The construction of a catalog of Brazilian Portuguese verbs

Márcia Cançado
Faculdade de Letras, UFMG
Antônio Carlos, 6627
Belo Horizonte, Brazil
mcancado@ufmg.br

Luisa Godoy
Faculdade de Letras, UFMG
Antônio Carlos, 6627
Belo Horizonte, Brazil
luisagodoy@gmail.com

Luana Amaral
Faculdade de Letras, UFMG
Antônio Carlos, 6627
Belo Horizonte, Brazil
luanalopes@ufmg.br

Abstract

In this paper we present the construction of a lexical-semantic resource for the study of Brazilian Portuguese (BP): a “catalog” of BP verbs. The purpose of this catalog is to serve as a complete source of data, in which we present a large amount of verbs and sentences (over 800 verbs in over 5500 sentences in a first volume). An important characteristic of our project is that it is not a mere listing of verbs. Besides actually listing them, we group the verbs into semantically and syntactically coherent classes, adopting the hypothesis presented in Levin (1993) that semantic properties of verbs determine the syntactic realization of their arguments. We propose representations for the verbs using predicate decomposition, so that predicate decomposition structures define the verb classes. The syntactic properties of the verbs are presented in sample sentences and are related to the predicate decomposition structures. The relevance of our catalog is to be a complete resource for lexical-semantic studies in BP, not only a listing of verbs, but also a classification and an exhaustive exemplification. This catalog will take the forms of a printed version (Cançado, Godoy and Amaral, to appear - first volume) and a digital database.

1 Introduction

In this paper we present and comment on an empirical project that has been developed in the field of lexical semantics, taking Brazilian Portuguese (BP) as a language to be described. This project is the construction of a lexical-semantic resource for the study of verbs in BP. We call it a “catalog” of the BP verbal lexicon, and it will take the forms of a printed version and a digital database.

The catalog is divided into a number of volumes, each one presenting verbs grouped according to semantic properties. Here, we present the first volume, which is ready to be published (Cançado, Godoy and Amaral, to appear). Subsequent volumes will be released in the future. This first volume contains 861 verbs of change from BP. All these verbs share a semantic property: they all denote some type of change. They are divided into 4 semantically and syntactically coherent classes à la Levin (1993): change of state, change of locative state, change of location, also known as “location verbs”, and change of possession, also known as “locatum verbs”. The class of change of state is the most representative one, with 685 verbs. We chose the verbs of change to start the construction of our catalog because we noted that they are very common verbs in the language, and they are the most studied verbs in the literature as well. Taking into account that there are actually approximately 6000 verbs in use in BP (Borba, 1990), we believe that the first volume of our catalog contains a significant amount of data.

2 The form of the catalog

Let us now describe and illustrate the form that our catalog takes. For each class of verbs, we present a listing of all the verbs we could find (the

---

1This notion is related with the primitive predicate BECOME. We will talk about that when we present the theoretical background of the catalog, in section 3.
verbs were collected mostly from an extensive BP verb dictionary: Borba, 1990) and of the semantic and syntactic properties of those verbs, as we illustrate below with the verb *quebrar* ‘break’, which belongs to the change of state class:

(1) O João quebrou o vaso.
the John broke the vase
‘John broke the vase.’
*(transitive causative with an agent subject)*

(2) A queda quebrou o vaso.
the fall broke the vase
‘The fall broke the vase.’
*(transitive causative with a cause subject)*

(3) O vaso ficou quebrado.
the vase became broken
‘The vase became broken.’
*(change of state entailment)*

(4) O vaso (se) quebrou.
the vase (SE) broke
‘The vase broke.’
*(intransitive inchoative with the optional se)*

(5) O vaso (se) quebrou com a queda.
the vase (SE) broke with the fall
‘The vase broke from the fall.’
*(intransitive inchoative with the optional se and the cause in adjunct position)*

(6) O João quebrou o vaso com um martelo.
the John broke the vase with a hammer
‘John broke the vase with a hammer.’
*(transitive causative with an instrument in adjunct position)*

(7) O vaso foi quebrado (pelo João).
the vase was broken (by-the John)
‘The vase was broken by John.’
*(passive construction)*

(8) O menino quebrou o braço (com a queda).
the boy broke the arm (with the fall)
‘The boy’s arm broke from the fall.’
*(patient-possessor alternation with an optional cause in adjunct position)*

(9) O João quebrou as paredes da casa com um ótimo pedreiro.
the John broke the walls of-the house with an excellent bricklayer
‘John had the walls of the house broken by an excellent bricklayer.’
*(agent-possessor alternation)*

