

NUMBER AGREEMENT IN BRAZILIAN PORTUGUESE: A FORMAL ANALYSIS IN THE FRAMEWORK OF LEXICAL- FUNCTIONAL GRAMMAR

A CONCORDÂNCIA DE NÚMERO NO PORTUGUÊS BRASILEIRO: UMA ANÁLISE FORMAL
NO QUADRO NA GRAMÁTICA LÉXICO-FUNCIONAL

LA CONCORDANCIA DE NÚMERO EN PORTUGUÉS BRASILEÑO: UN ANÁLISIS FORMAL EN
EL MARCO DE LA GRAMÁTICA LÉXICO-FUNCIONAL

Jessé de Sousa Mourão*

Universidade Estadual Vale do Acaraú

ABSTRACT: In this paper, I propose a formal description and analysis of verb and noun agreement in Brazilian Portuguese (BP). The formal account allows us to analyze, in the same computational grammar, agreement patterns from standard and popular varieties of BP. For that, the main types of agreement were identified and described based on agreement theories (STELLE, 1978; CORBETT, 2003; 2006). The formal account was based on Lexical-Functional Grammar (DALRYMPLE, HRISTOV, 2010; DALRYMPLE; WECHSLER; ZLATIĆ, 2000; 2003). In this work, it was assumed that BP uses two levels of lexicogrammatical features for number agreement: morphosyntactic features (CONCORD) and semantic features (INDEX), which license alternative and varied agreement patterns. The formalization was implemented computationally through the XLE system, and tests were carried out with exemplary sentences for the modeled patterns. As a result, the grammar assigned to the sentences the expected analyses, licensing the acceptable patterns and blocking the unacceptable ones.

KEYWORDS: Noun and Verb Agreement. Brazilian Portuguese. Lexical-Functional Grammar.

RESUMO: En este artículo, propongo un análisis y descripción formal de la concordancia nominal y verbal de número en portugués brasileño (PB). La formalización permite, en una misma gramática computacional, el análisis de patrones de concordancia de las variedades populares y cultas del PB. Para esto, los principales tipos de concordancia han sido identificados y descritos con base en teorías de concordancia (STELLE, 1978; CORBETT, 2003; 2006). El tratamiento formal se ha basado en la Gramática Léxico-Funcional (DALRYMPLE, HRISTOV, 2010; DALRYMPLE; WECHSLER; ZLATIĆ, 2000; 2003). En esta descripción, se asumió que el PB emplea dos niveles de rasgos léxico-gramaticales para la concordancia de número: rasgos morfosintácticos (CONCORD) y rasgos semánticos (INDEX), responsables por licenciar patrones variados y alternativos de concordancia. La formalización se ha implementado computacionalmente a través del sistema XLE, y se realizaron pruebas utilizando sentencias ejemplares de los patrones modelados. Como resultado, la gramática atribuyó a las sentencias los análisis esperados, licenciando los patrones aceptables y bloqueando los patrones inaceptables.

PALAVRAS-CHAVE: Concordância Nominal e Verbal. Português Brasileiro. Gramática Léxico-Funcional.

*Holds a Masters degree in Applied Linguistics from State University of Ceará (UECE) and a PhD in Linguistics from Federal University of Ceará (UFC). Assistant professor in the Letras Course at the State University of Vale do Acaraú (UEVA), carrying out teaching and research activities in the following subjects: English Teaching, Computational Linguistics and Generative Grammar. Email: jesse_mourao@uvanet.br.

RESUMEN: En este artículo, propongo un análisis y descripción formal de la concordancia nominal y verbal de número en portugués brasileño (PB). La formalización permite, en una misma gramática computacional, el análisis de patrones de concordancia de las variedades populares y cultas del PB. Para esto, los principales tipos de concordancia han sido identificados y descritos con base en teorías de concordancia (STELLE, 1978; CORBETT, 2003; 2006). El tratamiento formal se ha basado en la Gramática Léxico-Funcional (DALRYMPLE, HRISTOV, 2010; DALRYMPLE; WECHSLER; ZLATIC, 2000; 2003). En esta descripción, se supuso que el PB emplea dos niveles de rasgos léxico-gramaticales para la concordancia de número: rasgos morfosintácticos (CONCORD) y rasgos semánticos (ÍNDICE), responsables por licenciar patrones variados y alternativos de concordancia. La formalización se ha implementado computacionalmente a través del sistema XLE, y pruebas se han realizado utilizando sentencias ejemplares de los patrones modelados. Como resultado, la gramática ha atribuido a las sentencias los análisis esperados, licenciando los patrones aceptables y bloqueando los patrones inaceptables

PALABRAS-CLAVE: Concordancia Nominal y Verbal. Portugués Brasileño. Gramática Léxico-Funcional.

1 INTRODUCTION

Brazilian portuguese (hereinafter BP) is a heterogeneous language comprising a set of linguistic varieties influenced by different historical, geographic and social factors. Among these varieties, two groups are recognizably distinct: standard brazilian portuguese and popular brazilian portuguese, as defined by Leite (2008) and Holm (2009). The separation between these two varieties is not always so clear, precise and definitive. Therefore, it may be more appropriate to treat standard brazilian portuguese as a group of varieties, which I refer here to as standard varieties; and popular brazilian portuguese as another group, referred to as popular varieties.

One of the aspects that most explicitly distinguish these two large groups of varieties are the patterns of noun and verbal number agreement. Generally speaking, we can say that, for simple (uncoordinated) noun phrases, standard varieties require that determiners and modifiers agree in number (that is, exhibit the same morphosyntactic marks) with the head of the noun phrase, and that the verb agree in number with the subject noun phrase, as in example (1) below, where both determiner and verb display the plural mark, compatible with the mark in the noun/subject.

- (1) *O-s menino-s brinc-aram.*
 The-PL boy-PL play-PST.3PL
 ‘The boys played’

Popular varieties, on the other hand, dispense with this agreement requirement and allow the combination of elements without explicit number marks with marked elements, both at the inner level of the noun phrase and at the level of the clause (in the subject-verb relationship), as in the example (2) below, in which the plural mark was dispensed with in the noun and verb.

- (2) *Os menino-Ø brinc-ou.*
 The-PL boy-PL play-PST.1SG
 ‘The boys played’

Much of the literature that has dealt with this and other phenomena of linguistic variation involves works guided by the sociolinguistic perspective (cf. LUCCHESI, 2017), where the linguistic aspect is just one of the factors under analysis. The broader analysis takes into account several extralinguistic factors, such as social (age, gender, education, etc.), historical, geographic and pragmatic context variables. Although these studies are important and contribute to the understanding of language mainly as a social phenomenon, their theoretical-methodological approach is not intended to investigate the internal functioning of language. Non-transformational, generative-based formal linguistic theories, such as Lexical-Functional Grammar (hereinafter LFG) and Head-Driven Phrase Structure Grammar (HPSG) have had an important impact in this field, as they allow the formal modeling of the grammar of a language and, therefore, the construction of computational grammars, which can be used in several applications and natural language processing tasks.

For BP, we can mention the LFG computational grammars developed by Alencar (2013a) and Santos (2014), which model several non-trivial phenomena and structures of BP, but, due to their initial stage, still leave much of the language uncovered, both in terms of linguistic phenomena and lexicogrammatical coverage. Number agreement is one of those phenomena that needs more in-depth treatment. The grammars mentioned only cover constructions of the type exemplified in (1), the typical patterns of standard varieties, while constructions of type (2), associated with popular varieties, would be treated as ungrammatical by these grammars, considering their current state.

The construction of computational grammars, such as those based on the LFG, is relevant both for its theoretical potential, for the description and explanation of linguistic phenomena, and for its applied potential, by providing a basic and fundamental architecture for the computational processing of a language. Regarding this second aspect, it is important that a grammar be able to recognize and analyze as many constructions and forms of the language and its variants as possible. Hence the importance of modeling not only the standard variety of the language, here BP, but also its popular varieties.

In a broader project, not restricted to BP, Alencar (2013b) deals with a subset of the agreement patterns of Brazilian and European varieties of Portuguese. The analysis proposed by the author represents a relevant contribution to the treatment of constructions that are not restricted to standard varieties, but focuses on constructions that involve verbal agreement (of number, gender and person) with elements considered pronominal in the subject function (eg: *A gente esperamos*. 'We waited')¹. Agreement at a more general level, between noun phrases (not just pronominals) and verbs, with apparently conflicting number marks, is left open for future work. Thus, the present work is in line with that of Alencar (2013b), but restricts the analysis to the two great varieties of BP and the phenomenon of number agreement, aiming to expand the coverage of a computational LFG grammar of BP. For this, I propose a formal description of the nominal and verbal number agreement phenomenon in BP that allows, in the same computational grammar, the analysis and generation of constructions of the two linguistic varieties.

