective and productive.

From a syntactic perspective, “se” may be:

- ◆ a verb argument (direct or indirect object);
- ◆ a particle to signal an impersonal construction, with subject suppression;
- ◆ an expletive, with no syntactic function.

From a semantic perspective, “se” may be:

- ◆ a token playing a thematic role (Patient or Theme, in the case of direct object, and several thematic roles, in the case of indirect object);
- ◆ a marker for Agent suppression, rendering a construction impersonal;
- ◆ a marker for thematic role alternation;
- ◆ a marker for a new perspective on the verb meaning or a marker for a new meaning of the verb;
- ◆ an expletive, with no semantic function, in monosemic verbs.

In the following subsections, we discuss some of the main difficulties encountered in our analysis.

4.4.1 Alternation

Existing taxonomies attempt to explain cliticized verbs based on an assumed basic construction of direct transitive verbs and ditransitive verbs. This may be fruitful from a linguistic perspective, as it illuminates the paths of grammaticalization, whereby direct transitive verbs become intransitive verbs and ditransitive verbs become indirect transitive

verbs. However, this approach is not productive for corpus annotation or particularly useful for NLP, as indicators for alternation are not always detectable. Moreover, a “se” construction may not necessarily entail alternation of thematic roles, but could actually be construing a new meaning.

Non-argumental “se” seems to operate as a suppressor of a thematic role. Yet, it is not an easy task to classify this type of “se” based on the process of Agent obliteration it derives from. As far as our data show, verbs in non-argumental “se” constructions can be intransitive verbs and indirect transitive verbs. Neither direct transitive verbs nor ditransitive verbs occur in these constructions. Some of the intransitive verbs and indirect transitive verbs in non-argumental “se” constructions have Agents and Experiencers as subjects, while others have Themes or Patients as subjects. The same happens to alternations of verbs which have a predicative complement.

If only one alternation were grammaticalized through pronominalization (having an Agent or Theme/Patient as subject), thematic roles and the meaning of a verb could be predicted in the lexicon. However, some verbs have developed two alternations: one of them with Patient suppression (the Agent remaining), and the other one suppressing the Agent and promoting the Theme or Patient to subject position. This is the case of verbs such as “basear-se” (“to base”, “to rely”), “concentrar-se” (“to center”, “to concentrate”), and “manter-se” (“to keep”, “to stay”).

- (115) *Fãs de Madonna se concentram* em frente ao Hotel Copacabana Palace. (Agent as subject)
- (116) *Áreas de lazer aumentam* em São Paulo, mas ainda *se concentram* em regiões nobres. (Theme as subject)
- (117) Sobre a relação com a PF, sua assessoria diz que a *PCR* vai *se basear* em “diálogo, respeito mútuo, valorização do trabalho de investigação da polícia e cooperação”. (Agent as subject)
- (118) *A euforia se baseia* na valorização de 276% da divisa no ano. (Theme as subject)
- (119) O *ex-ministro* terá prazo para apresentação de defesa caso pretenda *manter-se* filiado ao partido. (Agent as subject)
- (120) O resultado é que a sua *participação* no bolo total das despesas da União praticamente *se manteve* estável. (Theme as subject)

4.4.2 Multiple Readings of Cliticized Verbs

A major challenge when annotating “se” in a corpus is to assess whether a pronominal verb has a direct transitive or ditransitive argument structure; or whether “se” is argumental or not; or whether “se” is impersonalizing a construction through subject suppression; or if “se” has no syntactic function and operates as a marker of new meanings of a verb or alternation of semantic roles. Depending on the context, verbs allow different readings and interpretations. This can be illustrated by the verb “apresentar-se” (respectively “to join”, “to perform”, “to be”, and “to present” in the following examples):

- (121) Quando **se apresentou** ao Santos, *Nilmar* não disputava uma partida oficial havia 14 meses. (Agent as subject, argumental “se”, direct object, Patient);
- (122) *Chiara Santoro e Daniel Gonçalves* **apresentam-se** no Música ao Meio-Dia. (Agent as subject, non-argumental “se”);
- (123) Eles também possuem um maior número de valvas no seu interior, os *vasos linfáticos* **apresentam-se** dilatados e exibem um aspecto nodular ou em colar de contas. (Theme as subject, non-argumental “se”, marker of new meaning, synonymous to “estar” (“to be”), with predicative complement);
- (124) **Apresentam-se**, a seguir, *dois símbolos* utilizados em desenhos topográficos. (impersonalizing “se”, non-argumental and postverbal noun phrase as object).

4.4.3 Cliticized Verbs Selecting a Complement

The literature approaches inchoative and middle voice alternations as processes which do away with an argument in the argument structure of direct transitive verbs, thereby yielding an intransitive verb. Inchoative alternation does away with the argument having the role of Cause and promotes the Theme as subject, whereas middle voice alternation does away with the argument having the role of Patient and keeps the Agent as subject. There are, however, many instances of pronominalization which, despite not having an argumental “se”, do not become intransitive verbs, selecting either a prepositional argument (phrase or clause) in the argumental structure, or a predicative complement. These complements can be relevant cues to distinguish different meanings, as seen in Section 4.3.

4.5 A Proposal for a More Fine-Grained and NLP-Productive Annotation of “se” in Corpora of Brazilian Portuguese

Based on the insights gathered from our analysis, we outlined a proposal for annotating “se” in corpora.

