

A Lexicon-Grammar of Brazilian Portuguese Predicative Adjectives

Ryan Martinez

Universidade Federal de São Carlos, Brazil
ryan.saldanha.martinez@gmail.com

Jorge Baptista

Universidade do Algarve, Portugal
INESC-ID Lisboa, Portugal
jbaptis@ualg.pt

Oto Araújo Vale

Universidade Federal de São Carlos, Brazil
otovale@ufscar.br

Abstract

This paper presents a syntactic lexicon of Brazilian Portuguese predicative adjectives that are not regularly derived from verbs. From the 7,000 most frequent adjectives in a large web corpus, 3,161 lexical items were selected and annotated with 36 syntactic properties. These properties were established through introspection and corpus evidence, covering argument structure, copular verbs, prepositions, transformations (e.g., raising, nominalization), semantic roles, and others. The resulting resource constitutes a machine-readable lexicon of predicative adjectives for Brazilian Portuguese.

1 Introduction

Different senses of a word often correlate with distinct syntactic patterns. In Brazilian Portuguese, for instance, *fofo* can mean either “kind” or “fluffy,” each corresponding to different argument structures. The first sense (1a) allows a clausal subject and complements, while the latter (1b) does not.

- (1) a. *Que João diga isso a Maria é muito fofo da parte dele para com ela.*
(That João said this to Maria was kind of him towards her)
- b. *Esse travesseiro é fofo.*
(This pillow is fluffy)

Other differences include the fact that the exclamative sentence *Seu fofo!* (“you kind/cute!”) refers to the first sense; also, nouns referring to a piece of language, such as *discurso* (discourse), or to a certain behavior, such as *comportamento* (behavior), can only take *fofo* in the former sense. When applied to *corpo* (body) or *textura* (texture), *fofo* can only convey the latter sense. Finally, *fofo* in syntactic patterns such as those in (2) can only mean “kind”.

- (2) a. *João foi fofo em fazer isso.*
(João was kind to do that)
- b. *Isso é fofo de se fazer*
(lit. “This is kind to do” = It is kind to do this)

This study presents a machine-readable lexicon of 3,161 predicative adjectives, selected from a list of the 7,000 most frequent adjectives found in a large-sized corpus, annotated with 36 syntactic properties capturing distinctions such as those presented above. These adjectives cannot be regularly derived from verbs. To the best of our knowledge, no such lexical-syntactical resource exists yet for Portuguese.

The following sections present our framework and methodological choices, followed by a description of the properties contained in the lexicon, their frequency, and relation with word sense (or lack thereof).

2 Theoretical and methodological framework

This study adopts the theoretical framework of Zellig S. Harris’s *Transformational Operator Grammar*, building on his foundational works (Harris, 1968, 1976, 1982, 1991). According to this approach, language structure results from a fundamental constraint governing word entry into the formation of a base sentence. Lexical items are thus divided into *operators*—words that select other words to form a sentence—and *elementary arguments*—words that do not themselves select arguments. This basic distinction, together with subsequent restrictions on word combination, determines the meaning of a given sentence.

A *base sentence* is the expression of a semantic predicate, determined by an operator and its arguments. In Portuguese, the base sentence of *predicative adjectives* consists of a clause contain-

ing a subject, a linking verb (typically *ser* or *estar*), and an adjective. In some cases, the predicate may also require a prepositional phrase as an essential complement (see 3).

Operators can be further classified according to the *number* (one, two, or three) and the *type* of their arguments (whether an operator or an elementary argument). *First-order operators* apply only to elementary arguments, that is, to non-operators (see 3a); whereas *second-order operators* apply to at least one other operator (see 3b), possibly to two, and may also select elementary arguments.

- (3) a. *João é truculento com Maria.*
(João is truculent towards Maria)
- b. *Que João faça isso é retrógrado.*
(It is retrograde that João do this)

All sentences of a given language can be derived from a set of base sentences and the operations applied to them (mainly reductions), in accordance with the transformational principles defined by Harris (1991). Accordingly, the operators and arguments in (3a) and (3b) may be combined and transformed into a more complex sentence:

- (4) *Que João seja truculento com Maria é retrógrado.*
(That João is truculent towards Maria is retrograde)

Sentences may undergo transformations that alter their form while preserving their informational content. Thus, (4) can be transformed into sentences such as (5).