2 NOMINAL AND VERBAL NUMBER AGREEMENT IN BP

The phenomenon of agreement in BP has been the subject of many sociolinguistic studies (SCHERRE, 1994, 1998; MOREIRA; VIANNA, 2018; LUCCHESI, 2015, 2017; MASCARELLO, 2010; BRANDÃO, 2011; MENDES, OUSHIRO, 2015; to name a few). What these works show is that there is a great variation in the patterns of nominal and verbal agreement in BP, and that such variation is motivated both by linguistic factors (such as the type and order of the elements that form the phrases) and by historical and social factors (such as the influence between languages, the education of speakers, etc.). For this article, I do not intend to discuss and analyze the reasons that underlie this variation. Actually, these patterns are assumed to be part of the language, legitimized by users of standard or popular varieties, susceptible to and lacking formalization.

Based on linguistic examples from *corpora* and my intuition as a BP speaker, it is possible to identify some agreement patterns that will be exemplified below².

At the internal level of the noun phrase, the identified patterns are:

I. Presence of a plural mark in all inflectionable elements that make up the noun phrase:

- (3) (a) *os fregueses* (All examples from Scherre (1994))
 'the clients'
 (b) *novas escolas*
 'new schools'

¹ *A gente* in this example is an expression morphosyntactically singular, but equivalent to the 1st person plural personal pronoun *nós* 'we'.

² Although this study is related to Variationist Sociolinguistics, as it deals with alternative patterns for the same linguistic structure, performance aspects of the BP speaker (such as age, memory, gender, social context, etc.) were not considered in the study. The structures were analyzed solely from the point of view of the speaker's competence.

(c) *as boas ações*

‘the good actions’

(d) *essas coisas todas*

‘all those things’

(e) *as partidas todas iguais*

‘the matches all equal’

II. Presence of a plural mark in some of the inflectionable elements of the noun phrase:

(4) (a) *essas estradas nova* (SCHERRE, 1994; exemplos a-c)

these roads new.F.SG

‘these new roads’

(b) *do meus pais*

of.M.SG my.M.1PL parents

‘of my parents’

(c) *as minhas duas filha*

the.F.PL my.F.1PL two daughters

‘my two daughters’

(d) *Os meus filho* (MOREIRA; VIANNA, 2018)

the.M.PL my.M.1PL son

‘my sons’

(e) *Os trãos elétrico* (LUCCHESI, 2015)

the.M.PL trucks electric.M.1SG

‘the sound trucks’

III. Presence of a plural mark in only one of the elements of the noun phrase:

(5) (a) *as codorna* (SCHERRE, 1994; exemplos a-c)

the.F.PL quail.F.SG

‘the quails’

(b) *as porta aberta*

the.F.PL door.F.SG open.F.SG

‘the open doors’

(c) *seus momento bom*

your.2PL moment good.SG

‘your good moments’

(d) *as primeira mudança* (MOREIRA; VIANNA, 2018)

the.F.PL first.SG change.SG

‘the first changes’

(e) *outras pessoa* (MASCARELLO, 2010)

other.F.PL person

‘other people’

IV. Absence of plural mark in all inflectionable elements of the noun phrase:

(6) (a) *dois risco verde* (all examples from Scherre (1994))

two scratche.SG green.SG

‘two green scratches’

(b) *uma porção de coisa interessante*

a bunch of stuff.SG interesting.SG

‘a bunch of interesting stuff’

At the level external to the noun phrase, in the relation between the verb and the subject, there is also variation, and the verb may or may not have the plural mark, regardless of whether all elements of the noun phrase have this mark or not, as we can see in the following examples:

- (7) (a) *As minhas filha era pequena.* (MOREIRA; VIANNA, 2018)
 the.F.PL my.F.PL daughter was young.F.SG
 'My daughters were young.'
- (b) *As menina fazia, (...).* (LUCCHESI, 2015; examples b-d)
 the.F.PL girl.F.SG made.3SG
 'The girls used to make, (...)'
- (c) *Os tróus elétrico saía doze hora.*
 the.M.PL trucks electric.M.SG left.3SG twelve hour.SG
 'The sound trucks left at twelve.'
- (d) *As minhas irmã são moderna.*
 the.F.PL my.F.PL sister.SG be.3PL modern.F.SG
 'My sisters are modern.'
- (e) *As menina tiraram retrato.* (BRANDÃO, 2011)
 the.F.PL girl.SG took.3PL picture
 'The girls took pictures.'
- (f) *Eles conhece Roma.* (SCHERRE, 1998; exemplos f-g)
 they.M.PL know.3SG Rome
 'They know Rome.'
- (g) *Os filho tá pedindo dinheiro.*
 the.M.PL son;child is asking.for money
 'The sons are asking for money. / The children are asking for money.'
- (h) *Outras pessoa vai modificando a comunidade.* (MASCARELLO, 2010)
 other.F.PL person go.3SG modifying the.F.SG community
 'Other people change the community.'

Another variation is observed with collective nouns in the subject function, as in the examples below, in which the verb may or may not exhibit the plural mark:

- (8) (a) *A -gente faz.*³
 we do.3SG
 'We do.'
- (b) *A -gente fazemos.* (MENDES, OUSHIRO, 2015)
 we do.1PL
 'We do.'
- (c) *O povo chegou.*
 the.M.SG people.M.SG arrived.3SG
 'People arrived.'

³ Currently there is a very consensual conception that the expression "a gente" in Brazilian and European Portuguese, as in examples 8a and 8b, went through a grammaticalization process and is already considered a first-person plural personal pronoun equivalent to the pronoun "Nós" (we), and not a noun phrase formed by a determiner and a noun (see ALENCAR, 2013b; ZILLES, 2005; LOPES, 2003; MARCOTULIO; VIANNA; LOPES, 2013). I agree with this analysis. In the formal treatment and implementation proposed here, to simplify the analysis, the expression "a gente" is still analyzed by means of the noun phrase rule. In the future, it will be necessary to correct or expand the lexicogrammatical properties of this expression so that it is filtered by the rule of a determiner phrase formed only by a pronoun. For this, it will be necessary to include the entire expression as a single lexical entry of the category PRO (pronoun), and assign to it the specifications suggested in this paper for noun heads of the type "povo". With this, "a gente" will be properly analyzed as a pronominal expression instead of nominal expression.

- (d) *O povo chegaram.*
 the.M.SG people.M.SG arrived.3PL
 'People arrived.'

Returning to what I mentioned earlier, some of these patterns are more common in certain varieties of BP. The examples in (1), (3) and (8a) are typical of standard varieties, while the others are more typical of popular varieties. In general, we find that the former employ a more redundant pattern of number agreement: if the head of the subject noun phrase displays a plural mark, then all determiners and inflectionable modifiers, as well as the main verb in the head of the verb phrase, must also display it. In the case of popular varieties, there is less redundancy, since a single explicit plural mark in one of the elements of the noun phrase (usually the determiner) is sufficient for the entire phrase and even the entire sentence.

Given the richness and variety of the language, other patterns must exist, but the above are the main patterns of number agreement that will be considered in this study for purposes of description, analysis and formalization.

3 THEORETICAL CONSIDERATIONS ON (NUMBER) AGREEMENT

A general conception of agreement, adopted here, is given by Steele (1978, p. 610, my translation), according to which agreement can be understood as "[...] a systematic covariance between a formal or semantic property of an element with the formal property from another". Corbett (2006) points out that the crucial point in Steele's conception is the notion of systematic covariance, that is, linguistic elements carry features that coincide with the features of other elements, but this coincidence and its possible variations are systematic: if an element varies in one aspect, another one also varies in that same aspect.

Several elements participate in this covariance and determine how the concordance relationship can occur. Corbett (2006), for example, understands agreement as a phenomenon that involves the following components: controller, target, domain, features and conditions. The controller is the element that determines agreement (like the subject noun phrase, which determines agreement with the verb, i.e., requires some specific verb form). The target is every element whose form is determined by the controller, such as the main verb (in a sentence) which varies its form depending on the subject. The domain is related to the syntactic environment in which the agreement relationship takes place (such as the noun phrase or the clause). The properties involved in agreement are the features (such as gender, number, and person). And conditions are properties (such as animacy, precedence, etc.) that do not have a formal manifestation, but act on controllers, targets and domains, and may have some effect on agreement.

All these elements are manifested and interact in different ways depending on the language. Thus, it is necessary to identify, for each language, which linguistic elements assume each of these roles, and how they interact to derive acceptable agreement patterns. At the same time, it is necessary that general principles be identified in order to compose a general theory of agreement applicable to any language. Although this work departs from more general theoretical conceptions of agreement, the main focus is the description of the particular properties of a language (BP).

The understanding that there is an element that controls and another that is the target of agreement is part of an asymmetric agreement conception (CORBETT, 2006). Asymmetry is suggested when one understands that the elements that converge in their grammatical properties and features do not play exactly the same role in this relationship. Consider the following sentence:

- (9) *Toda-s a-s criança-s pequena-s dorme-m.*
 all-PL the-PL child-PL young-PL sleep.3-PL
 'All the young children sleep.'

