

# Online resources for the syntactic-semantic classification of verbs: theory, methods and applications

Recursos online para a classificação sintático-semântica de verbos: teoria, métodos e aplicações

Luana Lopes Amaral  \*1

<sup>1</sup>Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil.

## Abstract

Verb classes are defined in lexical-semantics literature as syntactically coherent groups of verbs that also share meaning components. Thus, the investigation of verb classes leads to important insights into the syntax-semantics interface. In this perspective, starting from Levin's (1993) foundational work for English, many authors have developed analyses for the syntactic-semantic classification of verbs in different languages. The aim of this paper is to provide an overview of such studies, focusing on those that constitute online open access resources for the syntactic-semantic classification of verbs: VerbNet for English, BVI for Basque, AnCoraVerb for Spanish and Catalan, VerboWeb for Brazilian Portuguese, CROVALLEX for Croatian, and VALLEX for Czech. Thus, this paper reviews the general theoretical assumptions on verb classification shared by all the works presented, and shows the specific characteristics of each one, considering the body of data, the specific theoretical perspective, analyses and methodology, and the search engines they provide. Finally, it offers a comparative analysis of two specific verb classes (namely, image creation verbs and verbs of creation and transformation), drawing on data from the studies mentioned above. Although these resources assume a specific lexical-semantic approach for verb classification, they provide valuable data and analyses that can be useful for research in different theoretical perspectives.

**Keywords:** Verb classes. Event structure. Argument structure. Online resources. Digital lexicons. Verbs of creation.

## Resumo

Classes verbais são definidas na literatura em semântica lexical como grupos sintaticamente coerentes de verbos que também compartilham componentes de significado. Assim, a investigação de classes verbais leva a importantes descobertas sobre a interface sintaxe-semântica. Nessa perspectiva, a partir do trabalho fundador de Levin (1993) para o inglês, muitos autores desenvolveram análises para a classificação sintático-semântica de verbos em diferentes línguas. O objetivo deste artigo é fornecer uma visão geral de tais trabalhos, com foco naqueles que constituem recursos online de acesso aberto para a classificação sintático-semântica de verbos: VerbNet para o inglês, BVI para o basco, AnCoraVerb para o espanhol e o catalão, VerboWeb para o português brasileiro, CROVALLEX para o croata e VALLEX para o tcheco. O artigo revisa os pressupostos teóricos gerais sobre classificação de verbos compartilhados por todos os trabalhos apresentados e apresenta as características específicas de cada um, considerando o corpo de dados, as análises e metodologias específicas e os mecanismos de busca que eles fornecem. Por fim, o presente artigo oferece uma análise comparativa de duas classes verbais específicas (a saber, verbos de criação de imagem e verbos de criação e transformação), com base nos dados dos trabalhos citados acima. Embora esses recursos assumam uma abordagem léxico-semântica específica para classificação de verbos, eles fornecem dados e análises valiosos que podem ser úteis para pesquisas em diferentes perspectivas teóricas.

**Palavras-chave:** Classes verbais. Estrutura de eventos. Estrutura argumental. Recursos online. Léxicos digitais. Verbos de criação.

*Textolivre*  
Linguagem e Tecnologia

DOI: 10.35699/1983-3652.2022.38715

Session:  
Articles

Corresponding author:  
Luana Lopes Amaral

Section Editor:  
Daniervelin Pereira  
Layout editor:  
Carolina Garcia

Received on:  
February 10, 2022  
Accepted on:  
April 22, 2022  
Published on:  
July 6, 2022

This work is licensed under a  
"CC BY 4.0" license.



## 1 Introduction

Since Fillmore's (2003) first findings, it has been argued that semantic properties of verbs have an impact on syntactic structure. In his seminal work on lexical semantics, the author showed that only

\*Email: luanalopes@ufmg.br

the change of state verbs, such as *break*, can occur in the well-known causative alternation, while verbs which do not describe changes of state, such as *hit*, do not present the same pattern of change, as examples (1) and (2) illustrate, respectively.

- (1) a) John broke the stick.  
b) The stick broke.
- (2) a) John hit the tree (with a rock).  
b) \*The tree hit.

(FILLMORE, 2003, p. 126-128)

Following the idea that lexical semantic properties of verbs determine sentence form, Levin (1993) took over the immense endeavor of classifying English verbs according to the relation between their semantics and their syntax. In her study, she systematized the correlations first found by Fillmore (2003), showing that what happens with *break* and *hit* are properties shared with a semantically defined group of verbs, or a *verb class*.

- (3) Break verb class: *break, crack, rip, shatter, snap...*  
(4) Hit verb class: *bash, hit, kick, pound, tap, whack...*

(LEVIN, 1993, p. 7)

As Levin (1993, p. 17) herself puts it, “verb classes arise because a set of verbs with one or more shared meaning components show similar behavior”. Besides *break* verbs and *hit* verbs, many other classes emerged in Levin’s (1993) analysis, each one related to specific syntactic and semantic properties. Two of those classes are the image creation verbs, exemplified in (5), and the verbs of creation and transformation, exemplified in (6). These classes will be of special interest in this paper and will be further described in Section 4.

- (5) Image creation verbs: *embroider, doodle, illustrate, mark, paint, type...*  
(6) Verbs of creation and transformation: *build, create, compose, develop, transform...*
- (LEVIN, 1993, p. 169–179)

Levin’s (1993) work inspired the syntactic-semantic classification of verbs that continued to be developed in different theoretical perspectives, and also the study of verb classes in different languages. One of the most relevant consequences of her work was the creation of VerbNet (SCHULER, 2005; KIPPER et al., 2008), an online open access resource which gathers Levin’s (1993) classes and many others later developed and makes this information available in an easily searchable platform. Also, based on her initial ideas and inspired by VerbNet, many researchers around the world proposed similar classifications and resources for the verbal lexicon of languages other than English. BVI (ESTARRONA et al., 2015), AnCoraVerb (APARICIO; TAULÉ; MARTÍ, 2008; TAULÉ; MARTÍ; BORREGA, 2010), VerboWeb (CANÇADO; AMARAL; MEIRELLES, 2017), CROVALLEX (PRERADOVIC; BORAS; KISICEK, 2009; PRERADOVIC, 2010), and VALLEX (KETTNEROVÁ; LOPATKOVÁ; BEJČEK, 2012) are of particular interest in this paper, and they are all online resources, or digital lexicons, like VerbNet, for the analysis and syntactic-semantic classification of verbs in different languages (BVI for Basque, AnCoraVerb for Spanish and Catalan, VerboWeb for Brazilian Portuguese, CROVALLEX for Croatian, and VALLEX for Czech).