- (5) a. *João foi retrógrado em ser truculento com Maria*
(‘João was retrograde in being truculent towards Maria’)
- b. *Que João seja truculento com Maria é retrógrado da parte dele.*
(It is retrograde of João that he is truculent against Maria)

In this paper, we combine the theoretical principles of Harris with the methodological framework of the Lexicon-Grammar (Gross, 1975, 1981, 1988, 1996b). This approach consists in adopting a formalism that represents operators as rows in a table and grammatical features as columns. Each feature is marked as acceptable (‘+’) or unacceptable

(‘-’) for a given predicate. The result is a machine-readable lexicon with systematic syntactic annotation.

Crucially, the Lexicon-Grammar approach further proposes that any description be carried out systematically over a significant portion of the lexicon, treating the elementary sentences determined by the predicative lexical items as the basic units of analysis.

Our list of predicative adjectival operators was built as follows. Firstly, we extracted the list of all adjectives in the Brazilian partition of PtTenTen20 web text corpus (Kilgarriff et al., 2014), ordered by frequency, and using Sketch Engine¹. We then assessed the syntactic-semantic behaviour of the first 7,000 items, resorting both to intuition and corpora, and tested them for the following features: a) whether any sense of the adjective corresponded to a past participle occurring in a passive sentence, regardless of the copular verb being *estar* ‘be’, as in (6a), or *ser* ‘be’, as in (6b). If an equivalent active sentence could be identified, as in (6c), that sense was placed in a separate list.

- (6) a. *Esse ralo está entupido.*
(This is drain is clogged)
- b. *Esse ralo foi entupido pela sujeira.*
(This drain was clogged by dirt)
- c. *A sujeira entupiu o ralo*
(The dirt clogged the drain)

As suggested by (Gross, 1996a), these adjectives can be more economically described as resulting from transformations of the corresponding verbal construction (6c).

b) whether any sense of an adjective ending in *-nte* or *-or* corresponded to a verbal construction. Adjectives in *-nte* have been shown to be regularly derived from verbs (Laporte, 1992), while those in *-or* can arguably be accounted for by a similar analysis, as shown in examples (7a-7b). Senses that met this criterion were placed in a separate list.

- (7) a. *Que João faça isso é humilhante / consolador para Maria.*
(It is humiliating / consoling for Maria that João does this)
- b. *Que João faça isso humilha / consola Maria*

¹<https://www.sketchengine.eu/>

(That João does this humiliates / consoles Maria)

c) whether the adjective ending in *-vel* ‘-ble’ (8a) was equivalent to a passive verbal construction combining the modal verbs *poder* ‘may/can’ or *dever* ‘must’ with auxiliary verb *ser* ‘be’ and a past participle (8b):

- (8) a. *Essa doença é curável*
(This disease is curable)
b. *Essa doença pode ser curada*
(This disease can be cured)

According to (Leeman and Meleuc, 1990), these adjectives can be trivially derived from their verbal counterparts. Several of them have also been described for European Portuguese by (Freire, 1995).

d) whether any sense of an adjective readily accepted a copular construction with *ser* or *estar* in non-contrastive contexts, or without an implied classifier noun such as *tipo* ‘type’ or *natureza* ‘nature’ (Harris, 1976, p. 62) (9c). Adjectives that did not meet these conditions, as in (9b–9c), were classified as non-predicative, following the criteria proposed by (Rio-Torto, 2006; Veloso and Raposo, 2013). This group comprised 2,567 adjectives.

- (9) a. *João mandou um cartão postal para Maria*
(João sent Maria a postcard)
b. **Esse cartão é/está postal*
(This card is postal.)
c. **Esse cartão é (de tipo postal) (e não de outro tipo)*
(This card is (of a) postal (type) and not of another type.)

Each adjective that satisfied these criteria was subsequently examined for its distinct distributional properties, each of which is typically associated with a different meaning. In addition to analysing our corpus data and relying on linguistic intuition, we systematically consulted lexicographic sources—primarily (Borba, 2002)—to ensure that no relevant sense was overlooked.

Each distributional pattern corresponded to a separate lexical entry. To illustrate this procedure, consider the adjective *duro*, one of the most polysemous items in our dataset. Each distribution of (10) receives its own entry in our lexicon-grammar.

- (10) a. *Fazer isso é duro.*
(It is hard to do this)
b. *Que João faça isso é duro para com Maria.*
(It is rough towards Maria for João to do this.)
c. *A consistência desse objeto é dura.*
(The consistency of this object is hard.)
d. *João está duro.*
(João is financially broke.)

Adjectives formed with the prefixes *a-*, *anti-*, *des-*, or *i/in/im-*, that preserve the distributional constraints of their corresponding base adjective (11) were not assigned a separate lexical entry. Instead, the ability to take one of these prefixes was encoded in a dedicated column, labelled *Neg*.