Guidelines for annotating argumental “se”:

- ◆ Annotate “se” as direct or indirect object if “se” can be probed by substituting a first- or second-person pronoun and this substitution does not trigger a substitution of the

person of the subject. For example: “Ele se matou” (“He killed himself”) can be probed by “Ele os matou” (“He killed them”) and “Ele se concedeu o benefício da dúvida” (“He gave himself the benefit of the doubt”) can be probed by “Ele me concedeu o benefício da dúvida” (“He gave me the benefit of the doubt”);

- ◆ Annotate as a pronoun feature if the construction is clearly reciprocal. For example: “Eles se tratam por apelidos” (“They address each other using nicknames”); “Eles se deram as mãos” (“They held each other’s hands”);
- ◆ Annotate as a pronoun feature accusative case for direct objects and dative case for indirect objects. For example: “Eles se [acc] tratam por apelidos” (“They address each other using nicknames”); “Eles se [dat] deram as mãos” (“They held each other’s hands”);
- ◆ Annotate “se” with the feature “reflexive”;
- ◆ Annotate for “se” the pronoun features of person and number (third person for all instances and singular or plural for number).

Guidelines for annotating impersonalizing “se”:

- ◆ Annotate “se” as impersonalizing expletive pronoun;
- ◆ Do not annotate features (case, person, number);
- ◆ Do not annotate reflexive or reciprocal.

Guidelines for annotating non-argumental “se” with no syntactic function:

- ◆ Annotate “se” as expletive pronoun;
- ◆ If the verb can inflect for person, annotate the feature Reflex (reflexive); the subject in these constructions will be annotated as Agent or Experiencer in thematic role annotation. For example: “Ele se resignou” (“He resigned himself”); “Ele se informou a respeito” (“He found out about it”); “Ele se formou médico” (“He graduated as a doctor”);
- ◆ If the verb can only inflect for number and not for person, do not annotate the feature Reflex (reflexive); the subject in these constructions will be annotated as Theme or Patient in thematic role annotation. For example: “O espetáculo se caracteriza pela simplicidade” (“The show is characterized by its simplicity”);
- ◆ Annotate Rcp (reciprocal) as a pronoun feature; this feature may occur when the subject is Theme, cf. Subsection 4.3.2. For example: “Eles se reuniram” (“They came together”), “As cores se mesclaram” (“The colors merged”).

In order to support annotations of “se” following the UD guidelines, which are the guidelines used in the corpus Portinari-base, Table 6 provides the PoS and dependency relation tags as well as the features and values which are suggested by our proposal.

Table 6 – Annotation proposal for UD treebanks of Brazilian Portuguese

ANNOTATION PROPOSAL								
POS TAG	Dependency Relation	PronType	Case	Reflex	Person	Number	Example	Translation
PRON	obj	Pers	Acc	Yes	3	Sing or Plur	Ele se matou. Ele se considera feio. Ele se forçou a comer.	He killed himself. He considers himself ugly. He forced himself to eat.
PRON	obj	Rep	Acc	Yes	3	Sing or Plur	Eles se tratam por apelidos.	They call each other by nicknames.
PRON	iobj	Pers	Dat	Yes	3	Sing or Plur	Ele se atribuiu a culpa. Ele se permitiu fazer uma pausa.	He assigned the blame to himself. He allowed himself to take a break.
PRON	iobj	Rcp	Dat	Yes	3	Sing or Plur	Eles se deram as mãos.	They held hands.
PRON	expl:impers						Reclama-se de tudo. Perdeu-se o prazo. Prevê-se/Preveem-se bons resultados. Espera-se que chova. Pretende-se fazer isso.	People complain about everything. The deadline was missed. Good results are expected. It is expected to rain. We intend to do that.
PRON	expl	Pers		Yes	3	Sing or Plur	Ele se desesperou. Ele se casou com ela. Ele se diz arrependido. Ele se desculpou por ter mentido. Ele se vangloriou disso.	He despaired. He married her. He claims to be regretful. He apologized for having lied. He boasted about it.
PRON	expl	Rcp		Yes	3	Sing or Plur	Eles se casaram. Eles se divorciaram.	They got married. They got divorced.
PRON	expl				3	Sing or Plur	O livro se baseia em relatos. O livro se constitui de capítulos.	The book is based on accounts. The book consists of chapters.
PRON	expl	Rcp			3	Sing or Plur	As cores se mesclaram. As vozes se misturaram.	The colors blended. The voices mixed mingled.

Source: the authors.

5 Conclusions

Our experience in coordinating annotation projects in corpora of Brazilian Portuguese pointed to the pronoun “se” as a problematic issue and to the limitations of available taxonomies, since they do not consider important variables for the classification of “se”, such as the position of the thematic roles Theme and Patient (in the absence of Cause and Agent) with respect to the verb and the nature of the complements selected by pronominal verbs that are not intransitive (a prepositional phrase, a clause or a predicative). These limitations bring uncertainty to annotators, who have to make decisions regarding instances that do not match prototypical examples cited in the literature. In situations like these, each annotator can produce a different analysis and, if there is no agreement between the annotators, the consistency of the annotation is compromised, as are subsequent automatic tasks that make use of human annotation.

We have attempted to provide a more thorough description and systematization in the form of tables for three particular environments of “se”, namely argumental “se”, impersonalizing “se” and non-argumental “se” with no syntactic function. The endeavor proved hard, and the reason for this is a fact that is rarely discussed in the literature: the same

verb cliticized with the pronoun “se” can be classified into different categories depending on the context, i.e., there is more than one possible reading in many verbs cliticized with “se”.

While we have described several semantic functions of “se” with no syntactic function, we believe we have not covered the whole spectrum, which would certainly demand analyzing “se” occurrences in other corpora, ideally of other genres. As the corpus we have drawn on is a treebank annotated following the UD guidelines, we have relied on those guidelines to set up a proposal for annotating the various instances of “se” with tags and features that enrich their annotation.