The investigation of verb classes leads to important insights into the syntax-semantics interface, since each class is determined by correlations between verb meaning and sentence form. In this case, resources such as VerbNet can be viewed as valuable sources of information, both data and analyses, for Theoretical Linguistics. Besides, they are all accessible online, being a great example of research aligned with the growing interest of the scientific community in Open Science practices (SERRAPILHEIRA, 2020). Assuming the relevance of such research work, the aim of this paper is to provide an overview of the syntactic-semantic classification of verbs in different languages, focusing on the online open access resources: VerbNet, BVI, AnCoraVerb, VerboWeb, CROVALLEX, and VALLEX. Theoretical linguists are not very familiar with these resources, as they are mostly built for NLP (natural language processing) purposes, and the discussion concerning these data occurs mainly in Computational Linguistics. Thus, this paper also aims at bringing these resources into Theoretical Linguistics, especially considering the rich body of data and the rich theoretical analyses they provide. In order to provide a clear view of the use of such resources, we also present a comparative analysis

of image creation verbs and verbs of creation and transformation, drawing on data made available in the presented resources.

Besides the ones presented, there are other important resources that should be acknowledged. FrameNet (RUPPENHOFER et al., 2016), WordNet (FELLBAUM, 1998), and PropBank (BONIAL et al., 2014) are important lexical resources that are not specific to verbs, but analyze other aspects of the lexicon as well. Also, there are other studies which are focused on verbs, but take a typological perspective, such as the Leipzig Valency Classes Project (HARTMANN; HASPELMATH; TAYLOR, 2013; MALCHUKOV; COMRIE, 2015). We acknowledge the importance of these studies, but they are not going to be analyzed in this paper, as we focus here on online open access resources for the syntactic-semantic classification of verbs in specific languages<sup>1</sup>.

The paper is organized as follows: in Section 2, we review the theoretical background which is common throughout the studies hereby presented; Section 3 provides an overview of VerbNet, BVI, AnCoraVerb, VerboWeb, CROVALLEX, and VALLEX, including specific theoretical assumptions, methodology for data gathering and analysis, results and classes presented in each database, as well as comments on the practical search mechanisms. In Section 4, the comparative analysis of image creation verbs and verbs of creation and transformation is presented; Section 5 brings some final remarks.

## 2 Shared theoretical assumptions

This section reviews the general background for the syntactic-semantic classification of verbs. As all the works presented here are initially inspired by Levin (1993), they all share general theoretical assumptions, despite differing in more specific topics (to be presented in Section 3).

First of all, all these studies share the same assumption about what a verb class is and what the criteria for verb classification are. As explained in Section 1, verb classes are syntactically coherent groups of verbs which share semantic properties. According to De Clerck, Coleman, and Willems (2013), this type of classification is made assuming lexical-grammatical criteria, i.e., assuming that lexical semantics defines the behavior of verbs and the syntactic realization of arguments. An important aspect of the syntactic-semantic correlation which defines classes is that not all semantic properties are grammatically relevant. It means that the semantic properties which define verb classes are those which have an impact on the syntactic realization of arguments (also called verbal diathesis or valency).

Cançado, Amaral, and Meirelles (2018), drawing on Levin and Rappaport Hovav's (2005) analysis of English, show for instance that *motion verbs* cannot be a verb class in Brazilian Portuguese. Different verbs which describe the motion of an object do not have a uniform syntactic behavior, as the examples from Brazilian Portuguese illustrate:

(7) O maratonista correu

the marathon.runner ran

'The marathon runner ran.'

(8) O menino balançou a cortina.

the boy swung the curtain

'The boy swung the curtain.'

(9) O atacante arremessou a bola para o gol.

the forward threw the ball to the goal

'The forward threw the ball to the goal.'

(CANÇADO; AMARAL; MEIRELLES, 2018, p. 127)

As can be seen in the examples (7), (8) and (9), a verb describing a motion event can be either intransitive (as *correr* 'run'), transitive (as *balançar* 'swing') or ditransitive (as *arremessar* 'throw'). Hence, the semantic property of motion alone does not define a verb class, as these verbs do not have

<sup>1</sup> At the end of the paper, we provide an Appendix with a brief list of useful online resources for the study of verb classes, including those analyzed here and those that were left out.

a syntactically coherent behavior.

However, analyzing more specific aspects of the semantics of motion verbs, other types of generalizations arise. *Correr* 'run' is an agentive verb, and it describes the internal motion of the agent in itself; *balançar* 'swing' describes the manner of motion of an inanimate object, caused by an external force; and, finally, *arremessar* 'throw' describes the motion of an object through a path, also caused by an external force. These more general properties of the event described by these verbs are relevant for syntactic structure. So, agentive motion verbs will be intransitive like *correr* 'run', manner of motion verbs will be transitive like *balançar* 'swing', and verbs which describe a motion through a path will be ditransitive like *arremessar* 'throw' (LEVIN; RAPPAPORT HOVAV, 2005; LEVIN, 1993; MEIRELLES; CANÇADO, 2017).

For Levin (1993), *run* and *swing* belong to distinct subclasses of manner of motion verbs in English, and *throw* belongs to the *throw* verbs, a subclass within the more general class of verbs of throwing (which also includes the subclass of *pelt* verbs).

(10) Manner of motion verbs (LEVIN, 1993, p. 264):

*Run* verbs: *fly, gallop, hike, jog, run, sleepwalk...*

*Roll* verbs: *bounce, drop, float, move, roll, swing...*

(11) *Throw* verbs: *pass, pitch, shoot, sling, throw, toss...* (LEVIN, 1993, p. 146)

The main motivation to distinguish these classes lies in their syntactic behavior, as shown in (7), (8) and (9). Besides the basic transitivity, other syntactic properties define the classes. For instance, in both English and Brazilian Portuguese, only motion verbs of the *roll* subclass occur in a type of causative alternation, as in the alternate sentences *Bill rolled the ball down the hill/The ball rolled down the hill* (LEVIN, 1993, p. 265). In English, only motion verbs of the *throw* type occur in the dative alternation, as in the alternate sentences *Steve tossed the ball to Anna/Steve tossed Anna the ball* (LEVIN, 1993, p. 147). The dative alternation does not occur in Brazilian Portuguese with these verbs.