- (11) a. *Que João faça isso é típico da parte dele.*
(It is typical of João for him to do this)
b. *Que João faça isso é atípico da parte dele.*
(It is atypical of João for him to do this)

These procedures resulted in a lexicon-grammar comprising 3,161 predicative adjective entries.

The syntactic properties considered in this study draw on previous work conducted for other languages (Vendler, 1968; Picabia, 1978; Meunier, 1981; Rahaga, 1995; Meydan, 1995; Nam, 1996; Sklavouínou, 1997; Chung, 2003; Valetopoulos, 2003; Messina, 2019) as well as for European Portuguese (Casteleiro, 1981; Baptista, 2005; Carvalho, 2007). These properties are discussed in greater detail in the following section.

In addition to the formal properties, each argument was assigned a semantic role using the tagset commonly adopted in other Lexicon-Grammar descriptions of Portuguese (Talhadas et al., 2013; Baptista and Mamede, 2020).

A fraction of our classificatory judgements were reproduced by 31 respondents of a form on the relation between syntax and word sense. The participants were presented with pairs of sentences containing the same adjective in different structures and were asked, for each pair of sentences, whether the adjective had the same meaning in both sentences (yes/no). A total of sixteen adjectives in 32 pairs of sentences were assessed: *natural* (‘natural’), *sadio* (‘healthy’), *nutritivo*

(‘nutritious’), *aéreo* (‘aerial’ or ‘distracted’), *alegre* (‘happy’), *alto* (‘tall’), *inocente* (‘innocent’), *tímido* (timid), *horrível* (horrible), *cego* (blind), *mole* (‘soft’), *grátis* (free-of-charge), *quieto* (quiet), *racista* (racist), *louco* (crazy) and *bobo* (silly). The precision for this task in relation to the data in our lexicon-grammar was 0.831 and the accuracy was 0.8.

3 Results

The lexicon-grammar resulting from this study is made available to the scientific community at <https://tinyurl.com/bpadjlexgram>. The following subsections present the syntactic properties identified for each adjective and discuss their overall distribution, including observed regularities and irregularities.

3.1 Number and type of arguments

All predicative adjectives were classified according to their argument structure—whether they select one or two arguments—and whether each argument could be realized as a clause. Intransitive adjectives (i.e. without essential complements) were the most frequent, particularly those whose subject cannot take a subclause. This class, labelled **AN0**, is illustrated in (12). It contains 1,729 adjectives, representing 54.7% of all predicative adjectives.

- (12) a. *Esse objeto é luminoso.*
(This object is luminous.)
- b. * *É luminoso que João fez/faça isso.*
(It is luminous that João did that.)

Columns $N_0 = Hum$ and $N_0 = nHum$ indicate, respectively, whether the subject of the adjective may be a human noun or a non-human noun.

We also propose a subclass of AN0 to distinguish adjectives whose meaning changes substantially when applied to human nouns (13a) as opposed to non-human nouns (13b). Adjectives exhibiting this non-human reading are classified as **AN0n**, a group comprising 112 items.

- (13) a. *João é cabeludo.*
(João is hairy)
- b. *Essa história é cabeluda.*
(This story is absurd)

The second most frequent type consists of intransitive adjectives that take a clausal subject. This

class, labelled **AQ0**, is exemplified in (14), and comprises 551 predicative adjectives, representing 17.43% of the total.

- (14) *É claro que João fez isso.*
(It is clear that João did that.)

Subject clauses may appear in the indicative mood (column $N_0 = QueF_{ind}$), in the subjunctive mood ($N_0 = QueF_{subj}$), as an infinitival clause ($N_0 = V_{inf}$), or in the indicative mood introduced by the subordinating conjunction *se* (‘whether’) ($N_0 = SeF_{ind}$). These fine-grained syntactic properties have been systematically encoded in the lexicon-grammar.

The third most frequent type consists of transitive adjectives that do not accept a clausal complement. This class, labeled **AN2** (15), includes 291 adjectives, representing 9.2% of the total entries.

- (15) a. *Esse objeto está distante daquele objeto.*
(This object is distant from that object)
- b. * *Que João faça isso está distante daquele objeto*
(João doing this is distant from this object)
- c. * *Esse objeto está distante de que João faça aquilo*
(This object is distant from João doing this)

Similarly to the subject position, the columns ($N_1 = N_{hum}$) and ($N_1 = N_{nhum}$) specify the distributional constraints on this argument syntactic slot, indicating whether complements may be human or non-human nouns.