Our proposal for annotating syntactic relations and morphosyntactic features can be mapped onto an annotation of thematic roles, which is relevant for NLP. This mapping was highlighted throughout the article so that annotation of thematic roles can rely on syntactic cues. This should not be taken to imply, however, that syntactic annotation will be sufficient to guide the semantic one. Likewise, since many verbs with the clitic “se” allow for several interpretations, which can ultimately be solved by resorting to context, lexis on its own does not provide sufficient cues to determine the semantics of pronominal verbs (except when they are monosemic).

Therefore, we believe that our proposal for enriching syntactic annotation, subsequently complemented by thematic role annotation, will provide valuable data to train models for syntactic and semantic parsing, which will, in turn, yield resources for AI applications.

Finally, we would like to acknowledge the relevant contribution of the taxonomies in the literature, which, despite not being sufficiently comprehensive to account for all “se” phenomena, were enlightening enough to approach the corpus.

Acknowledgements

The authors would like to thank the reviewers for their valuable suggestions.

Funding

This work was carried out at the Center for Artificial Intelligence of the University of São Paulo (C4AI – <http://c4ai.inova.usp.br/>), with support by the São Paulo Research Foundation (FAPESP grant #2019/07665-4) and by the IBM Corporation. The project was also supported by the Ministry of Science, Technology and Innovation, with resources of Law N. 8.248, of October 23, 1991, within the scope of PPI-SOFTEX, coordinated by Softex and published as Residence in TIC 13, DOU 01245.010222/2022-44. Adriana Pagano has a grant from the National Council for Scientific and Technological Development (CNPq 404722/2024-5; 313103/2021-6) and Minas Gerais State Agency for Research and Development (FAPEMIG).

References

- CÂMARA JR., J. M. *Princípios de linguística geral*. Rio de Janeiro: Livraria Acadêmica, 1972.
- CANÇADO, M.; AMARAL, L. *Introdução à Semântica Lexical*: papéis temáticos, aspecto lexical e decomposição de predicados. Petrópolis: Editora Vozes, 2016. 256p. *E-book*.
- DURAN, M. S.; ALUÍSIO, S. M. O tratamento da partícula “se” para fins de anotação de papéis semânticos. In: JORNADA DE DESCRIÇÃO DO PORTUGUÊS, 2., 2011; SYMPOSIUM IN INFORMATION AND HUMAN LANGUAGE TECHNOLOGY (STIL), 8., 2011, Cuiabá. *Proceedings [...]*. Porto Alegre: Sociedade Brasileira da Computação, 2011. p. 24-26. Available at: http://www.nilc.icmc.usp.br/stil/stil2011_English/jdp/artigos/STIL2011_JDP_P4.pdf. Accessed on: 24 Sept. 2025.
- DURAN, M. S.; LOPES, L.; NUNES, M. G. V.; PARDO, T. A. S. The Dawn of the Porttinari Multigenre Treebank: Introducing Its Journalistic Portion. In: SYMPOSIUM IN INFORMATION AND HUMAN LANGUAGE TECHNOLOGY (STIL), 14., 2023, Belo Horizonte. *Proceedings [...]*. Porto Alegre: Sociedade Brasileira de Computação, 2023. p. 115-124. Available at: <https://sol.sbc.org.br/index.php/stil/article/view/25443/25264>. Accessed on: 24 Sept. 2025.
- DURAN, M. S.; SCARTON, C.; ALUISIO, S.; RAMISCH, C. Identifying Pronominal Verbs: Towards Automatic Disambiguation of the Clitic ‘se’ in Portuguese. In: WORKSHOP ON MULTIWORD EXPRESSIONS, 9., 2013, Atlanta. *Proceedings [...]*. Atlanta: Association of Computational Linguistics, 2013. p. 93-100. Available at: <https://aclanthology.org/W13-1014>. Accessed on: 24 Sept. 2025.
- GODOY, L. A reflexivização no português brasileiro e a decomposição semântica de predicados. 2012. 158 p. Doctoral Dissertation (PhD in Linguistic Studies) – Faculty of Letters, Federal University of Minas Gerais, Belo Horizonte, 2012.
- GODOY, L.; PINHEIRO, D. A rede gramatical das construções com se no português brasileiro. *Revista Soletras*, Rio de Janeiro, n. 45, p. 123-150, 2023. DOI: <https://doi.org/10.12957/soletras.2023.73491>.
- GUIBON, G.; COURTIN, M.; GERDES, K.; GUILLAUME, B. When Collaborative Treebank Curation Meets Graph Grammars. In: LANGUAGE RESOURCES AND EVALUATION CONFERENCE, 12., 2020, Marseille. *Proceedings [...]*. Marseille: European Language Resources Association, 2020. p. 5293-5302. Available at: <https://www.aclweb.org/anthology/2020.lrec-1.651>. Accessed on: 24 Sept. 2025.
- MARNEFFE, M. C. de; MANNING, C. D.; NIVRE, J.; ZEMAN, D. Universal Dependencies. *Computational Linguistics*, v. 47, n. 2, p. 255-308, 2021. MIT Press. DOI: https://doi.org/10.1162/coli_a_00402.
- MARTINS, A. M.; NUNES, J. Passives and *Se* Constructions. In: WETZELS, W. L.; COSTA, J.; MENUZZI, S. (ed.). *The Handbook of Portuguese Linguistics*. Oxford: John Wiley & Sons, 2016. p. 318-337. DOI: <https://doi.org/10.1002/9781118791844.ch17>.
- MELLO, H.; RASO, T.; OLIVEIRA JUNIOR, M. A.; SARDINHA, T. B.; FREITAS, C.; ALUÍSIO, S. M.; PARDO, T. A. S.; DURAN, M. S.; LEAL, S. E.; DAVIES, M.; GALVES-CHAMBELLAND, C. Brazilian Portuguese: Spoken, Written and Diachronic Corpora. In: CONGRESS OF THE ITALIAN SOCIETY OF LINGUISTICS, 54., 2021, Online. *Proceedings [...]*. Milan: Officinaventuno, 2022. p. 179-188. DOI: <https://doi.org/10.17469/O2106SLI000011>.
- NUNES, J. Ainda o famigerado SE. *DELTA*, v. 11, n. 2, p. 201-240, 1995.
- NUNES, J. O famigerado SE: uma análise sincrônica e diacrônica das construções com SE apassivador e indeterminador. 1990. 172 p. Thesis (Master's Degree in Linguistics) – Institute of Language