We show that some specific construction is either allowed or prohibited by all members within a class. For example, for the classes which do not participate in the causative-inchoative alternation, we list ungrammatical examples of that alternation. The verb *aparelhar* ‘equip’, a verb of change of possession (also known as locatum), does not participate in the causative-inchoative alternation and it cannot have a cause as subject, as we show in the examples marked with * below:

(10) O dentista aparelhou o consultório.
the dentist equipped the office
‘The dentist equipped the office.’
*(transitive causative with an agent subject)*

(11) *A necessidade de modernizar aparelhou o consultório.
the need to modernize equipped the office
*(transitive causative with a cause subject)*

(12) O consultório ficou com aparelho.
the office became with equipment
‘The office became with equipment.’
*(change of possession entailment)*

(13) O dentista aparelhou o consultório com aparelhos de aço inox.
the dentist equipped the office with equipment of steel stainless

*The alternations agent-possessor and patient-possessor are very productive in BP. See Cançado (2010). They are also known in the literature as “possessor raising”.*
‘The dentist equipped the office with stainless steel equipment.’

(cognate/instrument construction)

(14) *O consultório (se) aparelhou.
the office (SE) equipped

(intransitive inchoative with the optional se)

(15) O consultório foi aparelhado (pelo dentista).
the office was equipped (by-the dentist)
‘The office was equipped by the dentist.’

(passive construction)

(16) O dentista se aparelhou.
the dentist SE equipped
‘The dentist provided himself with equipment.’

(reflexive)

(17) O dentista aparelhou o consultório com a melhor loja de aparelhos odontológicos da cidade.
the dentist equipped the office with the best store of equipment orthodontic of-the city
‘The dentist had the office equipped by the best store of orthodontic equipment in the city.’

(agent-possessor alternation)

The syntactic properties reflexive, patient-possessor alternation and agent-possessor alternation are not exemplified when they do not occur (for example, we do not present a reflexive sentence for the verb quebrar in (1)-(9), nor an example of the patient-possessor alternation for the verb aparelhar in (10)-(17)) because they are not classificatory of the classes, they do not occur with all the verbs in a class and they also have pragmatic or other semantic restrictions (see section 3.2). The change entailment is an important semantic characteristic of the verbs, which distinguishes the semantics of the classes (e.g., distinguishes change of state from change of possession).

We provide the classes with an exhaustive exemplification, presenting carefully crafted examples of each one of the syntactic properties picked up for observation, for each verb. So, for each verb belonging to the class of change of state, for example, we present an example of a sentence with the semantic and syntactic properties of the class, just like was done for quebrar and aparelhar above. In the first volume of the catalog we present over 5500 examples.

Since our purpose is to describe the BP lexicon as a system, we test the limits of grammaticality and deal with negative and intuitive data. So, as illustrated above, we present examples carefully crafted by us, not empirical data collected from corpus. Our methodology is very close to the one adopted in the research done by Maurice Gross (Gross, 1975, 1981), although the theoretical frame differs a lot. Research based on intuitive/introspective examples has been much criticized with the ascendance of corpus linguistics, but we believe that this is the adequate methodology for the kind of research we develop and it holds many advantages, as argued for by Laporte (2008).

Our catalog contains, in the first volume dedicated to verbs of change, 861 verbs divided into 4 classes. The class of change of state, the most representative one, contains 685 verbs; the class of change of locative state contains 68 verbs; the class of change of location contains 15 verbs; and the class of change of possession contains 93 verbs. For each verb we list syntactic and semantic properties, as illustrated above, totaling over 5500 sentences.

Besides the presentation of the data, which is the most important part of our catalog, in the printed version we also present a theoretical explanation of our analysis (which will not be available in the digital database).

---

3We do not explain why there is such difference in the number of members of the classes of verbs of change, but the main hypothesis is that BP has a preference for lexicalizing change of state. The verbs from the class of change of location appear to be diachronically going to the class of verbs of change of locative state, another clue that BP likes having change of state verbs. For example, the verb acomodar ‘place in a comfortable position’ comes from the name cômodo ‘room’ (Cunha, 2010), so the original meaning of this verb would be ‘place in room’, but when forming a sentence with acomodar we can add any locative, not only one that is a room. For example, a mãe acomodou a filha na cama ‘the mother placed the daughter in bed in a comfortable position’.
3 Theoretical Background

Let us now take a look at the theoretical background of our work. Our catalog starts from Levin’s (1993) original insights about the importance of lexical description and her methodological ideas, taking the hypothesis of lexical semantic determinants of syntactic behavior seriously. We assume a strong version of Levin’s (1993) methodology, admitting that if two verbs or groups of verbs behave alike, they must belong to the same class. For example, the verbs quebrar ‘break’ and abrir ‘open’ are transitive, participate in the causative-inchoative alternation, and can be passivized. This similar syntactic behavior, according to Levin (1993), is a clue that these verbs belong to the same class and share some grammatically relevant semantic property. So, an important part of our work is to divide the verbs in semantically and syntactically coherent classes, showing that semantic properties of verbs determine their syntactic behavior. It is important to note that our catalog is not a transposition of Levin’s (1993) classes into BP. Our project is not a translation and all the verbs were collected directly from BP.