These examples demonstrate which type of syntactic and semantic properties are taken into account in the classification of verbs. Syntactic behavior is understood as the syntactic functions of arguments, considering also the variable behavior, or argument alternations. Hence, for the classification of verbs, the syntactic properties taken into account are *argument realization* properties, i.e., transitivity and argument alternations (LEVIN; RAPPAPORT HOVAV, 2005), also called diathesis or valency. Argument realization properties relate to which argument of a verb goes to which syntactic function (subject, object and oblique positions, as in examples in (7), (8) and (9)). Alternations are variations in these syntactic patterns, as the cases illustrating the causative alternation (*John broke the stick/The stick broke*), the manner of motion verbs alternation (*Bill rolled the ball down the hill/The ball rolled down the hill*) and the dative alternation (*Steve tossed the ball to Anna/Steve tossed Anna the ball*). In each of these cases, an argument of the verb may be in different syntactic positions, depending on the sentence (as the case of *the stick*, which appears in either subject or object positions in the causative alternation).

Semantically defined groups of verbs are understood as verbs that share semantic properties related to *event structure*, as the more general properties of motion verbs - agentivity, manner, external causation, and path (LEVIN; RAPPAPORT HOVAV, 2005). Event structure is a generalization over specific situations described by verbs, and groups of verbs may share the same event structure in the sense that they describe events of the same type, or which have the same types of participants and the same temporal contour. It is assumed that argument realization and argument alternations can be predicted from the event structure lexicalized by verbs and verb classes.

In more recent studies, linguists working with argument structure, in different theoretical perspectives, agree that semantic properties which govern argument realization are event structure properties (GOLDBERG, 1995; CROFT, 2012; RAMCHAND, 2018; BEAVERS; KOONTZ-GARBODEN, 2020). However, those more recent developments on the investigation of argument structure, which can be described as non-lexicalist, recognize that event structure may not be coded in verbal semantics, or at least not only in verbal semantics. In different ways, these studies analyze event structure as a property of both verbs and argument structure constructions, delegating many event structure prop-

erties to syntax or to a shared work between syntax and the lexicon. Without going into the details of those proposals, the important assumption to highlight, common in all frameworks, is that there is a strong link between event structure and argument realization.<sup>2</sup> The syntactic-semantic classification of verbs in the resources presented follows this path, despite assuming the lexicalist view that verbs and verb classes play a huge role in this relation.

Going back to the examples of motion verbs, *roll* verbs, *run* verbs, and *throw* verbs have different event structures: *run* verbs describe a type of event in which an agent acts and moves at the same time; *roll* verbs describe a type of event in which an agent causes the motion of an object; and *throw* verbs describe a type of event in which an agent causes the motion of an object through a path.

The semantic analysis of event structure assumes that verb meaning is compositional, so the meaning of verbs derives from a combination of smaller components of the event structure. More specifically, verb meaning comprises properties of smaller parts of the event, or subevents, and relations between subevents and participants, also known as force-dynamics or event complexity (LEVIN; RAPPAPORT HOVAV, 2005; CROFT, 2012). Rappaport Hovav and Levin (1998) propose to represent the semantics of verb classes using more abstract properties of the event structure, in a predicate decomposition metalanguage, which will be further explored in Section 3. The event structure for *run* verbs, for example (an agent acts and moves at the same time), can be represented as in (12). The *manner* component specifies the way in which an individual (x) acts, a type of motion in the case of *run*.

- (12) [x ACT<sub><MANNER></sub>]  
(RAPPAPORT HOVAV; LEVIN, 1998, p. 108)

Components of the event structure provide information about two important semantic properties widely used in verb classification: thematic roles and lexical aspect. Thematic roles are the semantic functions associated with the participants of an event (FILLMORE, 1968; JACKENDOFF, 1990; DOWTY, 1991). So, going back to motion verbs, the participant with control over the motion and who starts the motion is an Agent (the subject of the three types of verbs: *run*, *roll*, and *throw*). The participant who is in motion is usually called the Theme (the subject of *run* and the object of *roll* and *throw*). Finally, the place to which the Theme goes is characterized as the Goal (the oblique for *throw* verbs). There are many other types of thematic roles, as we will see in Section 3.

Lexical aspect, in its turn, is defined as the internal temporal structure of situations (VENDLER, 1967; DOWTY, 1979; VAN VALIN, 2005; ROTHSTEIN, 2004). In Vendler's (1967) traditional approach, verbs can be divided into four aspectual types (activities, accomplishments, achievements, and states), according to three characteristics: dynamicity, duration, and telicity. Dynamic situations (or events) are energetic and need change and force to take place; non-dynamic situations (or states) are non-energetic, so there are no changes in the course of time. Durative situations have long duration in time, and non-durative (or punctual) situations occur in a short period of time. Telic situations develop in the direction of a result or ending point, while atelic situations do not. So, activities (exemplified by *run* and *roll* verbs) are dynamic, durative and atelic; accomplishments (exemplified by *throw* verbs) are dynamic, durative and telic; achievements (exemplified by *break* in an intransitive form) are dynamic, punctual and telic; states (exemplified by verbs such as *know*, *love* and phrases such as *be pretty*) are non-dynamic, durative and atelic. Although Vendler's proposal is widely used in verb classification, verbs can have variable aspectual behavior. *Break*, for example, is an achievement in the intransitive form, but it can be an accomplishment in a transitive sentence. Besides, other types of aspects (in addition to the ones proposed by Vendler (1967)) were also discovered (as semelfactives; Smith (1991), and degree achievements, Dowty (1991)).

Syntactic and semantic properties that define verb classes are represented in a variety of ways. Types of representation of the syntactic structure include subcategorization frames, and abstract sentence structures. Types of representations for the event structure of verbs are generally given in the form of thematic grids, aspectual classes and predicate decomposition (as shown in (12)). In BVI, for instance, the syntactic-semantic properties of the Basque manner of motion verb *dardaratu* 'tremble' are represented as:

<sup>2</sup> See more details about different perspectives in Levin (2018).

(13) arg1: theme, - (abs)

The representation in item (13) signals that *dardaratu* has a single argument (a direct object; arg0 stands for subject and arg1 stands for object), bearing the Theme thematic role and marked with absolutive case (abs). These representations will be detailed in Section 3.