Studies, Unicamp, Campinas, 1990. Available at: <http://www.bibliotecadigital.unicamp.br/document/?code=vtls000045410>. Accessed on: 24 Sept. 2025.

SOUZA, E. de; FREITAS, C. Um pronome com muitas funções: descrição e resultados da anotação do pronome -se em um treebank segundo o esquema Universal Dependencies (UD) para Português. In: SYMPOSIUM IN INFORMATION AND HUMAN LANGUAGE TECHNOLOGY (STIL), 14., 2023, Belo Horizonte. *Proceedings* [...]. Porto Alegre: Sociedade Brasileira de Computação, 2023. p. 278-287. DOI: <https://doi.org/10.5753/stil.2023.233986>.

History

Date of submission: 31/07/2024.

Date of approval: 05/09/2025.

Responsible Editor

Andréa Machado de Almeida Mattos, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais/MG, Brazil. Lattes: <http://lattes.cnpq.br/7749222257907067>, ORCID: <https://orcid.org/0000-0003-3190-7329>, email: andreamattos@ufmg.br.

Authors' Contribution

Adriana Silvina Pagano contributed to the drafting and revision of this paper. Magali Sanches Duran contributed to the compilation, annotation and analysis of the corpus and drafting of the paper. Thiago Alexandre Salgueiro Pardo contributed to the drafting of the introduction and background sections as well as to the final proofreading.

Declaration of Conflict of Interest

The authors declare that there is no conflict of interest.

Data Availability Statement

The data analysed were retrieved from the treebank Porttinari 2.0 developed within the project POETISA, which is publicly available at: <https://sites.google.com/icmc.usp.br/poetisa/porttinari>.

Use of AI

No AI was used in the data analysis or drafting of this article.

Reviews

As part of the commitment made by the Brazilian Journal of Applied Linguistics to Open Science, the journal publishes the reviews issued regarding its published works, when authorized by all parties involved.

Appendix – Examples provided in the article followed by a free translation into English

- (1) A política agora **se parece** mais com o que era antes da crise.
Present politics **resembles** more what it was like before the crisis.
- (2) A mulher **se torna** humana ao mesmo tempo em que é objetificada.
A woman **becomes** human as she is objectified.
- (3) Quando começa a mexer em três setores, essa oportunidade **se perde** em função da falta de estrutura.
When it begins to impact three sectors, that opportunity **gets lost** due to lack of structure.
- (4) Em queda na pesquisa, Doria pode **se atropelar** na ambição e fracassar.
Falling in the polls, Doria can **get carried away** with his ambition and fail.
- (4a) Em queda na pesquisa, Doria pode **atropelar** João com sua ambição.
Falling in the polls, Doria can **run over** João with his ambition.
- (5) As garotinhas podem **se perguntar** agora: eu quero ser cabeleireira ou chanceler?
Little girls can now **ask themselves**: would I like to be a hairdresser or a chancellor?
- (5a) As garotinhas podem **perguntar** a Maria agora: você quer ser cabeleireira ou chanceler?
Little girls can now **ask** Maria: would you like to be a hairdresser or a chancellor?
- (6) Rocque diz que ainda **se considera** neoliberal, mas **se classifica** como “ex-banqueira e ex-economista”.
Rocque says she still **considers herself** a neoliberal, but **classifies herself** as an “ex-banker and ex-economist.”

- (6a) Rocque diz que ainda **considera** João neoliberal, mas o **classifica** como “ex-banqueiro e ex-economista”.
Rocque says she still **considers** João a neoliberal, but **classifies** him as an “ex-banker and ex-economist.”
- (7) O mercado **se apeg**a a isso e vê que está muito complicado operar, afirma.
The market **clings** to this and sees that it is very complicated to operate, he says.
- (7a) *O mercado os **apega** a isso e vemos que está muito complicado operar, afirma.
*The market **clings** them to this and we see that it is very complicated to operate, he says.
- (7b) Nós **nos apegamos** a isso e vemos que está muito complicado operar, afirma.
We **cling** to this and see that it is very complicated to operate, he says.
- (8) Sem JK, Frente Ampla **se reúne** no Rio para planejar estratégia.
Without JK, Frente Ampla **meets** in Rio to plan strategy.
- (8a) Sem JK, Frente Ampla **reúne-os** no Rio para planejar estratégia.
Without JK, Frente Ampla **recruits them** in Rio to plan strategy.
- (8b) Sem JK, nós **nos reunimos** no Rio para planejar estratégia.
Without JK, we **meet** in Rio to plan strategy.
- (9) A narrativa **se concentra** em alguns de seus filmes e despreza outros.
The narrative **focuses** on some of his films and ignores others.
- (10) O Brasil é um país que **se define** pela copulação incessante dos contrários.
Brazil is a country that **defines itself** by the ceaseless copulation of opposites.
- (11) O Brasil é um país **definido** pela copulação incessante dos contrários.
Brazil is a country **defined** by the ceaseless copulation of opposites.
- (12) O Brasil é um país que **se define** como uma copulação incessante de contrários.
Brazil is a country that **defines itself** as a ceaseless copulation of opposites.
- (13) Os terceiros colocados **se enfrentam** por uma chance na repescagem.
The third-placed teams **face each other** for a chance in the play-offs.
- (14) Na época, quatro cooperativas **se deram** as mãos para criar a Fecoagro.
At the time, four cooperatives **joined** forces to create Fecoagro.
- (15) Ao todo, 32 petroleiras **se inscreveram** para a disputa.
In all, 32 oil companies **entered** the bid.