In this sentence, all the words display the same grammatical mark. The asymmetry is understood when it is assumed that, in a domain of agreement, some features are attributed directly by the lexical or syntactic item itself, while others are determined by the features present in another lexical or syntactic element. In this sense, in sentence (9), it can be understood that the plural morphemes

in the noun phrase domain (*todas as crianças* ‘all children small’) were determined by the head *crianças* (children), signaling a plural referent, thus requiring that the other determiners, quantifiers and modifiers specify the same feature. The verb form *dormem* (sleep) is also determined by the noun, or more abstractly by the subject. Based on this, an initial conception is that in BP the elements which impose the agreement control are the head of the noun phrase and the subject of the sentence⁴. More precisely, we can say that in the noun phrase domain, the controller is the noun, and the other elements are the targets; whereas, in the domain of the clause (in the subject-predicate relation), the controller is the subject, and the verb is the target. The subject can be treated as a controller, because there are cases where the form of the verb is not determined by the noun, but by the whole subject; as in the case of subjects with coordinated singular noun phrases - see example (10) below -, where the verb does not agree in number with any of the individual nouns, but with the group, which constitutes the entire subject noun phrase. Since the roles of these elements are distinct, and there is a dependency relationship between them, this agreement is considered as asymmetric and unidirectional.

- (10) *A mãe e o filho desapareceram.*
 the mother and the son disappeared.3PL
 ‘The mother and the son disappeared’

In a symmetric conception, there would be no such unidirectionality in agreement. Rather, it is assumed to be multidirectional: features of lexical and syntactic components are assigned more or less independently by each item, and both controller and target features may have similar roles in agreement: both may contribute information to the syntactic structures.

The discussion about the asymmetry and symmetry of agreement is quite varied in the literature, both in the treatment of different languages and across theories, and is therefore not consensual. Corbett (2006), for example, argues that canonical agreement is asymmetric. Ackema and Neeleman (2013), Kathol (1999), Haug and Nikitina (2016) defend the symmetric character of agreement. In the analysis proposed here, the asymmetric conception of agreement, according to Corbett (2006), is not discarded. However, I argue that this asymmetry is not produced solely and directly at the level of formal expression, but in the relation of a semantic-referential level with the formal level. In this sense, directionality is not *formal expression* (controller) → *formal expression* (target), but *semantic-referential entity* (controller) → *formal expression* (target). In the sentence in (9), for example, it is not exactly the form of the noun that determines the form of determiners and modifiers, since the form of the noun itself has already been selected to express and be compatible with the plural referent of that noun. In BP, there are at least two lexical forms available for the lexeme CRIANÇA (CHILD): *criança* (child) e *crianças* (children). The choice for one of the forms depends on the referent, whether it denotes only one or more than one individual⁵. In this case, even the noun can be considered a target. Another argument would be the example in (10), in which the coordinated noun phrase *a mãe e o filho* (the mother and the son) has a singular morphosyntactic expression, but denotes more than one individual, and it is this plural referent that controls the form of the verb, inducing the selection of its plural form.

Due to this conception, asymmetry, in this work, will not be modeled directly, through the manipulation of the morphosyntactic features provided by the lexical elements. This means that the modeled agreement system will not capture asymmetry, but only the specifications of the agreement features that each element projects in the syntactic structure, responsible for generating valid syntactic constructions compatible with their referents. This type of treatment, commonly adopted in works within the framework of the LFG, is even referred to as a symmetric analysis of agreement, since agreement is conceived as a procedure of partial accumulation of information contributed by the elements of this relationship and unified in more complex syntactic structures (DALRYMPLE, MYCOCK, 2019; HAUG; NIKITINA, 2016).

The choice of this type of analysis is also motivated by the variability of agreement patterns and relationships between elements traditionally considered controllers and targets. In many cases, and across languages, the relationship between controller and target

⁴ Pronouns are special elements, since they can take on the role of both controller and target, as shown by Bresnan et al. (2016, p. 191), but they will not be discussed here.

⁵ A singular noun can also denote a class of individuals, as in the sentence *The child has their rights guaranteed by the constitution*, pointed out by Leonel Figueiredo de Alencar in a personal communication. Even in this case, the selected form is singular, since the referent can be conceptualized as singular: the class of individuals.

does not necessarily involve the absolute determination of features of agreement, in the sense that whenever there is a controller and a target, the target must obligatorily display the same features as the controller, and vice versa. Thus, the asymmetrical relationship is not so clear and homogeneous at the level of linguistic expression, as I exemplify below.

There are cases where targets don't formally specify or display the features specified by their controller, or may even display features that aren't present in the controller, as in the examples below:

- (11) (a) Walman (CORBETT, 2006, p. 171)

Pelen y-aikiri.

Dog PL-barked.

'The dogs barked.'

- (b) Spanish (BOSQUE; MORENO 1984, apud JOHNSON, 2014, p. 54)

Los jugadores queremos ir a París

the players.PL want.1PL to-go to Paris

'We the players want to go to Paris'

- (c) *Nós estamos cansadas. / Nós estamos cansados.*

we be.1PL tired.F.PL / we be.1PL tired.M.PL

'We are tired.'

- (d) *Precisamos de ajuda.*

need.1PL of help.

'We need help.'

In (11a), in the Walman language (a language spoken in Papua New Guinea), the controller *pelen* (dog) has no formal number mark, which only appears in the target verb *yaikiri* (bark). In the Spanish example (11b), the feature mismatch is in the subject-verb domain, in which the verb exhibits a person mark (1st person) different from the one explicitly present in the subject (3rd person). In the Portuguese example (11c), the predicative *cansadas/cansados* (tired) has gender marks that are not present in the subject, which is an underspecified pronoun in terms of gender. And in example (11d), the number and person marks present in the verb did not come from the controller (the subject), which is not even present in the sentence. These examples show that the agreement relationship, or more precisely the relationship between controller and target, is not as direct, superficial or morphosyntactically transparent as it seems in some cases, indicating that the speaker resorts to other types of information (semantic, referential and even pragmatic), beyond formal morphosyntactic properties. There are other cases where targets specify values different from the value assigned to the same feature of the controller, as in examples (4)-(7), already mentioned, and in examples in (12) below:

- (12) (a) This boy and girl are eating a pizza. (KING; DALRYMPLE, 2004)

- (b) Polish (CORBETT, 2003)

Sześć kobiet było smutn-ych

six woman.PL.GEN be.PAST-SG.NEUT sad-PL.GEN

'Six women were sad'.

- (c) *O prefeito e o secretário foram presos.*

the mayor and the secretary were arrested.M.PL

'The mayor and the secretary were arrested.'

In example (12a), from English, the verb is in the plural form, while its subject has only elements with singular value for number, as in example (12c), from Portuguese. The agreement, in this case, was not established individually with any of the syntactic elements that make up the subject, but with the plural meaning denoted by the set, suggesting that the agreement was semantically controlled. In example (12b), from Polish, the verb *było* (to be) has a singular number value, different from the controller *kobiet* (women) which is plural.

Another situation is when, in the same language, the targets have the option of specifying or not the features and values of the controller, thus having alternative patterns of agreement for the same structure, as in the examples in (8), already mentioned, and in (13), below:

- (13) (a) The committee has decided / have decided (CORBETT, 2003, p. 113)
 (b) *A maioria das pessoas (desistiu / desistiram).*
 the majority of people (gave.up.3SG / gave.up.3PL)
 'Most people have given up.'

In the examples in (8) and (13), the verb can agree both in the singular and in the plural with the subject. This usually happens when the noun or the entire noun phrase can denote both an individual and a group, and thus there are two acceptable possibilities of agreement, which can be treated as a case of semantic agreement (CORBETT, 2006).

Examples (11)-(13) are normally treated as a case of mismatch in agreement (CORBETT, 2003, 2006; WECHSLER; ZLATIĆ, 2000; 2003), when the target features and values don't match the controller's; or as a case of a mixed or hybrid agreement system (WECHSLER; ZLATIĆ, 2000; 2003), when agreement relations are not uniformly determined by morphosyntactic features originating in the controllers, and may be influenced by other linguistic dimensions.

All this heterogeneity and formal incompatibility observed both across languages and within a single language seems to come from various linguistic dimensions, which interact with the formal structure of the language, giving rise to apparently incompatible formal superficial relationships. Corbett (2006), among others, talks about how the phenomenon of agreement may involve lexical, syntactic, morphological and pragmatic aspects. It is this multiplicity of sources of (extra)linguistic information that influences the production and understanding of the language and allows the speaker diverse and alternative ways of formally structuring the linguistic expression. Therefore, any agreement theory will not be complete if it bases its premises and principles only on a single level of linguistic information, such as the syntactic or semantic level.