The fourth most frequent type of predicative adjective consists of those that accept a subject subclause but not a complement clause, a class labelled **AQN**, illustrated in (16). This group comprises 258 adjectives, representing 8.16% of the total.

- (16) a. *Que João faça isso é bárbaro (para) com Maria.*
(João doing this is barbarous towards Maria)
- b. * *Que João faça isso é bárbaro (para) com que Maria faça aquilo.*
(João doing this is barbarous towards Maria doing that)

There are 196 adjectives (6.2%) that accept both subject and complement subclauses; these were assigned to the class **AQ2**, exemplified in (17).

- (17) *Que João faça isso é facultativo para que Maria faça aquilo.*
(João doing this is facultative for Maria to do that.)

Complement clauses may occur in the indicative mood (column $N_1=QueF_{ind}$), in the subjunctive mood ($N_1=QueF_{subj}$), as an infinitival clause ($N_1=V_{inf}$), or in the indicative mood introduced by the subordinating conjunction *se* ('whether') ($N_1=SeF_{ind}$).

Finally, 136 adjectives (4.3%) accept a complement clause but not a subject clause. This class, labelled **ANQ**, is illustrated in (18). Compare this usage with that in (15), which expresses physical distance, whereas the former denotes temporal distance.

- (18) a. *João está distante de fazer isso.*
(João is distant from doing that)
b. * *Fazer isso está distante de fazer aquilo.*
(Doing this is distant from doing that)

The semantic roles of subjects are specified in column *SR0*. The most frequent role is **object** (also named **theme** in other semantic roles' taxonomies), which accounts for 2,681 cases (84.81%) and appears across all classes. Subjects with the role of **experiencer** total 330 (10.44%) and are particularly common in classes ANQ (e.g., *aprensivo para/de* 'apprehensive to') and AN0 (e.g., *sonolento* 'sleepy'). The possibility of forming sentences with a subject with an **agent** interpretation, as in (19), was identified for 816 adjectives (25.81%). However, such adjectives predominantly alternate with **object**-subject constructions, meaning that the subject does not necessarily denote an active participant in a given event.

- (19) *João foi traiçoeiro / vingativo / jocosos*
(João was treacherous / vindictive / jocular)

The semantic roles of complements are discussed below, in the subsection dedicated to prepositions.

3.2 Copular verbs *ser* and *estar*

In Portuguese, predicative adjectives may combine with the copular verbs *ser* and/or *estar*. Although their contrast is commonly described as one of **permanent** versus **temporary** state, the actual distinction is more nuanced. For instance, the adjective *temporário* ('temporary') preferentially selects *ser* (20a), whereas *morto* ('dead'), definitely denoting a **permanent** state, preferentially combines with *estar* (20b).

- (20) a. *Essa situação é/*está temporária.*
(This situation is temporary)
b. *João ?*é/está morto*
(João is dead)

Most adjectives (1,717) accept both copular verbs. A large number of them, however, combine exclusively with *ser* (1,399), while only a few select *estar* alone (45). In several cases, the choice of copular verb serves as an indicator of word sense. Adjectives that occur only with *estar* are exemplified by *tonto* ('dizzy') in (21a), from class AN0. Replacing *estar* with *ser* yields a different meaning ('stupid') as shown in (21b). This meaning is encoded as a distinct entry in the lexicon-grammar, belonging to class AQ0 rather than AN0.

- (21) a. *João está tonto.*
(João is dizzy)
b. *João é tonto*
(João is stupid)

3.3 Prepositions and semantic roles

Our study identified eight prepositions (Prep) that may introduce complements of adjectives: *a* ('to', 226 adjectives), *com* ('with', 266), *contra* ('against', 12), *de* ('of', 213), *em* ('in', 82), *para* ('for/to', 512), *para com* ('towards', 202), and *por* ('for', 25). These prepositions are not mutually exclusive, as many adjectives accept more than one without any significant change in overall meaning of the sentence. Most complements introduced by *a* (157, 69.47%), *de* (176, 82.63%), *em* (73, 89.02%), and *por* (20, 83.33%) are annotated in column *SR1* as semantic **objects**.

Thirty-five complements introduced by *a* are annotated in the columns *SR1* with **time** (22b) and/or **locative** (22a) semantic roles, often alternating with prepositions *de* or *com* (22b).

(22)

- a. *Esse lugar é anterior àquele lugar.*
(‘This place is prior to that place’)
- b. *Esse evento é simultâneo àquele/com
aquele evento.*
(‘This event is simultaneous with that
event’)

The preposition *de* introduces 19 **locative** complements, including adjectives that take **locative-source** complements, such as *nativo* (‘native’), *natural* (‘natural’), *originário* (‘originary’), and *oriundo* (‘coming from’). The preposition *em* introduces **locative** complements with adjectives such as *presente* (‘present’), *ausente* (‘absent’), and *imerso* (‘immersed’).