- (16) Está **se perguntando** por que o tratamento fiscal desse produto é diferente dos demais?
Are you **wondering** why the tax regime for this product is different from the others?
- (17) É quase a mesma proposta do A Bola que Pariu, feito por torcedoras que **se declaram** fanáticas e escrevem sobre seus próprios clubes.
It's almost the same proposal as A Bola que Pariu, conceived of by female soccer fans who **declare themselves** passionate and write about their own clubs.
- (18) Ainda que seu canto fosse as dores da matriarca, Sangaré conta ter percebido que as mulheres que a ouviam também **se viam** decifradas nele.
Even though her song was the pain of the matriarch, Sangaré says she realized that the women who listened to her also **saw themselves** deciphered in it.
- (19) Em resumo: **chora-se** de tanto e por qualquer coisa que o choro justo se perde na choradeira.
In short: everything is so much **cried upon** that the just lament is lost in the weep frenzy / we **cry** so much and for so many different reasons that the true cry is lost in the crying.
- (20) Segundo a PF, **trata-se** da maior apreensão da droga já realizada pela delegacia do aeroporto.
According to the Federal Police, **this is** the biggest drug confiscation ever made by the airport police.
- (21) Em momento algum, **fala-se** em privatização do ensino superior do RJ, resumiu Meirelles.
There is no talk whatsoever of privatizing higher education in Rio de Janeiro, summarized Meirelles.
- (22) **Sobe-se** uma pequena escada e encontra-se uma antiga casa com quartos separados do corredor por cortinas. (*uma pequena escada é subida)
You go up a small staircase and find an old house with curtains separating the rooms from the corridor. (*a small staircase is climbed)
- (23) **Corre-se** o risco de judicializar e desfazer. (*o risco é corrido)
One runs the risk of judicialization and dissolution. (*the risk is taken)
- (24) **Vive-se** hoje, ao contrário, uma saudável correção de rumos. (*uma correção é vivida)
Today, on the contrary, **we are experiencing** a healthy correction of course. (*a correction is experienced)
- (25) Como não desejar uma transformação em um país onde **se abriga** um permanente e vergonhoso abismo social?

How can we not long for a transformation in a country that is **home** to a permanent and shameful social abyss?

- (26) A isso **se aliou** a privatização de aeroportos.
This was followed by the privatization of airports.
- (27) Já no Norte e Sudeste, **ultrapassou-se** consideravelmente a marca de 100%.
However, in the North and Southeast, the 100% mark **was** considerably **surpassed**.
- (28) Por aqui, **tributa-se** menos renda e propriedade e mais bens e serviços.
Here, we **levy** less tax on income and property and more on goods and services.
- (29) A outra metade **se elegeria** por outro sistema, mais provavelmente o de voto em lista.
The other half **would be elected** by another system, most likely the list vote.
- (30) A euforia **se baseia** na valorização de 276% da divisa no ano.
The euphoria is **based** on the 276% appreciation of the currency over the year.
- (31) Isso **se repete** até que ela entra, seguida por três homens sem rosto.
This is **repeated** until she enters, followed by three faceless men.
- (32) Isso **se aplica** a inclinações políticas, gostos artísticos, preferências sexuais.
This **applies** to political leanings, artistic tastes and sexual preferences.
- (33) “Nossa corretora recomendou contratar apenas as coberturas que realmente **se encaixem** no nosso perfil e conforme a utilização do veículo”, diz a advogada.
“Our broker recommended that we only buy the insurance that really **fits** our profile and the use of the vehicle,” says the lawyer.
- (34) Nesse sentido, **encaixa-se** o trabalho apresentado pelo grupo Cão no evento de lançamento.
In this sense, the work presented by the group Cão at the launch event is particularly **fitting**.
- (35) Quanto mais o setor **se torna** competitivo, mais o recurso estará à disposição.
The more competitive the sector **becomes**, the more this resource will be available.
- (36) O Trench demitiu Miller em julho, quando **se tornou** pública a suspeita sobre o ex-procurador.
Trench fired Miller in July, when suspicions about the former attorney **became** public.
- (37) Por que não **se discute** melhor isso?
Why isn't that better **discussed**?
- (38) Em um dos cartazes, **lia-se**: “professores desmotivados, alunos prejudicados”.
One of the posters **read**: “Unmotivated teachers, harmed students”.