Another very important characteristic of our catalog is the metalanguage we use for representing the meaning of the verbs and their relevant semantic components: the semantic decomposition of verbs in primitive predicates, known as “predicate decomposition”. Predicate decomposition is a more formal metalanguage to represent meaning than natural language, and it is also less problematic and more complete than a lexical representation in terms of thematic grids (Levin and Rappaport Hovav, 2005). Predicate decomposition structures can accommodate distinct types of semantic information, both thematic and aspectual, contained in a verb’s meaning within a single representation. So it is not restricted to representing the semantic function of the arguments (like in thematic grids), but it represents also, and most importantly, the semantics of the event and its subparts. An important characteristic of the verbs of change is the presence of the predicate BECOME in their representations. So the common semantic feature that these verbs share is very clear in the semantic representations, as well as the differences between them, as we will see below.

An important point in the predicate decomposition representations is the opposition between structural semantic information and idiosyncratic semantic information. The introduction of the concept of a “root” or “constant” was a major step in lexical representation (Levin and Rappaport Hovav, 2005) and is now widely accepted. We assume (along with Grimshaw, 2005 and others) that every grammatically irrelevant semantic information belongs to the root, so the relevant semantic information is represented in the structural part. The structural part is shared by the verbs within the same class, so it does not only represent grammatically relevant semantic information, but also identifies a class.

In the catalog, we articulate the syntactic properties listed with the predicate decomposition structures proposed for each class. We motivate the existence of each component within a structure, in order to represent only those aspects of meaning which are relevant. So we try to propose well-motivated and economic structures (we represent in the structural part the relevant features leaving all the rest for the root). In the printed version of the catalog we provide explanation and independent evidence for the structures in (18)-(21).

Below, we present the representations for each verb class together with a short list of examples:

The structures of the verbs of change:

\[(18) \quad [\ [X_{VOLITION}] \ CAUSE \ [BECOME \ Y \ \langle STATE \rangle ] ] \]

(change of state: quebrar ‘break’, abrir ‘open’, machucar ‘hurt’)

\[(19) \quad [\ [X_{VOLITION}] \ CAUSE \ [BECOME \ Y \ \langle STATE \rangle \ IN \ Z ] ] \]

(change of locative state: acomodar ‘place in a comfortable position’, esconder ‘hide’, pendurar ‘hang’)

---

\*We are aware, however, that some properties of the root may be relevant to some syntactic properties, which we call “non-classificatory”. For example, the reflexive construction only occurs with verbs that select animate objects (Godoy, 2012). See section 3.2.
(20) \([X_{\text{VOLITION}}] \text{CAUSE } [\text{BECOME } Y \text{ IN } \langle \text{PLACE} \rangle] \)

(change of location: engaiolar ‘cage’, engarrafar ‘bottle’, ensacar ‘bag’)

(21) \([X_{\text{VOLITION}}] \text{CAUSE } [\text{BECOME } Y \text{ WITH } \langle \text{THING} \rangle] \)

(change of possession: aparelhar ‘equip’, amanteigar ‘butter’, colorir ‘color’)

All the structures contain the predicate BECOME, which represents the change. But the structures that occur with BECOME are different, signaling that each class denotes a different kind of change. All verbs of change are transitive and causative, so the representations show the change (the substructure which contains BECOME) and the agent/cause responsible for bringing that change about (X), as well as the relation between them, represented by the predicate CAUSE. Let us now present what kind of semantic and syntactic properties we can predict for the verb classes by observing the predicate decomposition structures.

3.1 Semantic properties

There are three kinds of semantic information that we can derive from those structures: the thematic roles of the arguments, lexical aspect, and the kinds of changes denoted by the verbs. All that information is given for each class of verbs in the catalog.