The prepositions *para* (376 occurrences, 72.87%), *com* (189, 71.05%), and their variant *para com* (186, 92.08%) are strongly associated with **beneficiary** and **victim** semantic roles. Twenty-six complements introduced by *a* and eleven introduced by *de* also receive a **beneficiary** or **victim** interpretation; most of these additionally accept other prepositions, especially *para* (23b), *com*, and *para com* (23a). The preposition *contra* is likewise predominantly associated with **victim** roles and is always interchangeable with another preposition (23c), most commonly *a* or *com*.

- (23) a. *João fazer isso é solidário para
com/com Maria*
(João doing this is solidary to Maria)
- b. *João fazer isso é nocivo para/a Maria*
(João doing this is harmful to Maria)
- c. *João é truculento com/contra Maria*
(‘João is truculent with / against Maria’)

The preposition *para* is the most regular marker of **experiencer** complements, occurring with 66 adjectives. In addition, seven complements introduced by *a* are also annotated as **experiencers**, and six of them likewise accept *para*.

- (24) *Que o João faça isso é familiar para/a
Maria.*
(lit. ‘That João does this is familiar to
Maria’)

3.4 Imperatives

Only adjectives that take human subjects may occur in imperative sentences, with the copular verb in the imperative mood (25). Imperative sentences can be analysed as reductions of performative second-order operators selecting human arguments, such as *order*, *ask*, *demand*, *command*, etc. (Harris, 1976, p. 157). This property is encoded in column $V_{copImpAdj}$. This construction often sounds more natural in the negative form, particularly when the adjective has a derogatory meaning (25b).

- (25) a. *Seja breve.*
(Be brief.)
- b. *Não seja travesso.*
(Don’t be mischievous.)

Demonyms (e.g., *carioca* ‘resident of Rio de Janeiro’) accept this construction with a slight difference in meaning: *seja carioca!* (be carioca!) does not mean that one should change their birth or residence place, but rather that they should act as a resident of Rio de Janeiro typically does. In total, 1,466 adjectives were found to allow this property.

3.5 Symmetry (Sim)

Symmetrical predicates involve two arguments that bear the same semantic role with respect to the predicate. As a result, they may switch syntactic positions (26a)-(26b) or be coordinated in a given argument slot (26c)-(26d) without altering the overall meaning of the expression. They also differ from mere *reciprocal* constructions in that the echo-complement is optional rather than obligatory (26c)-(26d).

- (26) a. *A lei X é consistente com a lei Y.*
(Law X is consistent with law Y)
- b. *= A lei Y é consistente com a lei X.*
(Law Y is consistent with law X)
- c. *= A lei X e a lei Y são consistentes (uma
com a outra)*
(Law X and law Y are consistent (with
each other))
- d. *= A lei Y e a lei X são consistentes (uma
com a outra)*
(Law Y and law X are consistent (with
each other))

Our lexicon-grammar includes 71 adjectives exhibiting this property. They frequently express

notions of adequacy (26) and equivalence (e.g., *homônimo* ‘homonymous’), as well as spatial (e.g., *vizinho* ‘neighboring’) and temporal (e.g., *síncrono* ‘synchronous’) relations.

3.6 Pre-nominal position (Adj N₀)

Most Portuguese adjectives may occur in both pre- and post-nominal positions within noun phrases without changing their meaning. Only 1,061 (33.66%) cannot appear in pre-nominal position (27).

- (27) *João é um menino banguela / *um banguela menino.*
(João is a toothless boy)

This study did not uncover fully consistent semantic correlations for this phenomenon. Gentilic and colour adjectives uniformly reject the pre-nominal position. However, among other adjectives, even close semantic pairs differ with respect to this property. For example, *palhaço* (‘clown-like’, ‘comic’) and *engraçado* (‘funny’), or *falso* (‘fake’) and the currently widespread Anglicism *fake*, display contrasting behaviour in this regard.