- (39) Para percorrer todos os lugares, próximos uns dos outros, **leva-se** em torno de 20 minutos.
It **takes** around 20 minutes to visit all the places which are close in distance to each other.
- (40) Este seria um lado positivo de **se ter** representantes de partidos políticos tradicionais na disputa.
This would be a positive side of **having** representatives from traditional political parties in the election.
- (41) “Este é um momento crítico de definição a respeito do VOD, tem que **se falar** as coisas agora”, diz Klotzel.
“This is a critical moment for decision-making on VOD, **we have to talk** about it now,” says Klotzel.
- (42) A grande corrupção é definida como abuso para **se atingir** altos níveis de poder.
High-level corruption is defined as abuse **to achieve** high levels of power.
- (43) Um dos temas mais instigantes deste Setembro Amarelo – mês em que **se celebra** a valorização da vida e a prevenção do suicídio – é o crescimento do número de jovens que se matam.
One of the most challenging themes of this Yellow September – the month in which **we celebrate** the value of life and the prevention of suicide – is the increase in the number of young people who kill themselves.
- (44) E não é dízimo que **se fala**, é Edirzímio!
And **the word for it** is not “tithe”, it’s “Edir-tithe”!
- (45) Aí eu aprendi que disco não **se faz** reunindo dez, 12, 14 músicas.
So I learned that **you cannot produce** a record by putting together ten, 12 or 14 songs.
- (46) Não é verdade o que **se fala**, que eu não tinha o direito de sair do país.
It’s not true what **they say**, that I had no right to leave the country.
- (47) O atrito entre os componentes é a causa daquele apito chato que **se ouve** ao frear.
The friction between the components is the cause of that annoying whistle **you hear** when braking.
- (48) **Sabe-se** até onde pode ir o STF quando exposto a insinuações tão constrangedoras.
We know how far the STF can go when exposed to such embarrassing insinuations.
- (49) **Supunha-se** que Nuzman conhecesse o sistema bancário e o uso de cartões de crédito.
Nuzman **was supposed** to be familiar with the banking system and the use of credit cards.

- (50) “Agora tem sol, nem **se pensa** que daqui a dois dias vai passar um tufão por aqui”, diz ela.
“It’s sunny now, **you can’t imagine** that in two days’ time a typhoon will strike here,” she says.
- (51) **Compreende-se** que a sociedade veja com receio as concessões feitas a grupos que causaram tanto sofrimento ao país.
It’s understood that society is wary of concessions made to groups that have caused the country so much suffering.
- (52) **Admita-se** que o assoprador de apito não tenha visto a braçada de Jô.
Let’s concede that the whistle blower may not have seen Jô’s stroke.
- (53) **Acréscete-se**, de todo modo, que a caracterização do crime de uso de informação privilegiada (“insider trading”, no jargão técnico) é tarefa complexa e relativamente nova no país.
In any case, **it should be added** that characterizing the crime of “insider trading” (used as a technical term) is a complex and relatively new task in the country.
- (54) **Recorde-se** que os mercados de ações e câmbio viveram um terremoto em 18 de maio, dia seguinte à primeira notícia da famigerada conversa entre o presidente Michel Temer (PMDB) e Joesley Batista.
It should be remembered that the stock and foreign exchange markets experienced an earthquake on May 18, the day after the first news of the infamous conversation between President Michel Temer (PMDB) and Joesley Batista.
- (55) Digamos que **se consiga**, com impostos, retirar metade da renda dos 10% mais ricos.
Let’s suppose **you succeed**, with taxes, in taking half of the income of the richest 10%.
- (56) Com ele **busca-se** mitigar os custos reputacionais brutais que acumularam.
With this, **we seek** to mitigate the huge reputational costs that have accumulated.
- (57) **Deve-se** trabalhar com isso, avisa.
We have to work with that, he warns.
- (58) Já a pasta do Planejamento pretende avançar na desestatização, a começar por Congonhas, com o qual **se pretende** obter R\$ 6 bilhões.
The Ministry of Planning, meanwhile, intends to move ahead with privatization, starting with Congonhas, which is **expected to** raise R\$6 billion.
- (59) São setores diferentes, não **se veem** empresas de saneamento fazendo essa migração.
These are different sectors; **we do not see** sanitation companies making this transition.
- (60) Permite-se no Brasil que **se agreguem** delatores indefinidamente a um mesmo tronco de acusações.

In Brazil, whistleblowers can **be added** indefinitely to the same set of charges.

- (61) Afinal, sua federação fez de tudo para naturalizar o garoto que **se mudou** para o país aos 13, com um contrato de formação com o Real Madrid.
After all, his federation did everything it could to obtain citizenship for the boy who **moved** to the country at the age of 13, with a training contract with Real Madrid.
- (62) Profissionais de todos os tipos vão precisar de novas habilidades para **se orientar** com êxito no meio dessas transformações econômicas.
Professionals of all kinds will need new skills to successfully **navigate** these economic transformations.
- (63) O grupo de trabalho criado pelo Conselho **se reúne** pela primeira vez na próxima semana, na quarta-feira (4/10).
The working group created by the Council will **meet** for the first time next week, on Wednesday (October 4).
- (64) “É impressionante o número de pessoas que estão **se aproximando**”, disse.
“It’s amazing how many people are **joining us**,” he said.
- (65) Não é só a identidade que **se embaralha**, mas a própria percepção do tempo.
It is not only identity that **becomes confused**, but also the very perception of time.
- (66) A cada ano, principalmente com a reforma da Previdência, o rombo **se estende** muito.
Every year, especially with the pension reform, the deficit **grows** significantly.
- (67) São universos que não **se misturam**.
These are universes that do not **mix**.
- (68) Como uma novela de TV brasileira, são vários personagens, em núcleos que **se misturam**.
As in a Brazilian TV soap opera, there are several characters in **interconnected** groups.
- (69) Os destinos de ambos **se separaram** no Equador.
The two men’s paths **split** in Ecuador.
- (70) Desde o início das investigações, de fato, as arestas na colaboração premiada **se somavam**.
Since the beginning of the investigations, in fact, the rough edges in the plea bargain were **piling up**.
- (71) No filme, que não situa a época em que **se passa**, os alvos políticos são resumidos num dos personagens, um líder da luta por terras.
In the film, which does not specify the period in which it **takes place**, the political targets are summarized in one of the characters, a leader in the struggle for land.