This incompatibility, then, should not be understood as non-agreement. According to Corbett (2006), even when there is some incompatibility between the controller and target features, the relationship should still be considered as agreement. This superficial disparity is a phenomenon that is present in many languages and that produces structures acceptable to speakers of those languages, being therefore licensed by their grammar, and by the agreement system in particular.

Concerning BP, the incompatibility in agreement is also present, as we have already observed, being more evident when we gather and contrast data from more than one language variety. The incompatibility observed in BP points to two types of agreement also identified by Corbett (2006): obligatory agreement and optional agreement. In the first case, agreement must be necessarily marked in the morphology of the flectionable targets; and in the second, this marking is optional. In this sense, standard varieties typically employ obligatory morphosyntactic agreement; while the popular varieties, the optional one. Next, I will present the LFG perspective for the treatment of agreement and a proposal to formalize these types of patterns of number agreement in BP.

3.1 THE FORMALIZATION OF NUMBER AGREEMENT IN PB VIA LFG

Lexical-Functional Grammar (LFG) is a non-transformational generative theory, which consists of a mathematically precise and formal model of grammar, which enables the formalization and computational implementation of grammars and linguistic phenomena of natural languages (see DALRYMPLE; MYCOK, 2019 for a very comprehensive and up-to-date exposition of the theory). LFG is also part of unification-based grammar models⁶, along with HPSG, GPSG, and others.

⁶ See Shieber (2003) for an introduction to unification grammars.

LFG studies also treat the agreement phenomenon as a process of feature co-specification involving a controller and a target (DALRYMPLE, HRISTOV, 2010; BRESNAN *et al.*, 2016). More specifically, LFG models the phenomenon of agreement through the formal mechanism of unification, as shown below, with examples from Portuguese.

In BP LFG grammars (ALENCAR, 2013a; SANTOS, 2014), determiners, nouns and verbs include the following information in their lexical entries (simplified here):

- (14)
- | | | |
|--|--|---|
| $\begin{array}{l} o \quad D \quad (\uparrow GEN) = MAS \\ \quad \quad (\uparrow NUM) = SG \\ \quad \quad (\uparrow SPEC) = DEF \end{array}$ | $\begin{array}{l} cavaleiro \quad N \quad (\uparrow PRED) = 'CAVALEIRO' \\ \quad \quad \quad (\uparrow NUM) = SG \end{array}$ | $\begin{array}{l} passa \quad V \quad (\uparrow PRED) = 'PASSAR < (\uparrow SUBJ) >' \\ \quad \quad \quad (\uparrow SUBJ \quad NUM) = SG \\ \quad \quad \quad (\uparrow SUBJ \quad PERS) = 3 \\ \quad \quad \quad (\uparrow TENSE) = PRES \end{array}$ |
| $\begin{array}{l} os \quad D \quad (\uparrow GEN) = MAS \\ \quad \quad (\uparrow NUM) = PL \\ \quad \quad (\uparrow SPEC) = DEF \end{array}$ | $\begin{array}{l} cavaleiros \quad N \quad (\uparrow PRED) = 'CAVALEIRO' \\ \quad \quad \quad (\uparrow NUM) = PL \end{array}$ | $\begin{array}{l} passam \quad V \quad (\uparrow PRED) = 'PASSAR < (\uparrow SUBJ) >' \\ \quad \quad \quad (\uparrow SUBJ \quad NUM) = SG \\ \quad \quad \quad (\uparrow SUBJ \quad PERS) = 3 \\ \quad \quad \quad (\uparrow TENSE) = PRES \end{array}$ |

Take as an example the first group of entries in (14). When instantiated⁷, they produce the F-structures (functional structures) in (15), below, which merge to produce the resolved functional structure for the sentence *O cavaleiro passa* (The knight passes), given in (16).

- (15)
- | | | |
|---|--|---|
| <i>O</i> | <i>cavaleiro</i> | <i>passa</i> |
| $\left[\begin{array}{l} \mathbf{PRED} \quad MAS \\ \mathbf{NUM} \quad SG \\ \mathbf{SPEC} \quad DEF \end{array} \right]$ | $\left[\begin{array}{l} \mathbf{PRED} \quad 'CAVALEIRO' \\ \mathbf{NUM} \quad SG \end{array} \right]$ | $\left[\begin{array}{l} \mathbf{PRED} \quad 'PASSAR < (\uparrow SUBJ) >' \\ \mathbf{SUBJ \quad NUM} \quad SG \\ \mathbf{SUBJ \quad PERS} \quad 3 \\ \mathbf{TENSE} \quad PRES \end{array} \right]$ |

- (16) *O cavaleiro passa.*

$$\left[\begin{array}{l} \mathbf{PRED} \quad 'PASSAR < SUBJ >' \\ \mathbf{SUBJ} \quad \left[\begin{array}{l} \mathbf{PRED} \quad 'CAVALEIRO' \\ \mathbf{NUM} \quad SG, \mathbf{PERS} \quad 3, \mathbf{SPEC} \quad DEF \end{array} \right] \\ \mathbf{TENSE} \quad PRES \end{array} \right]$$

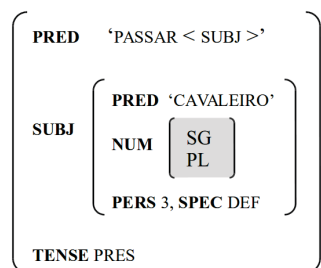
The F-structure in (16) was properly built only because the unification operation was satisfied (in addition to other well-formed properties not relevant to the discussion). In other words, the values for the number feature in the determiner and in the noun are the same (NUM = SG), which allows the unification of the two elements and, consequently, the good formation of the noun phrase. The verb also specifies an SG (representing singular) value for the number (NUM) attribute of the subject, the same value specified by the subject, thus resulting in a fully unified F-structure (with no conflicting values) for the sentence. The same situation would happen with the entries of plural items in (14), which would project a well-formed F-structure for the sentence *Os cavaleiros passam* (The knights pass). We thus have the phenomenon of number agreement captured through unification.

These specifications, however, only deal with canonical cases of morphosyntactic agreement (obligatory agreement). When instantiated in sentences like (17), below, such specifications contribute conflicting information and project the F-structure given in (18). In this structure, we have different values (SG and PL) for the same attribute, which produces a malformed (non-unified) structure, leading the grammar to reject the sentences in (17).

⁷ Instantiation, in LFG, is the process in which functional schemes (specifications) are assigned to syntactic constituents indicated by arrows (\uparrow and \downarrow), producing a set of expressions, called functional descriptions, that inform and guide the construction of functional structures (WESCOAT, 1989).

- (17) (a) *Os cavaleiro passa.*
 the.PL knight.SG pass.3SG
 ‘The knights pass.’
 (b) *Os cavaleiros passa.*
 the.PL knight.PL pass.3SG
 ‘The knights pass.’

- (18) *Os cavaleiro passa. / Os cavaleiros passa.*



Considering that the sentences in (17) are acceptable, at least in the popular varieties of BP, we can infer that there must be other lexicogrammatical properties operating in the language, in addition to those exposed in (15), allowing the grammar to attribute valid analyzes to both cases of obligatory and optional agreement, as exemplified in (17). Next, I present a proposal for the description and formalization of number agreement in BP that can handle these two cases.

4 THEORETICAL-FORMAL PROPOSAL

The proposal to be presented is based on the treatment given by Wechsler and Zlatić (2000, 2003), initially elaborated within the framework of the HPSG, but widely explored by other authors within the scope of LFG (ALENCAR, 2013b; BELYAEV; DALRYMPLE; LOWE, 2015; DALRYMPLE; HRISTOV, 2010; KING; DALRYMPLE, 2004; SADLER, 2003; SADLER; KUHN, 2007). Wechsler and Zlatić (2000; 2003) propose that agreement does not occur only through a single level of features, but through more than one level, which the authors call INDEX and CONCORD, both containing more specific features, such as number, person and gender. As mentioned in Dalrymple and Hristov (2010), these two categories are syntactic properties, however CONCORD is more closely linked to morphosyntactic features and INDEX is more related to semantic features.

By stipulating two levels of agreement, the possibilities of feature co-specification are expanded, and agreement can then be established at more than one level. Thus, when we have constructions with incompatible grammatical properties of agreement at one level, the compatibility may be still ensured at another level. Due to this greater flexibility, this multilevel system of features has been widely used, with good results, to explain the phenomenon of incompatibility or mixed agreement (cf. BRESNAN *et al.*, 2016). Alencar (2013b) follows a similar proposal for the treatment of agreement in BP, distinguishing between semantic features (called SIGMA features) and syntactic features (called PHI features). Here, we chose to use the INDEX and CONCORD features, since they are widely accepted in the LFG scenario.