3.7 Nominalization (Adj-n)

A total of 1,195 adjectives have a corresponding support-verb construction with *ter* (‘have’) plus a predicative noun morphologically associated with the adjective, as illustrated in (28):

- (28) a. *João é/está tuberculoso*
(João is tubercular)
b. *João tem tuberculose*
(João has tuberculosis)

An additional 861 adjectives, partially overlapping with the previous group, have a corresponding support-verb construction with *ser de* (‘be of’) plus a predicative noun (29):

- (29) a. *João é frio para com Maria*
(João is cold towards Mary)
b. *João é de uma certa frieza para com Maria*
(lit. João is of a certain coldness towards Maria)

Some adjectives (77 in total) may be preceded by a determiner in popular constructions with human subjects, yielding a direct nominalization of the

adjective (30). As noted in the literature (Carvalho, 2007, p. 118), this structure is predominantly associated with derogatory adjectives—a tendency confirmed by the present study:

- (30) a. *João é cretino para com Maria*
(lit. João is cretin towards Maria)
b. *João é um cretino para com Maria*
(João is a cretin towards Maria)

3.8 SeuAdj! and SeuNegAdj!

The possibility of forming exclamative sentences with these adjectives preceded by the possessive *seu* (‘you’) has been associated in Portuguese primarily with derogatory adjectives (Carvalho, 2007), functioning consensually as insults. Examples include *cafona* (‘corny’), *hipócrita* (‘hypocrite’), *picareta* (‘unprofessional’, ‘charlatan’), and *porco* (lit. ‘pig’, i.e., ‘unhygienic’).

This construction also applies to certain adjectives whose interpretation may be controversial and perceived as derogatory by some speakers, depending on their values—for instance, *promíscuo* (‘promiscuous’), *mulherengo* (‘womanizer’), or *apolítico* (‘apolitical’).² Finally, a small number of overtly positive adjectives can also occur in this structure, such as *gostoso* (‘sexually attractive’). In total, 549 adjectives were observed to allow this construction, plus an additional 79 when negation is applied, as indicated by the *Neg* feature and recorded in the *SeuNegAdj!* column, for example *insensato* (‘unreasonable’).

3.9 Restructuring

Three types of restructuring operations (Gross, 1975, p. 142–143), (Guillet and Leclère, 1981, p. 100) were identified and described in this study. The first, known as *Subject Raising*, consists of extracting the *subject* from a subject subclause and placing it in the subject position of the matrix clause, while the remainder of the subclause becomes a complement introduced by the preposition *em* (‘in’), retaining its verb in the infinitive, as shown in (31). This structure is encoded as *SubRais* and applies to 497 items.

- (31) a. *Que João diga isso é redundante.*
(lit. That João says this is redundant)
b. *João é redundante em dizer isso.*

²Other adjectives may fit this pattern in prejudiced language, examples of which we refrain from providing here.

(João is redundant in saying this)

The second type of restructuring is known as *Object Raising* (also referred to as “Tough construction”). It consists of extracting the *object* complement of the subject subclause and promoting it to the subject position of the matrix clause, while the remainder of the subclause is moved to a complement introduced by the preposition *de* (‘of’) and the verb is put in the infinitive, as illustrated in (32). This structure is encoded as *ObjRais* and applies to 742 adjectives.

- (32) a. *Fazer isso é baixo*
(lit. Doing this is low = dishonest)
- b. *Isso é baixo de (se) fazer*
(lit. ‘This is low to do’ = It is dishonest to do this)

We also identified restructuring operations (33) involving clauses introduced by *o modo como* (‘the way’).

These expressions formally constitute a noun phrase headed by *modo* taking as a modifier a type of relative clause. The finite relative clause (33a) may be reduced to an infinitival clause modifying the noun *modo* (33b). Under certain adjectives which, by their meaning, can modify both the **agent** and the **manner** of the action, these noun phrases can undergo restructuring such that the subject of the subordinate clause becomes the subject of the matrix clause, while the remainder of the clause surfaces as a prepositional phrase attached to the adjective and introduced by *em* (‘in’). Both the finite and infinitival variants of the construction may undergo this process (33c–33d). Finally, *modo* with the infinitive modifier may also undergo zeroing, as illustrated in (33e). This transformation, encoded as *ModoRais*, applies to 136 adjectives.

- (33) a. *O modo como João faz isso é detalhista.*
(The way João does this is meticulous)
- b. *O modo de João fazer isso é detalhista.*
(The way of João doing this is meticulous)
- c. *João é detalhista no modo como faz isso.*

(João is meticulous in the way as he does this)

- d. *João é detalhista no modo de fazer isso.*
(João is meticulous in the way of doing this)
- e. *João é detalhista em fazer isso.*
(João is meticulous in doing this)

3.10 Echo complements

Certain prepositional phrases following the adjective consistently refer to one of the arguments of the adjective’s subject subclause. We refer to these as **echo complements**, expanding on a hypothesis presented by (Baptista, 2005, pp. 157–162) and originally proposed by (Guillet and Leclère, 1981, pp. 116–117). We identified three types, based on the preposition that introduces them and coreference constraints: (a) *da parte de N₀* (‘from N₀’), (b) *para com N₁* (‘towards N₁’), and (c) *para N₀* (‘for N₀’). The first type refers to an **agent** of the subject clause (34). This possibility is registered in column $F_0 V_{cop} Adj dpd N^0$ (502 adjectives).