- (72) A novela **se passa**, principalmente, no Tocantins, incluindo sua capital, Palmas.
The soap opera is mainly **set** in Tocantins, including its capital, Palmas.
- (73) Isso pode **se dar** de várias maneiras, que incluem a participação em projetos de pesquisa e a colaboração com cientistas.
This can **be achieved** in a variety of ways, including participating in research projects and collaborating with scientists.
- (74) Há três anos, **abalou-se** com a morte do pai, Patrocino, após um AVC.
Three years ago, he/she **was devastated** by the death of his/her father, Patrocino, after he suffered a stroke.
- (75) Ela **se emocionava** extraordinariamente com qualquer coisa.
She used to **get** unusually **emotional** about anything.
- (76) Os ânimos **se acalmaram** dias depois.
Tempers **calmed down** a few days later.
- (77) Quando a reportagem perguntou por que ele precisava da informação, ele **se atrapalhou**.
When the reporter asked him why he needed the information, he **fumbled**.
- (78) O São Paulo **se enervou** em campo, mas quase venceu nos acréscimos com cabeçada de Militão.
São Paulo **got nervous** on the pitch, but almost won it in injury time with a header from Militão.
- (79) Juan não parece **se intimidar** – viu de perto a guerra civil de Angola, em 1985, e ajudou o Haiti durante a epidemia de cólera e o terremoto de 2010.
Juan does not seem to be **intimidated** – he witnessed Angola’s civil war up close in 1985 and helped Haiti during the cholera epidemic and the 2010 earthquake.
- (80) Alckmin **se irritou** e disse que o “novo” na política é “falar a verdade”.
Alckmin **became irritated** and said that what is “new” in politics is “telling the truth.”
- (81) Portanto, não **se surpreenda** se aparecerem cada vez mais atores chineses ou sino-americanos em blockbusters hollywoodianos.
So don’t **be surprised** if you see more and more Chinese or Chinese-American actors appearing in Hollywood blockbusters.
- (82) Seria um baque nas contas públicas, **queixaram-se** os economistas.
It would be a blow to public finances, economists **complained**.
- (83) A verdadeira questão é: por que um periódico supostamente científico **se presta** a isso?

The real question is: why would a supposedly scientific journal **lend itself** to this?

- (84) Sobre a relação com a PF, sua assessoria diz que a PGR vai **se basear** em “diálogo, respeito mútuo, valorização do trabalho de investigação da polícia e cooperação”.
Regarding the relationship with the Federal Police, his advisors say that the Attorney General’s Office will **be based on** “dialogue, mutual respect, appreciation of the police’s investigative work, and cooperation.”
- (85) Mas deu espaços à equipe de Cuca, escalada para **se aproveitar** disso.
But he gave space to Cuca’s team, which was set up **to take advantage** of that.
- (86) **Desculpou-se** pela confusão.
He/she **apologized** for the confusion.
- (87) Sua esperança é que o eleitor gaste tempo para **se informar** e “ir além do discurso”.
His hope is that voters will take the time **to get informed** and “go beyond the rhetoric.”
- (88) O senado vai **se posicionar** sobre essa decisão.
The Senate will **take a position** on this decision.
- (89) Quando um dos guerrilheiros **se rendeu**, ele gritou: “Sou Che Guevara, e valho mais a vocês vivo do que morto”.
When one of the guerrillas **surrendered**, he shouted: “I am Che Guevara and I am worth more alive than dead.”
- (90) Existe expectativa de que o presidente **se reúna** ainda neste sábado com Mouco e Moreira Franco.
The president is expected **to meet** with Mouco and Moreira Franco this Saturday.
- (91) Assim como Gustavo Franco, **aproximou-se** do Partido Novo.
Like Gustavo Franco, he **joined** the New Party.
- (92) Em 2011, Twain **se casou** com Frédéric Thiébaud, o ex-marido da amiga com quem seu ex a traiu.
In 2011, Twain **married** Frédéric Thiébaud, the ex-husband of the friend with whom her ex had cheated on her.
- (93) Maldonado havia **se juntado** a um protesto de indígenas mapuche que reivindicam um território na Província de Chubut, no sul do país, hoje sob posse da companhia italiana Benetton.
Maldonado had **joined** a protest by Mapuche indigenous people claiming territory in the province of Chubut, in the south of the country, currently owned by the Italian company Benetton.
- (94) Os capítulos **se desdobram** em conflitos psicológicos da protagonista e as novas funções.

The chapters **unfold around** psychological conflicts of the protagonist and her new roles.