An important implication of the assumption that event structure determines argument realization is that different syntactic properties may emerge if a single verb is able to describe different types of events. As polysemy is a pervasive process in the world's languages, a single verb normally does not display the same type of event structure in all its uses. According to the hypothesis that event structure determines argument realization, as the verb changes its meaning, it can also change its syntactic properties, as long as the change in meaning affects semantically relevant properties of the verb (AMARAL, 2016). Thus, lexical-semantic theories of verb classification treat verbal polysemy as the property of verbs which are able to belong to different classes (LEVIN; RAPPAPORT HOVAV, 2013; CANÇADO; AMARAL, 2016). The Croatian verb *potisnuti*, for instance, has four entries in CROVALLEX, each one associated with a different meaning and a different class:<sup>3</sup>

(14) 1 *potisnuti* (*pòtisnuti*) 'press down'

class: put\_result

2 *potisnuti* (*pòtisnuti*) 'force something to stay inside'

class: conceal

3 *potisnuti* (*pòtisnuti*) 'force to retreat (in battle)'

class: force

4 *potisnuti* (*pòtisnuti*) 'force when to move, remove; push away'

class: battle

A final important theoretical assumption for verb classification is the postulation of different levels of classification, more specifically classes and subclasses. As proposed by Levin (1993), different levels of classification are needed in order to fully understand the verbal lexicon of a language. This happens because different types of grammatically relevant semantic properties might crosscut different classes and also might be more general or more specific. An interesting example provided by Cançado and Gonçalves (2016), and also part of the classification of Brazilian Portuguese in VerboWeb, is the case of reciprocal verbs. Reciprocity is considered to be a grammatically relevant semantic property of verbs, since it determines the occurrence of the reciprocal participant as a single plural argument or as two distinct arguments in syntax (GODOY, 2008). The examples in (15) and (16) are from VerboWeb; in (a) the reciprocal participant is a single plural argument whereas in (b) it appears as two distinct arguments in syntax:

(15) a) O **casal** brigava sempre.

the couple argued always

'The couple argued all the time.'

b) O **marido** sempre brigava com a **mulher**.

the husband always argued with the wife

'The husband argued with his wife all the time.'

(16) a) O técnico conectou **os computadores**.

the repairman connected the computers

'The repairman connected the computers.'

b) O técnico conectou **um computador** ao **outro**.

the repairman connected one computer to the other

'The repairman connected one computer to the other.'

Reciprocity crosscuts other relevant semantic properties. As can be seen from the examples (15) and (16), reciprocity can be a property of the Agent, as the case of *brigar* 'argue' (two individuals are intentionally involved in a reciprocal action), or it can be a property of the Patient, as the case

<sup>3</sup> CROVALLEX does not provide English translations for the verbs. The translations in example (14) are from Google Translator, based on the definition for each entry given in CROVALLEX.

of *conectar* 'connect' (two objects are reciprocally affected by an action). Thus, although reciprocity is relevant for syntax, it is subordinated to other more general properties, such as intentionality and change of state. For that, Cançado and Gonçalves (2016) propose that this property should be considered in a fine-grained classification, the level in which one can find subclasses in VerboWeb's analysis. In VerboWeb's analysis, the verb *brigar* belongs to the class of internally caused verbs, and to the subclass of reciprocal verbs with reciprocity in the subject; and *conectar* belongs to the optionally agentive change of state class, and to the subclass of reciprocal verbs with reciprocity in the object.

Although each resource presented here has its own classes, subclasses and properties, they share the assumptions presented in this section, following Levin (1993). And, as demonstrated with the Basque example in (13), they also provide representations for both syntactic and semantic properties of verbs. One can also find in these resources the syntactic-semantic properties that define each class, as well as several examples of verbs, classes, and subclasses for the object language. As Levin (1993, p. 10) argues, "the existence of ties between verb behavior and verb meaning is not particular to English". For this reason, a great deal of research has been developed in order to analyze such ties in many different languages. In the next section, we will present specifically the analyses for English, Basque, Spanish, Catalan, Brazilian Portuguese, Croatian, and Czech, which assume the same perspective as Levin (1993).

### 3 Online resources for the syntactic-semantic classification of verbs

This section is dedicated to providing a general overview of the online open access resources for the syntactic-semantic classification of verbs. These are VerbNet, BVI, AnCoraVerb, VerboWeb, CROVALLEX, and VALLEX, which present an analysis of the verbal lexicons of 7 different languages: English, Basque, Spanish and Catalan, Brazilian Portuguese, Croatian, and Czech, respectively.

#### 3.1 VerbNet (English)

VerbNet (SCHULER, 2005; KIPPER et al., 2008) is the most complete and developed online resource for the syntactic-semantic classification of verbs, being a great inspiration for those which came next. It "provides detailed syntactic semantic descriptions of Levin classes" for the English language (KIPPER et al., 2008, p. 21). The current version (3.4) presents 4570 verbs in 6791 senses, 329 classes, and 272 subclasses. It also focuses on NLP applications, although its vast quantity of data, together with thorough theoretical analyses, are easily accessible for linguists. It can be accessed at <https://uvi.colorado.edu/> (accessed on: 24 Nov. 2021) together with other important lexical resources, such as FrameNet and PropBank (which are not focused on verbal classification). VerbNet's analysis draws heavily on Levin (1993), but brings several new theoretical analyses, representations, and classes (KIPPER et al., 2008).

VerbNet provides an entry for each verb class. In the entry for a class, one can find the set of members of the class, the thematic roles for the arguments of the verbs, as well as selectional restrictions on the arguments, and a representation of the syntactic argument realization patterns of the class. VerbNet's analyses, following Levin (1993), incorporates different degrees of granularity, so classes can be further subdivided in distinct subclasses. Classes are labeled by the name of a representative verb, and numbers are used to signal the level of granularity. The class entry provides a sentence example of a representative verb for each argument realization pattern, as well as a semantic description of the event structure in terms of predicate decomposition. Specific entries for individual verbs provide information about specific semantic components of that verb (not shared by other class members) and also links to other resources.

As Levin and Rappaport Hovav (2005) explain, predicate decomposition is a representation of the meaning of verbs, which are taken to be decomposable semantic expressions. In this perspective, it is assumed that the meaning of verbs is built from a group of smaller semantic components, traditionally considered to be primitives, which can be represented in a predicate-argument structure using English verbs as a form of metalanguage. In VerbNet, predicates such as DO, CAUSE, BE, and HAVE are used to describe the semantic components of the meaning of verbs in a specific class. *Run* verbs, for

example, are described by the predicate MOTION, which is defined as “an agent or a theme in motion”. This representation is more specific than the one given in (12), as it includes the motion component of the verbal semantics. As an exemplification, we show in Table 1 the information provided in VerbNet for the class of *run* verbs, and in Table 2 the entry for the verb *crawl* (which belongs to the *run* class).