Thematic roles are defined by Jackendoff (1990) as argument places in the predicate decomposition structure. X as the first argument of CAUSE receives the cause thematic role, X modified by VOLITION receives the agent thematic role, Y as the argument of BECOME receives the patient thematic role, and Z, argument of IN, receives the locative thematic role. All the verbs of change have a patient object, the verbs of change of state may have cause or agent subjects, the verbs of change of locative state, change of location and change of possession have agent subjects, and the verbs of change of locative state also have a third argument, which is a location.

Lexical aspect can also be defined in terms of predicate decomposition structures, as done by Dowty (1979). All the verbs of change denote accomplishments, since they denote causative events, with two subevents related by the predicate CAUSE in the predicate decomposition structure. When the verbs participate in the causative-inchoative alternation (only verbs of change of state), the inchoative counterparts denote achievements, since they only denote the final result, the change of state.

Each kind of root, together with other elements, like the prepositions, will indicate if the change in question is a change of state, locative state, location or possession. For each verb in the catalog we show the sentence with the change entailment, so the verbs of change of state ([BECOME Y ⟨STATE⟩]) entail the sentence become state, verbs of change of locative state ([BECOME Y ⟨STATE⟩ IN Z]) entail become state in Z, the verbs of change of location ([BECOME Y IN ⟨PLACE⟩]) entail become in location, and the verbs of change of possession ([BECOME Y WITH ⟨THING⟩]) entail become with thing.

3.2 Syntactic properties

Taking the hypothesis that the semantic properties of verbs determine their syntactic behavior, we are able to link certain predicate decomposition structures with certain syntactic properties.

The causative-inchoative alternation only occurs with the verbs of change of state, so it is sensible to the substructure [BECOME Y ⟨STATE⟩]. Passive and instrument adjunction occur with agentive verbs, so these constructions are sensible to the substructure [X_{\text{VOLITION}}]. The adjunction of a cause will only occur with verbs that permit a cause subject, so verbs which have the substructure [X_{\text{VOLITION}}].

The other syntactic properties, reflexive, patient-possessor alternation and agent-possessor alternation, are sensible to the structure, but also have pragmatic and other kind of semantic restrictions. For example, the reflexivization occurs with agentive verbs ([X_{\text{VOLITION}}]), but it also requires that the verb selects an animate object, as proposed by Godoy (2012). For those different kinds of restrictions, the verbs within a class do not have exactly the same behavior in respect to these syntactic properties. We are aware of that and that fact is accounted for in our catalog. Syntactic properties that we call “non-classificatory”
are sensible to semantic properties or pragmatic properties that crosscut verb classes or that do not belong to all the verbs in a class. This does not invalidate the hypothesis that the verb classes are syntactically coherent, as we explain in length in the printed version of the catalog.

The verbs of change of state show a particularity in respect with their X argument. While all other classes have agent subjects, verbs of change of state may have agents or causes as subject, as observed by Cançado (2005, 2010). That is why we represent VOLITION between parentheses in the representation of that class ([X(VOLITION)]). This property actually divides the class of change of state verbs into four subclasses: strictly volitive, optionally volitive, non-volitive (a class pointed out by Cançado and Franchi, 1999), and inchoative. The verbs of change of state belong to a broad class because all of them participate in the causative–inchoative alternation and entail become state as showed by Cançado and Godoy (2012). Subclasses arise because these verbs differ with respect to the agentivity of the external argument, and also with respect to the occurrence of the clitic se in the intransitive-inchoative sentences. As we showed in example (4), optionally volitive verbs like quebrar ‘break’, strictly volitive verbs (like legalizar ‘legalize’) and non-volitive verbs (like depimir ‘depress’) may or may not occur with se in intransitive sentences. Inchoative verbs (like apodrecer ‘rot’), however, never occur with se (*a maçã se apodrece). The optionality of se in the three first subclasses of verbs of change of state seems to be dialectological, implying no meaning difference. The prohibition of se in intransitive sentences with inchoative verbs, however, seems to be a consequence of the fact that these verbs are basically intransitive, as argued for by Cançado and Amaral (2010).

The classes of verbs of change of locative state, change of location and change of possession are also differentiated by syntactic properties. The class of change of locative state is the only one with three arguments: a mãe acomodou a menina na cama ‘the mother placed the daughter in bed in a comfortable position’, as stated in Godoy (2012). The verbs of the class of change of location form sentences with a cognate location, like:

\[ o homem engaiolou os pássaros em uma gaiola de ouro ‘the man caged the birds in a golden cage’ \]

and form a different kind of reflexive, which we call “middle reflexive” (also known as “naturally reflexive”; Reinhart and Reuland, 1993). The verbs of the class of change of possession form sentences with a cognate possession, like: o dentista aparelhou o consultório com aparelhos de aço inox ‘the dentist equipped the office with stainless steel equipment’ and form canonical reflexives.