In the proposal presented here, agreement will be modeled considering only the following syntactic categories: determiners, nouns and verbs. Regarding the domains of agreement, I will only consider the domain of simple noun phrases (not coordinated phrases) and the domain of clauses (subject-verb relationship)⁸. The analysis will also be restricted to grammatical number features, leaving aside for the time being other agreement features, such as person and gender.

⁸ Corbett (2006) identifies at least 4 domains of agreement: a) the internal domain of the noun phrase; b) the domain external to the noun phrase, but still within the sentence; c) the domain that extends beyond the clause, but still within the level of the sentence; and d) domains that can go beyond the sentence level.

The first claim to be made is that the BP agreement system makes requirements for both INDEX and CONCORD features. This means that lexical items can specify these two sets of features in their entries simultaneously. Therefore, for this variety of Portuguese, there will be two levels at which agreement can be satisfied, allowing varied and alternative patterns of agreement.

This agreement system in BP is even introduced by Dalrymple and King (2004), who suggest the following specification of number features for singular and plural determiners:

- (19) (a) Singular determiners:
 (↑ CONCORD NUM) = SG
 (↑ INDEX NUM) = SG
 (b) Plural determiners:
 (↑ CONCORD NUM) = PL
 (↑ INDEX NUM) = PL

My proposal follows this system, but with some adjustments and additional developments. Dalrymple and King (2004) do not show how the specification would be set for other categories (such as nouns, adjectives and verbs); moreover, the specification given in (19) does not seem entirely adequate, as it would exclude constructions of the popular varieties of BP, as *o povo chegaram* (People arrived.), in which the SG value of INDEX would conflict with the INDEX value of *povo* (people), which, according to my analysis, must be plural (PL), which is evidenced with the plural agreement between the verb and the subject.

With this and other accounts, I propose the following feature number agreement specification for the categories determiners, nouns and verbs in BP⁹:

Determiners

- (a) Singular determiners:
 (↑ CONC NUM) = SG
 (b) Plural determiners:
 (↑ CONC NUM) = PL
 (↑ INDEX NUM) = PL

The specification of the determiners is similar to that of Dalrymple and King (2004), but is different in the information of the singular determiner, having only CONC(ORD) features for number. The plural determiner includes both levels CONCORD and INDEX, with a redundant specification of the PL (plural) value for NUM. The distinction between the two types comes from the fact that the plural determiner is more restrictive, hence imposing its number features on both levels. The singular determiner, on the other hand, specifies only one level of agreement (CONCORD), with no restriction on the INDEX level, which allows it to be combined with nouns with an INDEX feature, regardless of its value. With this, the grammar can provide valid analyzes for sentences of the type in (8), repeated below in (20), which have a morphosyntactically singular noun phrase, but semantically plural, as can be seen from the plural verb agreement.

- (20) (b) *A -gente fazemos.* (MENDES, OUSHIRO, 2015)
 we do.1PL
 ‘We do.’
 (b) *O povo chegaram.*
 the.M.SG people.M.SG arrived.3PL
 ‘People arrived.’

⁹ I am grateful to Leonel Figueiredo de Alencar for his comments and revision of the article, which helped mainly in improving the implemented lexical specifications, eliminating the ambiguities of spurious analyzes that an initial version of the proposal still had.

Now let's see the specification of number features for nouns:

Nouns

Type 1 (nouns with plural denotation and singular form (ex.: *povo, gente, galera* 'people')):

(a) Singular nouns:

@(DEFAULT (↑ CONC NUM) SG)

(↑ INDEX NUM) = PL

Type 2 (All other nouns)

(a) Singular nouns:

@(DEFAULT (↑ CONC NUM) SG)

(b) Plural nouns:

(↑ CONC NUM) = PL

(↑ INDEX NUM) = PL

For nouns, it is necessary to stipulate at least two sets of distinct features. The first set corresponds to a subgroup of nouns generally referred to as collective nouns (Type 1), understood as nouns that denote a collection or a group of individuals, despite their singular forms (SCHWARZSCHILD, 1996)¹⁰. The second set corresponds to all other nouns (Type 2).

Let's start the analysis with the group of Type 2, which represents the most general class and with the largest number of elements. For these nouns, the suggested specification also includes the two levels of agreement (CONCORD and INDEX), but the way this set of features is specified differs between singular and plural nouns.

For the singular form, these nouns only specify a default number assignment. This assignment is encoded in the @(DEFAULT (↑ CONC NUM) SG) specification, which represents the call to the default assignment template, presented in (21a) and which unfolds in the specifications in (21b):

- (21) (a) DEFAULT (D V) = { D D~V | D=V }
 (b) {(↑ CONC NUM) (↑ CONC NUM) ~ SG} | (↑ CONC NUM) = SG }

The result of this assignment is: if no other noun phrase element specifies a different value for the feature CONC NUM, the noun specifies the SG value for that feature. Through this assignment, this type of noun (with singular form) can receive a plural value from another element, such as, for example, the plural determiner, or it will have a singular value at that level, compatible with elements that also have this value. And, if there is no other element assigning values to the CONC NUM feature, as when the noun phrase is made up of a single word (a noun), it will assign the value SG.

In order to better exemplify these properties, take the following examples:

- (22) (a) *a codorna*
 the.F.SG quail.F.SG
 'the quails'
 (b) *as codorna* (SCHERRE, 1994)
 the.F.PL quail.F.SG
 'the quails'

¹⁰ There is no consensual definition for the class of collective nouns in linguistic literature, as shown by Gardelle (2019). Thus, we are assuming this more general conception of collective noun, but restricting the proposed specification to a subgroup of these nouns. That is, not all nouns with a singular form that denote a plural referent will necessarily specify the suggested features. The specification is therefore idiosyncratic: those collective nouns that license agreement both morphosyntactically (in the singular) and semantically (in the plural) will have this property specified in their lexical entry.

In (22a), the noun specifies the value SG by default, since the determiner does not assign a different value. In example (22b), on the other hand, the noun does not receive the value SG by default, but the PL value assigned by the determiner instead.

With these properties I claim that singular nouns have a singular morphosyntactic form, formally encoded at the CONCORD level ((↑ CONC NUM) = SG). This allows them to be combined with determiners with the same property, which happens in example (22a). They do not, however, have number information at a level closer to the semantic one, requiring that this information come from elsewhere. In (22b), for example, the interpretation of a plural referent for the item *codorna* (meaning more than one quail) does not come from the noun itself, but from the determiner that assigns plural information to the noun phrase. We can say that the singular noun in BP, with the exception of collective nouns, is underspecified for the number value at the semantic level. To make this formalization even clearer, I give next the informal interpretation of the functional scheme of the singular noun. The noun says to the grammar: I can assign a singular number value to my functional structure through my singular form, or I can say nothing about my number and require it to be given by another element within the same functional structure which I'm part of.

Plural nouns, on the other hand, function similarly to determiners. Both are more restrictive and require two levels of agreement (CONCORD and INDEX), with the same value for the number feature. This captures the intuition that nouns with a plural form are morphologically marked with a plural feature besides denoting plural entities. Thus, plural nouns can only be combined, in the noun phrase domain, with elements that specify the same set of features and values or with elements that specify at least one of the compatible features. As a result, constructions such as those in (23), below, are blocked by the grammar, since agreement is not satisfied at any of the levels.

- (23) * *a codornas*
the.F.SG quail.F.PL

Let us now move on to Type 1 nouns, which represent a smaller class of elements, a subgroup of collective nouns.

The specifications of these nouns have different information from the other group only in the lexical entry of their singular forms. At the CONCORD level, these nouns have the same default assignment as other singular nouns. Also, this type of noun specifies the PL value for the INDEX NUM level. So, with this type of noun, it is possible to have a noun phrase with different number values: SG, at the CONCORD level, and PL at the INDEX level. This mixed specification captures the notion that such nouns have a singular form that can denote both an entity and more than one individual. The singular denotation is captured at the CONCORD level; and the plural denotation, at the INDEX level. With this specification, the constructions in (8), repeated below as (24), are licensed by the grammar.

- (24) (a) *A gente faz.*
(b) *A gente fazemos.* (MENDES, OUSHIRO, 2015)
(c) *O povo trabalha.*
the.M.SG people.M.SG work.3SG
'People work.'
(d) *O povo trabalham.* (BECHARA, 2001, p.555 apud MAIA, 2011)
the.M.SG people.M.SG work.3PL
'People work.'

Here it is necessary to consider the agreement both in the internal domain of the noun phrase and in the external clause domain to understand the specifications of these nouns. Thus, firstly in the domain of the noun phrase, the combination of the determiner with the noun produces two alternative options for combining features, demonstrated in (25), below:

(25) (a) *o povo / a gente*



(b) *o povo / a gente*

In the first option, in (25a), the features are compatible at the CONCORD level, which results in the assignment of 'CONC NUM = SG' to the noun phrase. In the second option, in (25b), we have only one INDEX feature contributed by the noun, which results in the assignment of 'INDEX NUM = PL' to the noun phrase. As a result, this specification produces a noun phrase with two options available for the external agreement (CONC NUM = SG or INDEX NUM PL). Because of this, these noun phrases can be combined with both singular and plural verbs, as in the examples in (24).