**Table 1.** Entry for the class of *run* verbs in VerbNet.

|                                    |   |
|------------------------------------|---|
| Class                              | run-51.3.2  |
| Member Verb Lemmas (class members) | <i>crawl, jump, roar, step, stomp, troop...</i> <sup>4</sup>  |
| Syntactic structure                | <p>NP V<br/>Example: The horse ran.</p> <p>NP V PP.location<br/>Example: The horse ran to the barn.</p> <p>NP V PP.location<br/>Example: The horse jumped over the fence.</p> <p>There V PP NP<br/>Example: There jumped out of the box a little white rabbit.</p> <p>There V PP NP<br/>Example: There jumped a little white rabbit out of the box.</p> <p>PP.location V NP<br/>Example: Out of the box jumped a little white rabbit.</p> |
| Semantic structure <sup>5</sup>    | <p>Roles/Selectional restrictions:<br/>Theme [+animate +machine]<br/>Trajectory [+concrete]<br/>Initial_Location [+concrete]<br/>Destination [+concrete]</p> <p>Semantics:<br/>HAS_LOCATION (e1<sup>6</sup>, Theme, ?Initial_Location)<br/>MOTION (ë2, Theme, Trajectory)<br/>¬ HAS_LOCATION (ë2, Theme, ?Initial_Location)<br/>HAS_LOCATION (e3, Theme, ?Destination)</p>  |
| Subclasses                         | <p>run-51.3.2-1<br/>run-51.3.2-2<br/>run-51.3.2-2-1</p>   |

Source: VerbNet (<https://uvi.colorado.edu/verbnet/run-51.3.2>, accessed on: 24 Nov. 2021).

Other information provided in class entries are dependence parse trees, for the syntactic structures, and force dynamic representations, for the event structure. Force dynamic representations of the event structure are based on the work of Croft (2012), Croft and Vigus (2017), and Kalm, Regan, and Croft (2019). Lexical aspect is not explicitly stated, but it can be derived from semantic representations. According to Kipper et al. (2008), the predicates used in the semantic representation have temporal definitions, so they also provide aspectual information for the classes. MOTION, for instance, describes activities. According to Kipper et al. (2008, p. 25),

<sup>4</sup> The complete list can be found at <https://uvi.colorado.edu/verbnet/run-51.3.2> (accessed on: 25 Nov. 2021).

<sup>5</sup> The Semantic Structure in Table 1 illustrates the semantic representation of the syntactic pattern NP V PP.location. Different semantic representations are provided for the other patterns of argument realization.

<sup>6</sup> In VerbNet’s semantic representation for the events (Generative Lexicon event structures), “e” is the event variable; “e” indicates stative situations and “ë” indicates processes (dynamic situations). A number accompanying the event variable indicates the phase of the event, in the sense that e1 precedes e2, which precedes e3, and so on. More detail about such representations can be found in Brown, Pustejovsky, et al. (2018) and Brown, Bonn, et al. (2019).



**Table 2.** Entry for the verb *crawl* in VerbNet.

| Verb                    | <i>crawl</i>                   |
|-------------------------|--------------------------------|
|                         | +speed                         |
| Verb features           | +bodily_manner<br>+orientation |
| WordNet correspondence  | crawl%2:38:00                  |
| FrameNet correspondence | Self_motion                    |

Source: VerbNet (<https://uvi.colorado.edu/verbnet/run-51.3.2>, accessed on: 24 Nov. 2021).

[e]ach predicate is associated with an event variable E that allows predicates to specify when in the event the predicate is true (start(E) for the preparatory stage, during(E) for the culmination stage, and end(E) for the consequent stage). Aspect is captured by this event variable argument present in the predicates. For example, verbs that denote activities or processes (e.g., motion verbs), have predicates referring to the during(E) stage of the event.

The database also provides theoretical explanations and definitions for all the properties listed in the entry of a verb or verb class. The search in VerbNet's database can be done using different parameters. One can search by verb, class (given that they are familiar with the class labels) and properties of the classes. A list of the classes is also provided. However, there is not a complete list of verbs.

### 3.2 BVI (Basque)

The Basque Verb Index (ESTARRONA et al., 2015) follows VerbNet's model and it is also focused on NLP applications. It can be accessed at <http://ixa2.si.ehu.eus/e-rollda/> (accessed on: 24 Nov. 2021), together with other resources, using the e-ROLda tool. It also links information to VerbNet, as well as to PropBank, FrameNet and WordNet. BVI provides "information about the syntactic and semantic structure of verbs as well as examples of use" for the Basque language, according to the introduction on the e-ROLda website.

The database has 1211 verb entries and provides a complete list of verbs (some verb entries include more than one sense for the same verb). Despite following VerbNet's model, BVI does not provide explicit information about verb classes, nor class entries, although classes can be derived from the information provided in the database.

Information for each verb entry comprises a translation in English (which is also a FrameNet, VerbNet, and/or PropBank correspondent) and a syntactic-semantic representation in the form of a thematic grid, with information on the number of verb arguments, the thematic roles of such arguments and also syntactic functions of arguments, as well as information about case marking (as in example (13)). Example sentences from corpora are also provided in each verb entry. In some entries, there is information on selectional restrictions for animate/inanimate, human/non-human, concrete/non-concrete. Moreover, in some of the entries it is also possible to get access to the Multilingual Central Repository, with the translation of the Basque verb into English, Catalan, Portuguese, Spanish, and Galician. Of all the resources presented here, BVI is the only one which offers a multilingual translation and comparison. The entry also specifies if the analysis was manual or automatic. More complex entries also provide information about different senses of the verb and different argument realization patterns. As an exemplification, Table 3 presents the information for the entry of the Basque verb *defenditu*. The analysis presented in the entry in Table 3 was carried out automatically.

As shown in (13), the syntactic-semantic information in BVI displays thematic roles, syntactic functions and case morphology for the verb's arguments. Arg0 is the subject, arg1 is the object and arg2 is the oblique. Erg, abs, and abl are short for ergative, absolutive, and ablative case marking, respectively. The thematic roles associated with those functions are Agent, Theme, and Locative

**Table 3.** Entry for *defenditu* in BVI.

| Verb   | <i>defenditu</i>  |
|--|---|
| FrameNet-PropBank<br>correspondence/English<br>translation | protect   |
| Syntactic-semantic information                             | arg0: agent, - (erg)<br>arg1: theme, - (abs)<br>arg2: loc, - (abl)  |
| Examples from corpora<br>(EPEC-RoSem corpus)               | Bestetik, Roldanen abokatuek guztiz gezurtatu dute atzo beren defendituak Felipe Gonzalez eta Narcis Serra espiatzeko agindu zuenik. <sup>7</sup> |

Source: BVI ([http://ixa2.si.ehu.eus/e-rolda/galdera.php?adi\\_eu=defenditu](http://ixa2.si.ehu.eus/e-rolda/galdera.php?adi_eu=defenditu); accessed o 24/11/2021).