- (34) *Que João_i faça isso é consistente da parte dele_i*
(João doing that is consistent of him)

The second type typically refers to an **object** in the subject clause (35). It is encoded as $F_0 V_{cop} Adj para com N^1$ and applies to 211 adjectives.

- (35) *Que João cobre isso de Maria_i é rigoroso para com ela_i.*
(lit. ‘For João to demand this from Maria is rigorous towards her.’)

The third type of echo complement, *para N₀* (‘for N₀’) (36), may refer to either **agent** (36a) or **object** (36b) of the subject subclause. These constructions are encoded as $F_0 V_{cop} Adj para N^0$ and were identified for 204 adjectives.

- (36) a. *Que João_i faça isso é vergonhoso para ele_i.*
(João doing so is shameful for him)
- b. *Que João_i adoça é frequente para ele_i.*
(João falling ill is frequent for him)

3.11 Appropriate nouns

Appropriate nouns (N_{ap}), as defined and discussed in detail by (Harris, 1976; Guillet and Leclère, 1981; Laporte, 1997), are nouns that are highly redundant within a given construction and are therefore frequently omitted. The adjective *quadrado* ('square') (37) illustrates a case in which two distinct senses, (37b) and (37d), are differentiated by means of different appropriate nouns: *formato* 'format' (37a) and *comportamento* 'behaviour' (37c).

- (37) a. *O formato desse objeto é quadrado.*
(The shape of this object is squared)
- b. *Esse objeto é quadrado.*
(This object is squared)
- c. *O comportamento de João é quadrado.*
(lit. João's behavior is square)
- d. *João é quadrado.*
(João is a square)

The distribution of N_{ap} exhibits some regularity, giving rise to major subclasses of adjectives, the most frequent of which are recorded in the lexicon-grammar as N_{ap} =comportamento ('behavior', 1,380 occurrences, 43.66%), N_{ap} =discurso ('discourse', 1,355, 42.87%), N_{ap} =mente ('mind', 1,132, 35.81%), N_{ap} =aparência ('appearance', 436, 13.79%), and N_{ap} =origem ('origin', 330, 10.44%). Other N_{ap} were identified for only a few adjectives or for a single adjective. These are recorded in the column N_{ap} =outros. Smaller groups include *cor* ('colour', 45 instances), *formato* ('format', 38), and *tamanho* ('size', 36). An example of an N_{ap} specific to a single adjective is *veneno* ('venom'), which occurs only with *peçonhento* ('venomous').

4 Conclusion

This paper has presented a machine-readable syntactic description of Brazilian Portuguese predicative adjectives that cannot be regularly derived from verbs. We described the construction of the lexicon, its syntactic properties and their interpretation, and highlighted the observed regularities. We expect this resource to be useful for a wide range of NLP tasks and to offer principled criteria for enriching annotation labels and guidelines.

Acknowledgements

Ryan Martinez: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal

de Nível Superior - Brasil (CAPES) - Finance Code 001. Jorge Baptista developed his research at the Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa, INESC-ID Lisboa – Human Language Technology Laboratory (INESC-ID Lisboa/HLT). Work supported by national funds through Fundação para a Ciência e a Tecnologia, I.P. (FCT) under projects UID/50021/2025 (DOI: <https://doi.org/10.54499/UID/50021/2025>) and UID/PRR/50021/2025 (DOI: <https://doi.org/10.54499/UID/PRR/50021/2025>). Oto Araújo Vale and Ryan Martinez: 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.

References

- Jorge Baptista. 2005. *Sintaxe dos predicados nominais com ser de*. Fundação Calouste Gulbenkian & Fundação para a Ciência e a Tecnologia, Lisboa.
- Jorge Baptista and Nuno Mamede. 2020. *Dicionário Gramatical de Verbos do Português Europeu*. Universidade do Algarve.
- Francisco da Silva Borba. 2002. *Dicionário de usos do português do Brasil*. (Editora Ática).
- Paula Cristina Quaresma da Fonseca Carvalho. 2007. *Análise e representação de construções adjetivais para processamento automático de texto: adjetivos intransitivos humanos*. Ph.D. thesis, Universidade de Lisboa.
- João Malaca Casteleiro. 1981. *Sintaxe transformacional do adjetivo – regência das construções completivas*. INIC, Lisbon.
- Gohsran Chung. 2003. *Analyse des constructions à double nominatif/accusatif par l'opération de restructuration en coréen Classification syntaxique des constructions à adjectifs sôngsang*. Ph.D. thesis, Université Paris-Est.
- Helena Maria Serras Reis Silva Freire. 1995. *Determinação e formalização das propriedades sintáticas de adjetivos terminados em-vel*. Master's thesis, Universidade de Lisboa.