Therefore, for the sentences in (24a) and (24c), agreement was established morphosyntactically (at the CONCORD level), capturing the meaning that the referent of the noun was conceptualized as an entity. For the sentences in (24b) and (24d), agreement is blocked at the CONCORD level, since the values of the noun phrase and the verb are incompatible. There is an agreement, however, that takes place semantically at the INDEX level, capturing the meaning that the referent of the noun has been conceptualized as several individuals.

The stipulation of these two levels of agreement also allows for a treatment consistent with the *Agreement marking principle* proposed by Bresnan *et al* (2016, p. 191): "Agreement is driven by a syntactic feature of the controller, if the controller has such a feature. If the controller lacks such a feature, then the target agreement inflection is semantically interpreted as characterizing the controller denotation".

Following this formalization, this principle applies both to the internal domain of the noun phrase and to the clause domain. In the first case, exemplified in (22b), the controller's CONCORD feature is not selected, since the singular noun does not have the NUM PL feature, thus not being able to induce the form of the determiner, with the inflection of the determiner being responsible for conveying the plural denotation of the noun. In the second case, exemplified with the sentences in (24b, d), the controller (the subject) also does not have a NUM PL morphosyntactic feature; nevertheless, the target (the verb) may present a morphosyntactic mark that characterizes the plural denotation of the subject.

Regarding the plural forms of these nouns (Type 1), they function like other nouns. Morphological plural marking simply produces a redundant (semantic and morphosyntactic) plural number assignment for such nouns. Thus, the selection of the item *povos* (people) excludes any interpretation of a singular referent.

It is important to note that the level of agreement selected for these nouns will depend on each variety (whether standard or popular) and even on issues of discourse register (such as literary and non-literary registers) or structural factors (such as distance between controller and target, aspects related to the properties previously mentioned as conditions of agreement, and which were not modeled in this formalization). The grammatical prescriptive tradition, for example, tends to associate the patterns in (24b,d) with popular varieties or with contexts of use determined by the literary register, where such structures are acceptable for reasons of style and poetic license, which allow for relaxation of the canonical grammatical standards (CUNHA, 1986; AMARAL *et al*, 2003; BECHARA, 2001; TERRA, 2002). These cases are even generally treated as a figure of speech, called *syllipsis* (when agreement is made logically and not formally), configuring a type of deviation from the canonical standards of the language, which require agreement at the level of form.

Now let's see the proposed specification for verbs:

Verbs

(a) Singular verbs

(↑ SUBJ CONC NUM)

(b) Plural verbs

{(↑ SUBJ CONC NUM) (↑ SUBJ CONC NUM) ~ PL (↑ SUBJ INDEX NUM) =c PL | (↑ SUBJ CONC NUM)=c PL}

The specification of verb agreement features is different from that of other elements. This is because the verb is located in another domain of agreement. Whereas determiners, nouns, and nominal modifiers specify features and values restricted to their own local syntactic domain, the verb specifies features that are linked to features of its external argument, the subject. Note that, in the given formalization, the features CONC and INDEX of the verb are attributes of the functional structure of the subject, represented by '↑ SUBJ'. This also formalizes the syntactic property of BP in which the verb establishes agreement with only one of its arguments: the one with the syntactic function of subject.

For the singular verb specification, we only have a restrictive property with respect to number. The functional scheme '(↑ SUBJ CONC NUM)' simply requires the noun to convey some value for the feature CONC NUM. A notation of the type '(↑ SUBJ CONC NUM)', in the LFG formalism, encodes the information that the singular noun is underspecified for the value of CONC NUM, being able to accept any value provided by another element of the same functional structure. A notation of the type '(↑ CONC NUM) = SG' symbolizes an equation of the defining or attributive type, that is, one that defines and projects attributes and values into the functional structure which it is part of. A notation of the type '(↑ SUBJ CONC NUM)' encodes an existential constraint, that is, one that does not project any specific value, but only requires that a value be given by another element of the same functional structure. With this specification, the grammar licenses constructions of the types exemplified in (26a) and (17), repeated below as (26b,c):

- (26) (a) *O cavaleiro passa.*
 (b) *Os cavaleiro passa.*
 (c) *Os cavaleiros passa.*

In all examples of (26), all subjects assign a value to the feature CONC NUM, satisfying the verb requirement. In (26a), the subject has a CONC NUM feature with SG value, and in (26b,c) the subjects have a CONC NUM feature with PL value, compatible with the verb. In (26c) this feature came from both the determiner and the noun, and in (26b) the feature came from the determiner only. Through this formalization, it follows a generalization that in BP, more specifically in popular varieties, a single plural morphological mark in the noun phrase is necessary to guarantee the agreement and grammaticality of a simple sentence, since this mark can project its information to the whole phrase. And as a result, the noun phrase can establish agreement with the verb.

As for the plural form of nouns, there is a more complex specification, as plural verbs need to make restrictions at both levels (INDEX and CONCORD). The specification is then implemented through a disjunction. The first element of the disjunction stipulates that the subject can have a CONC NUM feature with a value different from PL, along with an INDEX NUM feature with a PL value. That is, the subject can simultaneously have the information CONC NUM = SG and INDEX NUM = PL. This allows plural verbs to combine with subjects that have Type 1 nouns, even if the subject has a CONC NUM = SG, as in examples (24b) and (24d), repeated below in (25):

- (27) (b) *A gente fazemos.* (MENDES, OUSHIRO, 2015)
 (d) *O povo trabalham.*

The second disjunct '(↑ SUBJ CONC NUM)=c PL', in turn, simply requires that the noun have a CONC NUM feature with PL value. That is, the plural verb can alternatively be combined with subjects that do not have distinct values for INDEX NUM and CONC NUM, given that the value of CONC NUM for that subject is PL. Thus, the sentences exemplified in (28) are licensed:

- (28) (a) *Os meninos chegaram.*
 the.M.PL boys arrived.3PL
 ‘The boys arrived.’

- (b) *As menina tiraram retrato.* (BRANDÃO, 2011)
 the.F.PL girl took.3PL picture
 ‘The girls took pictures.’

It should be noted that the specification of the INDEX features in the plural verb must be done by means of a restrictive equation ($(\uparrow \text{SUBJ INDEX NUM}) = \text{c PL}$) and not simply by a defining equation. The assignment of ‘INDEX NUM = PL’ to the subject, via the verb, would lead the grammar to license the constructions in (29) and (30) below, since both the determiner and the singular noun do not specify any value for INDEX. This is also true for the CONCORD value, where the default singular value of the noun could be overridden by the plural value assigned by the verb, if the verb specification were a defining one.

- (29) * *O cavaleiro esperam.*
 the.SG knight.SG wait.3PL

- (30) * *Cavaleiro esperam.*
 knight.SG wait.3PL

With this whole set of specifications for determiners, nouns and verbs, the grammar that contains them will provide valid analyzes for the main types of number agreement patterns in BP listed throughout this text, covering part of the variability of this phenomenon in both popular and standard varieties of the language.

4.1 COMPUTATIONAL IMPLEMENTATION AND RESULTS OF THE PROPOSED FORMALIZATION

In order to test and verify its consistency, all the formalization proposed and described above was submitted to computational implementation. To this end, all specifications presented were converted to the LFG-XLE¹¹ notational language (see CROUCH *et al*, 2011 on XLE notation) and implemented in the XLE environment (CROUCH *et al*, 2011). For this implementation, I used an LFG-XLE minigrammar from BP (MOURÃO, 2020) containing a component of syntactic rules and a lexicon with items belonging to the categories that were the object of this study, enriched with the proposed specifications.

For testing this grammar, two sets of sentences were used: a set of grammatical sentences (positive test), reproduced in (31) and a set of ungrammatical sentences (negative test), reproduced in (32). The sentences of the positive test were elaborated based on the agreement patterns acceptable in BP and modeled in the grammar, and that, therefore, should be analyzed by the implemented grammar. The negative test sentences were constructed based on unacceptable patterns, which violate agreement at some level, and that should not receive any analysis.

- (31) Positive test set
 #01: *o cavaleiro espera*
 #02: *os cavaleiros esperam*
 #03: *o povo espera*
 #04: *os povos esperam*
 #05: *os cavaleiros espera*
 #07: *os cavaleiro espera*
 #08: *os povos espera*

¹¹ LFG-XLE refers to the LFG formalism implemented in the Xerox Linguistic Environment (XLE), which is a computational environment for implementing lexical-functional grammars. Therefore, there are adaptations to the symbols of the classical LFG theory and even additions to the notation.