(loc).

The search in the database can be done by verb, by sense, by English equivalent, by the number of arguments required, and by thematic structure. However, there is no class list or class search. A highlight of this resource is that English translations of the database structure, as well as translations of the data, make it easily accessible for those who are not fluent in Basque, but have interest in the Basque data. The database also provides a list of Basque verbs, with links to verb entries. The search in BVI, however, is not easy for the researcher who is not familiar with the database annotation. BVI is a corpus-based lexicon, which means that the list of verbs comes from a semiautomatic search in large Basque annotated corpora.

### 3.3 AnCoraVerb (Spanish and Catalan)

AnCoraVerb\_ES (for Spanish) and AnCoraVerb\_CA (for Catalan) (APARICIO; TAULÉ; MARTÍ, 2008; TAULÉ; MARTÍ; BORREGA, 2010), as the BVI, are corpus-based lexicons. However, AnCoraVerb lexicons are not focused on NLP applications, and do not present automatic analyses. “AnCora-Verb lexicons focus on syntactic functions, arguments and thematic roles of each verbal predicate taking into account the verbal semantic class and those alternations in diathesis where the predicate can participate” (APARICIO; TAULÉ; MARTÍ, 2008, p. 261). Both lexicons can be accessed at <http://clic.ub.edu/corpus/lexicons/> (accessed on: 24 Nov. 2021), together with the AnCora corpus. AnCoraVerb\_ES presents 2647 Spanish verbs, and AnCoraVerb\_CA presents 2143 Catalan verbs. A verb with more than one meaning may belong to more than one class.

The theoretical basis of the classification in AnCoraVerb follows Levin’s (1993) initial assumptions, as described in Section 2. For the representation of syntactic-semantic information, Rappaport Hovav and Levin’s (1998) predicate decomposition model is adopted. The semantic information is provided in the form of a Lexical Semantic Structure (LSS), built with the predicate decomposition metalanguage. The LSS determines argument structure, including alternations, and also provides information on aspect and thematic roles. In Rappaport Hovav and Levin’s (1998) representations, an LSS comprises three types of components: primitive predicates (such as CAUSE, BECOME and ACT), the root (or constant), which has an ontological category and specifies the meanings of individual verbs, and variables, which represent the arguments of verbs. Each class has its own LSS, however class entries are not provided in the database. The LSS in (17) represents the event structure for the transitive-causative class of verbs (APARICIO; TAULÉ; MARTÍ, 2008, p. 265):

(17) [x CAUSE [BECOME [y <STATE >]]]

Verbs in this class, such as *romper* ‘break’, *abrir* ‘open’ and *cerrar* ‘close’, have two arguments, a Cause argument (x), and a Patient argument (y) (called a Theme in AnCoraVerb’s analysis). They

<sup>7</sup> More examples are provided in [http://ixa2.si.ehu.eus/e-rolda/galdera.php?adi\\_eu=defenditu](http://ixa2.si.ehu.eus/e-rolda/galdera.php?adi_eu=defenditu) (accessed on: 24 Nov. 2021). The examples do not have English translations.

are also defined as accomplishment verbs. In the database, the structure in (17) is not given, but it is indicated with a letter-number code: A11. The letter indicates the aspectual type of the class: accomplishments (A), achievements (B), states (C) and activities (D), and the number indicates the form of the predicate decomposition structure (as well as the granularity of the classification). The structure in (17) is LSS A11, and a structure for an achievement verb, such as [BECOME [*PLACE*]], is LSS B21.

Table 4 presents an example with the verb *abismar* in Spanish. The entry of a specific verb includes its coded LSS, information about different senses associated to the same verb, the syntactic functions and thematic roles of the verb's arguments, the alternations in which the verb occurs, and also example sentences. Verbal entries also provide links to several other resources, such as PropBank and VerbNet.

**Table 4.** Entry for *abismar* in AnCoraVerb\_ES.

| Verb                             | <i>abismar</i>  |          |          |       |      |      |     |    |      |     |    |      |     |
|----------------------------------|---|----------|----------|-------|------|------|-----|----|------|-----|----|------|-----|
| PropBank correspondent (sense 1) | immerse.01  |          |          |       |      |      |     |    |      |     |    |      |     |
| VerbNet class                    | put-9.1   |          |          |       |      |      |     |    |      |     |    |      |     |
| Syntactic-semantic information   | <u>lss: A11.transitive-causative</u>  |          |          |       |      |      |     |    |      |     |    |      |     |
|                                  | <table border="1"> <thead> <tr> <th>Function</th> <th>Argument</th> <th>Theme</th> </tr> </thead> <tbody> <tr> <td>subj</td> <td>arg0</td> <td>cau</td> </tr> <tr> <td>cd</td> <td>arg1</td> <td>tem</td> </tr> <tr> <td>cc</td> <td>argM</td> <td>loc</td> </tr> </tbody> </table> | Function | Argument | Theme | subj | arg0 | cau | cd | arg1 | tem | cc | argM | loc |
| Function                         | Argument  | Theme    |          |       |      |      |     |    |      |     |    |      |     |
| subj                             | arg0  | cau      |          |       |      |      |     |    |      |     |    |      |     |
| cd                               | arg1  | tem      |          |       |      |      |     |    |      |     |    |      |     |
| cc                               | argM  | loc      |          |       |      |      |     |    |      |     |    |      |     |
| Example                          | -   |          |          |       |      |      |     |    |      |     |    |      |     |
| PropBank correspondent (sense 2) | immerse.01  |          |          |       |      |      |     |    |      |     |    |      |     |
| VerbNet class                    | put-9.1   |          |          |       |      |      |     |    |      |     |    |      |     |
| Syntactic-semantic information   | <u>lss: B21.unaccusative-state</u>  |          |          |       |      |      |     |    |      |     |    |      |     |
|                                  | <table border="1"> <thead> <tr> <th>Function</th> <th>Argument</th> <th>Theme</th> </tr> </thead> <tbody> <tr> <td>subj</td> <td>arg1</td> <td>tem</td> </tr> <tr> <td>cd</td> <td>argM</td> <td>loc</td> </tr> </tbody> </table>   | Function | Argument | Theme | subj | arg1 | tem | cd | argM | loc |    |      |     |
| Function                         | Argument  | Theme    |          |       |      |      |     |    |      |     |    |      |     |
| subj                             | arg1  | tem      |          |       |      |      |     |    |      |     |    |      |     |
| cd                               | argM  | loc      |          |       |      |      |     |    |      |     |    |      |     |
| Example                          | muchos otros que se han abismado definitivamente en el caos y el drama. <sup>8</sup>  |          |          |       |      |      |     |    |      |     |    |      |     |