- Maurice Gross. 1975. *Méthodes en syntaxe: régime des constructions complétives*. Herman, Paris.
- Maurice Gross. 1981. Les bases empiriques de la notion de prédicat sémantique. *Langages*, 63:7–52.
- Maurice Gross. 1988. Methods and tactics in the construction of a lexicon-grammar. *Linguistics in the morning calm*, 2:177–197.
- Maurice Gross. 1996a. Les verbes supports d'adjectifs et le passif. *Langages*, pages 8–18.
- Maurice Gross. 1996b. Lexicon-Grammar. In *Concise Encyclopedia of Syntactic Theories*, pages 244–259. Pergamon.
- Alain Guillet and Christian Leclère. 1981. Restructuration du groupe nominal. *Langages*, 63:99–125.
- Zellig Sabbetai Harris. 1968. *Mathematical Structures of Language*. John Wiley & Sons, New York.
- Zellig Sabbetai Harris. 1976. *Notes du cours de syntaxe*. Éditions du Seuil, Paris.
- Zellig Sabbetai Harris. 1982. *A Grammar of English on Mathematical Principles*. John Wiley and Sons, Inc., New York.
- Zellig Sabbetai Harris. 1991. *Theory of Language and Information: A Mathematical Approach*. Clarendon Press, Oxford.
- A. Kilgarriff, M. Jakubíček, J. Pomikalek, T.B. Sardinha, and P. Whitelock. 2014. Pttenten: A corpus for portuguese lexicography. In T.B. Sardinha and T. de Lurdes São Bento Ferreira, editors, *Working with Portuguese Corpora*, pages 111–128. Bloomsbury Academic.
- Éric Laporte. 1992. Adjectifs en-ant dérivés de verbes. *Langue française*, 96:30–43.
- Eric Laporte. 1997. L'analyse de phrases adjectivales par rétablissement de noms appropriés. *Langages*, 31(126):79–104.
- Danielle Leeman and Serge Meleuc. 1990. Verbes en tables et adjectifs en -able. *Langue française*, 87:30–51.
- Simona Messina. 2019. The predicative adjective and its propositional arguments: A lexicon-grammar classification. *Linguisticæ Investigationes*, 42(2):234–261.
- Annie Meunier. 1981. *Nominalisations d'adjectifs par verbes supports*. Ph.D. thesis, Paris 7.
- Métiyé Meydan. 1995. *Transformations des constructions verbales et adjectivales: élaboration du lexique-grammaire des adjectifs déverbaux*. Ph.D. thesis, Paris 7.
- Jee-Sun Nam. 1996. *Classification syntaxique des constructions adjectivales en coréen*. John Benjamins Publishing Company, Amsterdam.
- Lélia Picabia. 1978. *Les constructions adjectivales en français: systématique transformationnelle*, volume 11. Librairie Droz.
- Baholisoa Simone Rahaga. 1995. *Lexique-grammaire du malgache: constructions adjectivales*. Ph.D. thesis, Paris 7.
- Graça Rio-Torto. 2006. Para uma gramática do adjetivo. *Alfa: Revista de Linguística*, 50(2):103–129.
- Elisávet Sklavouínou. 1997. *Etude comparée de la nominalisation des adjectifs en grec moderne et en français*. Ph.D. thesis, Paris 8.
- Rui Talhadas, Nuno Mamede, and Jorge Baptista. 2013. Semantic roles for portuguese verbs. In *32nd International Conference on Lexis and Grammar*, pages 127–132.
- Freiderikos Valetopoulos. 2003. *Les adjectifs prédicatifs en grec et en français: de l'analyse syntaxique à l'élaboration des classes sémantiques*. Ph.D. thesis, Université Paris 13.
- Rita Veloso and Eduardo B Paiva Raposo. 2013. Adjetivo e sintagma adjetival. In *Gramática do português*, volume 2, pages 1359–1493. Fundação Calouste Gulbenkian.
- Zeno Vendler. 1968. *Adjectives and nominalizations*. Mouton, Paris.