4 Some important observations

Here, we should mention a work similar to ours, the database called VerbNet\(^5\), which is a current descriptive project that was also inspired by Levin’s (1993) work. At first sight, our catalog might be seen as a translated version of it, not only because VerbNet is inspired in Levin’s verbal classes, but also because it uses predicate decomposition as semantic information. However, our catalog bears important differences in relation to that database. First, in VerbNet, the only syntactic property listed in the verbs entries is a sort of a subcategorization frame, and our catalog presents several syntactic properties that group the verbs into classes. Besides, we assume that predicate decomposition structures play a much more important role: first, they define the verb classes; second, thematic and aspectual properties can be derived from them (Jackendoff, 1990; Dowty, 1979). VerbNet does not assume a theoretical background in proposing representations – thematic roles and predicate decomposition structures are used only as semantic descriptions.

Another departure from VerbNet, and also from Levin (1993), is an obvious but not as trivial aspect as one might think – the fact that our catalog describes BP, not English. We take on the hypothesis that even when translation can occur, lexicalization might differ, if one is observing two different languages. Levin and Rappaport Hovav (1995) present the example of the two verbs, English blush and its Italian version arrossire, which are good translations of each other (e.g., they may have the same truth conditions), however they seem to belong to different classes in each lan-

---

\(^5\)http://verbs.colorado.edu/ mpalmer/projects/verbnet.html
guage, *blush* behaving as an unergative in English and *arrossire* behaving as an unaccusative in Italian. The analysis provided by VerbNet of the English verbal lexicon, therefore, cannot be trivially applied to the BP verbal lexicon. Plus, BP is an interesting alive language, whose grammar differs much from European Portuguese (for example, it was shown in Raposo, 1998 that the pronominal system of BP is very different from that of the other European Romance languages), which is something that is not accounted for in many studies that refer to Romance languages as a whole.

5 Applications in linguistic research

After presenting the project we developed for BP, we point out some of the applications of our catalog in linguistic research. First of all, any researcher who wants to study verbs in BP would benefit from a listing of the verbs of the language and their properties already classified into semantic classes. Also, our catalog will be the base for building an electronic database. We plan on building an online verb database where researchers could search verbs by their names, their thematic roles, their syntactic properties, their representations, and so on. For example, if one wants to study the passive construction in BP, he/she can go to our online database and search for verbs that occur in the passive construction. The result of the search will show all the verbs of change (in this first volume) that have this syntactic property as well as sample sentences. Another important application of our catalog is that it can be used by researchers who are not native BP speakers. With a list of verbs and sentences at hand one could, for example, compare BP with his/her own language in terms of lexicalization patterns, syntactic constructions available, classes of verbs, and so on. It would also be very helpful for the study of language typology, since the researcher could easily acquire a great amount of data without having to collect it.

6 Conclusion

In this paper, our purpose was to present the construction of a lexical-semantic resource for the study of verbs in BP. We wanted to show how we use semantic and syntactic properties of verbs to group and to separate them in classes and subclasses, relating this grouping and some cross-classifications that might occur to components in a semantic predicate decomposition representation. The listing of syntactic properties exhaustively exemplified, along with the central role played by well-motivated and parsimonious predicate decomposition structures are the two central characteristics of our project. The main purpose of an exhaustive exemplification is to provide a complete and straightforward source of data of the BP verbal lexicon, which we hope is going to be very helpful for researchers. In doing this exhaustive exemplification, we also try to leave little space for the refutation, on an empirical ground, of the existence of the classes we propose and of the hypothesis of grammatically relevant components of lexical meaning that determine syntactic behavior.

It is worth noting that the catalog is not an isolated project. The group of lexical semantics, coordinated by Cançado, has been doing descriptions of the BP verbal lexicon and studying lexical representations for over 17 years now, so the catalog contains a sort of a compilation of the main results of her work, both individual and advising students, in the field of lexical semantics.

Acknowledgments

We thank CNPq and FAPEMIG the financial support for Márcia Cançado; we thank CNPq the financial support for Luisa Godoy; and we thank CAPES the financial support for Luana Amaral. Besides, we must thank Diógenes Evangelista for help with formatting the paper.

References


Cancado, Márcia; Godoy, Luisa and Amaral, Luana. to appear. Catálogo de verbos do português brasileiro, volume 1 (A catalog of Brazilian Portuguese verbs, volume 1). (printed version of the catalog presented in this paper)