Source: AnCoraVerb ([http://clic.ub.edu/mbertran/ancora/lexentry.php?file=abismar.lex.xml&lexicon=AnCoraVerb\\_ES](http://clic.ub.edu/mbertran/ancora/lexentry.php?file=abismar.lex.xml&lexicon=AnCoraVerb_ES); accessed on: 24 Nov. 2021).

As in BVI, arg0 is the subject (subj), arg1 is the object (cd) and arg2 is the oblique (cc). Cau, tem and loc are short for the thematic roles Cause, Theme and Locative. ArgM stands for adjunct; adjuncts are indicated in some cases, along with verbal arguments.

AnCoraVerb is available in English, and English translations for verbs are provided by means of PropBank correspondents. Both AnCoraVerb\_ES and AnCoraVerb\_CA provide the complete list of verbs. Specific search tools include search by syntactic structure, thematic roles and verb. There is no search by class.

### 3.4 VerboWeb (Brazilian Portuguese)

VerboWeb (CANÇADO; AMARAL; MEIRELLES, 2017) can be accessed at [www.letras.ufmg.br/verboweb](http://www.letras.ufmg.br/verboweb) (accessed on: 10 Mar. 2022). As VerbNet, VerboWeb is not corpus-based. Most of VerboWeb's verbs come from Borba's (1990) dictionary of Brazilian Portuguese verbs. And differently from VerbNet and BVI, but similarly to AnCoraVerb, it is not focused on NLP applications. Although the authors do recognize its possible applications in NLP, VerboWeb comes as the result of theoretical

<sup>8</sup> AnCoraVerb does not provide English translations for examples.

lexical-semantic work, with the main goal of providing a representation of the speakers' knowledge about the verbal lexicon, in line with Levin's (1993) original goals. VerboWeb is also the only one of those resources which specifies a variant within a language, presenting a syntactic-semantic classification of the Brazilian Portuguese verbal lexicon. The database currently presents 1571 verbal entries (including distinct entries for polysemous verbs), 19 classes, and 8 subclasses.

As AnCoraVerb, VerboWeb uses Rappaport Hovav and Levin's (1998) predicate decomposition model. Based on previous studies for Brazilian Portuguese, VerboWeb develops the authors' meta-language for predicate decomposition and presents an event structure representation for each class. Classes are also defined by four other main properties: the common semantic content in the class, basic syntactic structure, thematic role structure, and basic lexical aspect. Each class is also determined by a group of argument realization properties. The name of a class is given considering the aspectual classification and the event structure properties of the verbs; they are crafted in order to briefly explain the semantic content of the class. All this information is provided in the class entry, together with explanations and bibliographical references for theoretical concepts.

**Table 5.** Entry for the class of Activity: internally caused verbs (unergative) in VerboWeb.

| Class                        | Activity: internally caused verbs (unergative)   |
|------------------------------|--|
| Class properties             | <ul style="list-style-type: none"> <li>- Common semantic content in the class: do/make an event</li> <li>- Basic syntactic form: [NP V] (intransitive verb)</li> <li>- Thematic role structure: Agent</li> <li>- Predicate decomposition structure: [X DO &lt;EVENT&gt;]</li> <li>- Basic lexical aspect: activity</li> <li>- Licenses cognate object</li> <li>- Licenses an adjunct equivalent to the cognate object</li> </ul> |
| Verbs belonging to the class | <p>Acenar ~ Wave<br/>           Andar ~ Walk<br/>           Apitar 2 ~ Whistle<br/>           Arrostar ~ Burp<br/>           Assobiar/Assoviar ~ Whistle<br/>           Balbuciar ~ Babble<br/>           Berrar ~ Shout<br/>           Bocejar ~ Yawn<br/>           Bradar ~ Scream<br/>           Brigar ~ Fight...<sup>9</sup></p>   |

Source: VerboWeb ([http://www.lettras.ufmg.br/sistemas/verboweb\\_cliente/ver\\_class\\_e.php?id=14#](http://www.lettras.ufmg.br/sistemas/verboweb_cliente/ver_class_e.php?id=14#); accessed on: 24 Nov. 2021).

VerboWeb also provides information on subclasses and non-classificatory (or multiclass) properties. Subclasses are a subgroup of verbs within a class, which have a specific property derived from a specific semantic content, as the case of reciprocal verbs exemplified in Section 2. Subclasses have more properties than classes, so the subclass of reciprocal verbs (reciprocity in subject) has all the properties of the class to which it belongs (activity: internally caused verbs (unergative)), as in Table 5, plus the properties shown in Table 6.

Non-classificatory (or multiclass) properties are syntactic constructions which crosscut different classes, such as reflexivization, middle voice, and other properties driven by non-lexical factors. The database provides entries for non-classificatory properties, which also contain explanations and bibliographical references for theoretical concepts, as well as the list of verbs which accept the property.

Finally, the entry for an individual verb presents an example of the basic use of the verb, its

<sup>9</sup> The complete list of verbs in this class can be found in [http://www.lettras.ufmg.br/sistemas/verboweb\\_cliente/ver\\_class\\_e.php?id=14](http://www.lettras.ufmg.br/sistemas/verboweb_cliente/ver_class_e.php?id=14)(accessed on: 24 Nov. 2021).