- #09: *os povo espera*
 #10: *os povo esperam*
 #11: *o povo esperam*
 #12: *cavaleiro espera*
 #13: *cavaleiros esperam*
 #14: *cavaleiros espera*

- (32) Negative test set
 #01: *o cavaleiro esperam*
 #02: *o cavaleiros esperam*
 #03: *o cavaleiros espera*
 #04: *o povos esperam*
 #05: *o povos espera*
 #06: *cavaleiro esperam*

The sentences were submitted to the grammar, and the XLE system returned analyzes compatible with the proposed formalization. All positive test sentences received exactly one valid analysis¹².

I present, below, some examples of analyzes generated by the grammar (only the simplified functional structures):

(33) **Image 1:** F-structure for sentence #01

"o cavaleiro espera"	
	[PRED 'ESPERAR<[1:CAVALEIRO]>']
5	3[PRED 'CAVALEIRO']
6	4[CONC [NUM SG]]
16	1[SUBJ
17	2[GEN MAS, PERS 3]
22	12]

Source: the author with XLE.

Image 2: F-structure for sentence #02

"os cavaleiros esperam"	
	[PRED 'ESPERAR<[1:CAVALEIRO]>']
5	3[PRED 'CAVALEIRO']
6	4[CONC [NUM PL]]
16	1[SUBJ
17	2[INDEX [NUM PL]]
22	12[GEN MAS, NUM PL, PERS 3, SPEC DEF]

Source: the author with XLE.

Image 3: F-structure for sentence #07

"os cavaleiro espera"	
	[PRED 'ESPERAR<[1:CAVALEIRO]>']
5	3[PRED 'CAVALEIRO']
6	4[CONC [NUM PL]]
16	1[SUBJ
17	2[INDEX [NUM PL]]
22	12[GEN MAS, PERS 3, SPEC DEF]

Source: the author with XLE

Image 4: F-structure for sentence #11

"o povo esperam"	
	[PRED 'ESPERAR<[1:POVO]>']
5	3[PRED 'POVO']
6	4[CONC [NUM SG]]
16	1[SUBJ
17	2[INDEX [NUM PL]]
22	12[GEN MAS, PERS 3]

Source: the author with XLE

As for the negative test sentences, they were all blocked by the grammar and did not receive any valid analysis. It is worth mentioning sentences #02, #03, #04 and #05, which were blocked due to the conflict of features in the determiner and in the noun.

Based on the modeled patterns, it is possible to say that in the popular varieties, a single explicit mark for the number feature is sufficient in one of the components of the agreement domain, the controller or the target, while the others are optional. This single mark, however, must be given in the first component, which is why structures such as sentences #02 to #05 of the negative test are not licensed. The mark given in the second element is different from the one in the first component which seems to have greater

¹² In LFG, a valid analysis refers to the output of applying a grammar to some linguistic expression. This analysis includes at least two paired parsing representations: a tree structure representation (called c-structure) and its respective functional representation (f-structure), in the form of a matrix of attributes and values, like the one given in (16).

strength in assigning the number value to the whole phrase. This proposal is compatible with the analyzes of several other authors, according to whom the first position in the noun phrase, in some varieties of BP, is the privileged locus for plural marking: if there is a determiner, it receives the plural mark; if not, the element that occupies that position receives the mark (GUY, 1981; SCHERRE, 1988,1994; COSTA, SILVA, 2006; ANDRADE, 2003).

The pattern given in #02, #03, #04 and #05 may appear in some varieties, as shown by Brandão (2011, p. 168), who shows the examples in (34), in which the plural mark in the noun phrase formed by the two last words is given only on the second word.

- (34) (a) *somos nós o culpados*
 be.1PL we the.M.SG guilty.M.PL
 ‘we are the guilty ones’
 (b) *sei lá essa influências*
 know.1SG there this influences
 ‘I don’t get these influences’
 (c) *só se for sonho bons*
 only if be.2SG dream good.PL
 ‘only if it’s good dreams’

However, the same author shows that these occurrences have an extremely low frequency in relation to other patterns. She says that pre-nuclear constituents, such as the determiners in the examples above, are the least susceptible to deletion of the number mark, presenting a much lower frequency of occurrence in relation to deletion in other elements of the noun phrase in other positions. Due to this low frequency, this pattern was not included in the formalization proposed here.

5 FINAL REMARKS

In this paper, I set out to develop a theoretical discussion and formal analysis of BP agreement patterns. The discussion held and the results achieved, although preliminary, allow us to catch the non-trivial nature of the agreement phenomenon in this language. When we perform superficial analyses on the standard varieties of BP, agreement seems relatively simplified and homogeneous. Elements involved in a relation of agreement display whenever possible the same morphosyntactic marks of grammatical features. But when we investigate popular varieties, we come across structures that seem to violate grammatical properties of the language. The patterns get more variable, and many are semantically equivalent inside the same variety or even across varieties.

To capture this variability and ensure the grammaticality of these patterns in a single grammar that encompasses both linguistic varieties, I proposed that number agreement in BP mobilizes two levels of features: one closer to the morphosyntactic level (CONCORD) and another closer to the level semantics (INDEX). It is this duplicity of levels that creates alternative possibilities for combining lexical and syntactic forms, allowing that even when there is a failure in agreement between these elements in one of the levels, agreement can still be satisfied in another.

It is also possible to infer from this proposal that the agreement patterns typically associated with standard varieties are those that are established more at the morphosyntactic level (CONCORD), while popular varieties use more semantic agreement (INDEX). While the popular varieties tend to be satisfied with a semantically determined agreement, at the same time they are less morphosyntactically marked, the standard varieties tend to make formally explicit the agreement features whenever possible, hence their greater redundancy in the marking of these features. That is why speakers of standard varieties might prefer the structure *a gente faz* (we do), in which agreement was established at the morphosyntactic level, even though the noun has a plural denotation. Speakers of popular varieties can select the structure *a gente fazemos* (we do), dispensing with morphosyntactic agreement, but maintaining agreement at the semantic level.

Another conclusion is that singular nouns, as well as singular verbs, are semantically underspecified for the number feature. Although they have a singular form, semantically they are neither plural nor singular, and therefore they can be syntactically combined with elements that specify any of these values, or even take one of these values from another element, such as the singular noun that takes a plural value from plural determiners. At the morphosyntactic level, the noun with singular form specifies a singular value only by default, whenever a different value is not assigned by some other element with which it establishes agreement.

Agreement in BP, therefore, involves cases of both mandatory and optional morphosyntactic agreement. The morphosyntactic obligatoriness is typical of standard varieties, while optionality is typical of popular varieties. It is clear that the choice for one or another pattern, in the use of the language, can have extralinguistic motivations and even consequences (such as sociocultural issues), but these problems are outside the modeled linguistic system and the scope of this work.

With the proposed formalization, the main highlighted patterns of agreement of both varieties are grammatically valid. And considering the perspective of natural language processing, the two varieties become computationally treatable, with potential to receive automatic processing and analysis. The modeled computational grammar fragment, which encompasses the two varieties, provides a basic but fundamental resource in natural language processing tasks and applications. The analysis produced by this grammar provides crucial linguistic information to other types of language processing, and none of the varieties is left out of this treatment.

Finally, it is important to mention that the scope of the phenomenon under analysis was relatively restricted, being limited to the treatment of the number agreement feature, in the syntactic domain of uncoordinated noun phrases, without modifiers, and of simple sentences. For further refinement of the proposal, it is necessary to expand the analysis to include more complex structures, such as coordinated noun phrases, with modifiers and other determiners, in addition to complex sentences. It is also necessary to formalize and implement, based on the same proposal, other agreement features, such as gender and person, which will impose their own constraints.

REFERÊNCIAS

- ACKEMA, P.; NEELEMAN, A. Subset controllers in agreement relations. *Morphology*, v. 23, n. 2, p. 291–323, 2013.
- ALENCAR, L. F. de. BrGram: Modelação computacional de padrões variáveis de concordância em português. *Revista de Estudos da Linguagem*, v. 21, n. 1, p. 43–110, 2013b.
- ALENCAR, L. F. de. BrGram: uma gramática computacional de um fragmento do português brasileiro no formalismo da LFG. In: *BRAZILIAN SYMPOSIUM IN INFORMATION AND HUMAN LANGUAGE TECHNOLOGY*, 9., 2013a, Fortaleza. Anais ... Fortaleza, Sociedade Brasileira de Computação, Fortaleza, 2013. p. 183-188.
- AMARAL, E. et al. *Novas Palavras: português*. 2. ed. São Paulo: FTD, 2003.
- ANDRADE, P. Ribeiro de. *Um fragmento da constituição sócio-histórica do português do Brasil: variação na concordância nominal de número em um dialeto afro-brasileiro*. 2003. Dissertação (Mestrado em Linguística) - Universidade Federal da Bahia, Salvador.
- BECHARA, E. *Moderna Gramática Portuguesa*. 37. ed. Revista e ampliada: Rio de Janeiro: Editora Lucerna, 2001.
- BELYAEV, O.; DALRYMPLE, M.; LOWE, J. Number mismatches in coordination: an LFG analysis. In: BUTT, M.; KING, T. H. (ed.). *Proceedings of the LFG15 Conference*. Stanford, CA: CSLI Publications, 2015. p. 26-46.
- BRANDÃO, S. Concordância nominal em duas variedades do português: convergências e divergências. *Veredas-Revista de Estudos Linguísticos*, v. 15, n. 1, 2011.
- BRESNAN, J. et al. *Lexical-Functional Syntax*. 2nd. ed. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2016.