- (95) Esse sistema **baseia-se nas** entregas urbanas efetuadas por pessoas, caminhando, de bicicleta ou mesmo por outros meios de transporte disponíveis, a preços razoáveis e com redução de trânsito e poluição.
This system is **based** on urban deliveries made by people on foot, by bicycle, or even by other available means of transportation, at reasonable prices and with reduced traffic and pollution.
- (96) Isso **se aplica** a inclinações políticas, gostos artísticos, preferências sexuais.
This **applies** to political leanings, artistic tastes, and sexual preferences.
- (97) A classificação da equipe para a Copa do Mundo de 2018 **se apoia** nisso.
The team's qualification for the 2018 World Cup is **based** on this.
- (98) Não se descarta que parte do aumento **se deva** a um incremento na notificação.
It cannot be ruled out that part of the increase is **due** to an increase in reporting.
- (99) O resultado é que a sua participação no bolo total das despesas da União praticamente **se manteve** estável.
The result is that its share of total Union expenditure has **remained** practically stable.
- (100) **Tornou-se** praxe entre as seguradoras oferecer auxílio-residência, que inclui serviços como eletricista e encanador.
It has **become** common practice among insurance companies to offer home assistance, which includes services such as electricians and plumbers.
- (101) **Encontrava se** no país em sua tentativa de exportar a revolução socialista ao mundo.
He/she **was** in the country in his/her attempt to export the socialist revolution to the world.
- (102) Muitos dos 60 pastores reunidos ali **se diziam** ex-militares e exigiram do deputado que mostrasse como faria para criar um governo “firme”.
Many of the 60 pastors gathered there **claimed to be** former military personnel and demanded that the congressman show how he would create a “firm” government.
- (103) Uma porta-voz da Qiaobi, Xu Chunyan, não **se demonstrou** arrependida.
A spokesperson for Qiaobi, Xu Chunyan, **showed** no remorse.
- (104) Chamou também altos comandos e veteranos do Exército boliviano, que **se mostraram** divididos quanto à celebração.
He also summoned senior commanders and veterans of the Bolivian Army, who **were** divided over the celebration.
- (105) Mas falta recheio para a mistura, que acaba **se mostrando** um pastelão de vento.

But there is no filling for the mixture, which ends up **being** a windy pastry.

- (106) Ele vai se esforçar sozinho e, **sabendo-se** aluno da USP, ganhará orgulho misterioso e estudará sozinho.
He will make an effort on his own and, **knowing** that he is a student at USP, he will gain a mysterious pride and study on his own.
- (107) Já o americano explosivo **se prova** na trama muito mais tático e focado do que seus acessos de raiva deixavam transparecer.
The explosive American, on the other hand, **proves to be** more tactical and focused in the plot than his outbursts of anger would suggest.
- (108) Se alguém **se sentiu** ofendido, peço perdão, afirmou o jogador do Corinthians.
If anyone **felt** offended, I apologize, said the Corinthians player.
- (109) Na visão de Mourão, eleições mostram que população **se deixou** levar por promessas fáceis.
In Mourão's view, the elections show that the population **let themselves** be carried away by easy promises.
- (110) Em uma época em que se precisa diminuir e redistribuir o bolo, a voz dos mais vulneráveis não **se faz** ouvir.
At a time when the pie has to be cut smaller and shared differently, the voices of the most vulnerable **go** unheard.
- (111) Estava na Califórnia **se formando** mestre-ervejeira quando soube da compra da Colorado pela Ambev.
She was in California **graduating** as a master brewer when she heard about Ambev's purchase of Colorado.
- (112) Em 2005, a empresa passou a **se chamar** JBS (iniciais do nome de Zé Mineiro) e iniciou a internacionalização, com a compra da Swift Argentina.
In 2005, the company **changed its name** to JBS (initials of Zé Mineiro's name) and began to internationalize with the purchase of Swift Argentina.
- (113) Em artigo, escreveu que **se sentia** perplexo e amedrontado, submetido a humilhação e vexame sem precedentes.
In an article, he wrote that he **felt** perplexed and frightened, subjected to unprecedented humiliation and vexation.
- (114) Sob esse aspecto, o exemplo dos terraplanistas **se afigura** quase benigno.
In this respect, the example of the flat-earthers **seems** almost benign.
- (115) Fãs de Madonna **se concentram** em frente ao Hotel Copacabana Palace.
Madonna's fans **gather** in front of the Copacabana Palace Hotel.

- (116) Áreas de lazer aumentam em São Paulo, mas ainda **se concentram** em regiões nobres.
Leisure areas are increasing in São Paulo, but they are still **concentrated** in wealthy neighborhoods.
- (117) Sobre a relação com a PF, sua assessoria diz que a PGR vai **se basear** em “diálogo, respeito mútuo, valorização do trabalho de investigação da polícia e cooperação”.
Regarding the relationship with the Federal Police, his advisors say that the PGR will **be based** on “dialogue, mutual respect, appreciation of police investigation work and cooperation.”
- (118) A euforia **se baseia** na valorização de 276% da divisa no ano.
The euphoria is **based** on the 276% appreciation of the currency in the year.
- (119) O ex-ministro terá prazo para apresentação de defesa caso pretenda **manter-se** filiado ao partido.
The former minister will have a deadline to present his defense if he intends **to remain** affiliated with the party.
- (120) O resultado é que a sua participação no bolo total das despesas da União praticamente **se manteve** estável.
The result is that its share of total Union expenditure has **remained** practically stable.
- (121) Quando **se apresentou** ao Santos, Nilmar não disputava uma partida oficial havia 14 meses.
When he **joined** Santos, Nilmar had not played an official match for 14 months.
- (122) Chiara Santoro e Daniel Gonçalves **apresentam-se** no Música ao Meio-Dia.
Chiara Santoro and Daniel Gonçalves **perform** at Música ao Meio-Dia.
- (123) Eles também possuem um maior número de valvas no seu interior, os vasos linfáticos **apresentam-se** dilatados e exibem um aspecto nodular ou em colar de contas.
They also have a greater number of valves inside them, the lymphatic vessels **are** dilated and have a nodular or beaded appearance.
- (124) **Apresentam-se**, a seguir, dois símbolos utilizados em desenhos topográficos.
In the following, we **present** two symbols used in topographic drawings.