- CORBETT, G. G. *Agreement*. Cambridge University Press, 2006.
- CORBETT, G. G. Agreement: terms and boundaries. In: GRIFFIN, W. E. (ed.). *The Role of Agreement in Natural Language*: TLS 5 Proceedings. Texas Linguistics Forum, v. 53, 2003, p. 109-122.
- COSTA, J.; SILVA, M. C. Nominal and verbal agreement in Portuguese: an argument for distributed morphology. In: COSTA, J.; FIGUEIREDOSILVA, M. C. (ed.). *Studies on agreement*. Amsterdam: John Benjamins, 2006. p. 25-46.
- CROUCH, et al. *XLE Documentation*. Palo Alto Research Center (PARC). Disponível em: https://ling.sprachwiss.uni-konstanz.de/pages/xle/doc/xle_toc.html. Acesso em: 30 set. 2023.
- CUNHA, C. F. *Gramática da Língua Portuguesa*. 11. ed. Rio de Janeiro: FAE, 1986.
- DALRYMPLE, M.; HRISTOV, B. Agreement patterns and coordination in Lexical Functional Grammar. In: BUT, M. KING, T. H. (ed.). *Proceedings of the LFG10 Conference*. Stanford, CA: CSLI Publications, 2010. p. 186-206.
- DALRYMPLE, M.; MYCOCK, L. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford: Oxford University Press, 2019.
- GARDELLE, L. *Semantic Plurality*: English collective nouns and other ways of denoting pluralities of entities. Amsterdam: John Benjamins, 2019. (Current Issues in Linguistic Theory, Vol. 349)
- GUY, G. R. *Linguistic variation in Brazilian Portuguese*: aspects of phonology, syntax and language history. 1981. Tese (Doutorado em Linguística) - University of Pennsylvania, Pennsylvania, 1981.
- HAUG, D. T. T.; NIKITINA, T. Feature sharing in agreement. *Natural Language & Linguistic Theory*, v. 34, n. 3, p. 865-910, 2016.
- HOLM, J. A. The genesis of the Brazilian vernacular: insights from the indigenization of Portuguese in Angola. *PAPIA - Revista Brasileira de Estudos do Contato Linguístico*, v. 19, n. 1, p. 93-122, 2009.
- JOHSON, C. *Deconstructing and Reconstructing Semantic Agreement: A Case Study of Multiple Antecedent Agreement in Indo-European*. PhD dissertation, Graduate School of The Ohio State University, 2014.
- KATHOL, A. Agreement and the syntax-morphology interface in HPSG. In: LEVINE, R. GEROGIA, G. (ed.). *Studies in Contemporary Phrase Structure Grammar*, Cambridge: Cambridge University Press, 1999. p. 223-274.
- KING, T. H.; DALRYMPLE, M. Determiner agreement and noun conjunction. *Journal of Linguistics*, v. 40, n. 1, p. 69-104, 2004.
- LEITE, M. Q. Português culto e popular: entrelaçamentos. In: *I SIMELP-Simpósio Mundial de Estudos de Língua Portuguesa*, 2008. (comunicação).
- LOPES, C. R. S. *A Inserção de “a Gente” no Quadro Pronominal do Português*. Frankfurt am Main: Vervuert, 2003.
- LUCCHESI, D. Parâmetros Sociolinguísticos do português brasileiro. *Revista Da ABRALIN*, v. 5, n. 1/2, 2017.
- LUCCHESI, D. A variação na concordância verbal no português popular da cidade de Salvador. *Estudos Linguísticos e Literários*. n. 52, ago-dez, p. 166-204, 2015.
- MAIA, J. P. *Variação na Concordância Verbal com nomes coletivos em Fortaleza-Ce*. Dissertação (Mestrado em Linguística) – Programa de Pós-Graduação em Linguística, Universidade Federal do Ceará, Fortaleza, 2011. 111p.

- MARCOTULIO, L., J.; VIANNA, C.; LOPES. Agreement patterns with “a gente” in Portuguese. *Journal of Portuguese Linguistics*, n. 12 2013, pp. 125–149.
- MASCARELLO, L. Variação na Concordância Verbal de Terceira Pessoa do Plural na Fala dos Florianopolitanos Nativos na Costa da Lagoa. *Work. pap. linguíst.*, n.esp.: p. 57-68, 2010.
- MENDES, R. B.; OUSHIRO, L. Variable number agreement in brazilian portuguese: an overview. *Language and Linguistics Compass*, Malden, v. 9, n. 9, p. 358-368, 2015.
- MOREIRA; VIANNA. A concordância nominal no português brasileiro falado em Nova Iguaçu. *Caligrama*, Belo Horizonte, v. 23, n. 2, p. 99-120, 2018.
- MOURÃO, Jessé. *PB-LFG-grammar*: fragmento de gramática LFG-XLE do português brasileiro. 16 abr. 2020. Disponível em: <https://github.com/jessemourao/PB-LFG-grammar>. Acesso em: 05 jan. 2023.
- SADLER, L. Coordination and Asymmetric Agreement in Welsh. In: BUTT, M.; KING, T. H. (ed.). *Nominals: Inside and Out*. Stanford, CA: CSLI Publications, 2003. p. 85-118.
- SADLER, L.; KUHN, J. Single conjunct agreement and the formal treatment of coordination in LFG. In: BUTT, M.; KING, T. H. (ed.). *Proceedings of the LFG07 Conference*. Stanford, CA: CSLI Publications, 2007. p. 302-322.
- SANTOS, A. F. dos. *Uma Gramática LFG-XLE para a Análise Sintática Profunda do Português* (Tese de Doutorado). Universidade Federal do Ceará, Fortaleza, Ce, Brazil, 2014.
- SCHERRE, M. M. P.; NARO, A. J. Sobre a concordância de número no português falado do Brasil. In: RUFFINO, GIOVANNI (org.) *Dialettologia, geolinguística, sociolinguística*. (Atti del XXI Congresso Internazionale di Linguistica e Filologia Romanza) Centro di Studi Filologici e Linguistici Siciliani, Università di Palermo. Tübingen: Max Niemeyer Verlag, v. 5, 1998. p. 509- 523.
- SCHERRE, M. M. P. Aspectos da concordância de número no português do Brasil. *Revista Internacional de Língua Portuguesa* (RILP) - Norma e Variação do Português. Associação das Universidades de Língua Portuguesa. v. 12, p. 37-49. dez. 1994.
- SCHERRE, M. M. P. *Reanálise da concordância nominal em português*. 1988. Tese (Doutorado em Letras) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1988.
- SHIEBER, S. M. *An introduction to unification-based approaches to grammar*. Brookline, Massachusetts: Microtome Publishing, 2003. Reissue of Shieber, Stuart M. 1986. *An introduction to unification-based approaches to grammar*. Stanford, California: CSLI Publications.
- SCHWARZSCHILD, R. *Pluralities*. Dordrecht: Springer Science & Business Media, 1996. (Studies in Linguistics and Philosophy, vl. 61)
- STEELE, S. Word order variation: a typological study. In: GREENBERG, J. H.; FERGUSON, A.; MORAVCSIK, E. (ed.). *Universals of Human Language IV: Syntax*. Stanford: Stanford University Press, 1978. p. 585-623.
- TERRA, E. *Curso Prático de Gramática*. São Paulo: Scipione, 2002.
- WESCOAT, M. T. *Practical Instructions for Working with the Formalism of Lexical Functional Grammar*, 1989. Manuscript, Xerox PARC. Disponível em: <http://www.essex.ac.uk/linguistics/external/lfg/FAQ/Introductions.html>. Acesso em: 15 nov. 2021

WECHSLER, S.; ZLATIĆ, L. A Theory of Agreement and Its Application to Serbo-Croatian. *Language*, v. 76, n. 4, p. 799–832, 2000.

WECHSLER, S.; ZLATIĆ, L. *The many faces of agreement*. Stanford, Calif: Center for the Study of Language and Information, 2003.

ZILLES, A. The development of a new pronoun: The linguistic and social embedding of a gente in Brazilian Portuguese. *Language Variation and Change*, v. 17, n. 1, p. 19-53, 2005.



Received on August 31, 2021. Approved on July 21, 2022.