**Table 6.** Entry for the subclass of Reciprocal verbs (reciprocity in subject) in VerboWeb.

|                                 |   |
|---------------------------------|---|
| Subclass                        | Reciprocal verbs (reciprocity in subject)   |
| Subclass properties             | - The subject argument has a plural denotation<br>- Licenses the discontinuous form (argument factoring)  |
| Verbs belonging to the subclass | Brigar ~ Fight<br>Contracenar ~ Act (with other actors)<br>Conversar ~ Talk<br>Dialogar ~ Dialogue<br>Duelar ~ Duel<br>Flertar ~ Flirt<br>Lutar ~ Fight<br>Prosear ~ Chat |

Source: VerboWeb ([http://www.lettras.ufmg.br/sistemas/verboweb\\_cliente/ver\\_subclasse.php?id=30#](http://www.lettras.ufmg.br/sistemas/verboweb_cliente/ver_subclasse.php?id=30#); accessed on: 24 Nov. 2021).

class and class properties, its subclass and subclass properties, as well as examples for each argument realization pattern for the specific verb. An example of non-classificatory properties is also provided, if the verb occurs in some type of non-classificatory construction. Table 7 illustrates a verb entry in VerboWeb, with the verb *brigar*.

**Table 7.** Entry for *brigar* in VerboWeb.

|                                     |   |
|-------------------------------------|---|
| Verb                                | <i>brigar</i>   |
| English translation                 | fight   |
| Example                             | <i>O casal brigava sempre.</i><br>'The couple argued all the time.'   |
| Class                               | Activity: internally caused verbs (unergative)  |
| Class properties (with examples)    | - Common semantic content in the class: do/make an event<br>- Basic syntactic form: [NP V] (intransitive verb)<br>- Thematic roles structure: Agent<br>- Predicate decomposition structure: [X DO <EVENT>]<br>- Basic lexical aspect: activity<br>- Licenses cognate object: <i>O casal brigou uma briga feia.</i><br>- Licenses an adjunct equivalent to the cognate object: <i>O casal brigou feio.</i> <sup>10</sup> |
| Subclass                            | Reciprocal verbs (reciprocity in subject)   |
| Subclass properties (with examples) | - The subject argument has a plural denotation<br>- Licenses the discontinuous form (argument factoring): <i>O marido sempre brigava com a mulher.</i>  |

Source: VerboWeb ([http://www.lettras.ufmg.br/sistemas/verboweb\\_cliente/ver\\_verbo.php?id=1314](http://www.lettras.ufmg.br/sistemas/verboweb_cliente/ver_verbo.php?id=1314); accessed on: 24 Nov. 2021).

The database provides translations in English for most of the data and the interface, which makes it accessible for those who are not speakers of Portuguese. However, English translations are still in progress and many verbs, as well as explanations for classes and properties, are still only available in Portuguese.

VerboWeb offers a very elaborate search engine. Simple search in the database can be done by

<sup>10</sup> Examples of properties do not have English translations in VerboWeb.

verb, by English equivalent, by class, by subclass, and by non-classificatory property. The advanced search allows one to search for specific properties of verbs, such as aspect, thematic role structure, and specific argument alternations. An alphabetical list of verbs is provided, and it is also possible to access lists of classes, subclasses, class properties, subclass properties, and also non-classificatory properties.

### 3.5 CROVALLEX (Croatian)

The Croatian Verb Valency Lexicon (CROVALLEX) (PRERADOVIC; BORAS; KISICEK, 2009; PRERADOVIC, 2010) can be accessed at <http://theta.ffzg.hr/crovallex/> (accessed on: 24 Nov. 2021). CROVALLEX contains 1739 verbs in 72 classes and their subclasses (based on VerbNet classes) and aims at providing a “high-quality machine-readable lexicon of Croatian verbs” (PRERADOVIC; BORAS; KISICEK, 2009, p. 533). Thus, its focus is on NLP applications, although it is easily accessible also for human readers. Like VerboWeb, this is not a corpus-based lexicon. CROVALLEX picked out its verb entries from the Croatian frequency dictionary of Moguš, Bratanić, and Tadić (1999), and the database presents all the verbs found in the dictionary that have frequency higher than 11. This resource differs from the others in considering the frequency of a verb as an important factor in the analysis.

The theoretical approach adopted in CROVALLEX is the Valency Theory of Tesnière (1959). More specifically, it assumes the Functional Generative Description (FGD) approach of Hajičová, Panevová, and Sgall (2002), cited in Preradovic, Boras, and Kisicek (2009). The fundamental assumptions remain those of Levin (1993), as described in Section 2. Some terminology and representations, however, differ in relation to the other resources presented. The syntactic properties of classes are represented in terms of valence frames, which correspond to the syntactic configuration of the verb’s arguments. Valence frames specify syntactic functions for arguments, including case markers and prepositions, and also specify obligatory and optional arguments, as well as typical modifiers. The semantic properties of classes are represented in terms of functors, or deep roles, which correspond to thematic roles in other approaches. The functors used in CROVALLEX are Agent-AGT, Patient-PAT, Recipient-REC, among others. Although they also present aspectual classifications, aspectual categories differ from the most common Vendlerian classification. In CROVALLEX there are two main aspectual classes, perfective verbs and imperfective verbs (typically described in the literature on Slavic languages); the dual aspect verbs may be perfective or imperfective depending on contextual factors. According to Croft (2012), the perfective/imperfective labels are used to define a distinction between temporally bounded events (perfective) and temporally unbounded events (imperfective).

The database presents entries for verbs. These entries include information about the semantic and syntactic properties of verbs and their classes: valence frames and functors. The verb entry also indicates its class and provides corpus examples for each valence frame. Table 8 shows as an example the entry for the Croatian verb *analizirati (analizirati)* ‘analyze’<sup>11</sup>.

**Table 8.** Entry for *analizirati (analizirati)* in CROVALLEX.

| Verb                                   | <i>analizirati (analizirati)</i>  |
|--|---|
| Aspect                                 | dual  |
| Frame (syntactic-semantic information) | AGT <sub>0_or_1</sub> <sup>obl</sup> PAT <sub>4</sub> <sup>obl</sup> LOC <sub>u+6</sub> <sup>typ</sup><br>TWHEN <sub>tijekom+2</sub> <sup>typ</sup> |
| Example                                | Dubrovački stručnjaci su analizirali brodove, sidrenja i sidrišta u gradskom akvatoriju tijekom prošle godine.                                      |
| Class                                  | Assessment/price  |

Source: CROVALLEX ([http://theta.ffzg.hr/crovallex/data/html/generated/word-entries/analizirati-\(analizirati\).html](http://theta.ffzg.hr/crovallex/data/html/generated/word-entries/analizirati-(analizirati).html); accessed on: 24 Nov. 2021).

<sup>11</sup>Data in CROVALLEX do not have English translations. The translation of *analizirati (analizirati)* is from Google Translator.

In the frame, AGT, PAT and LOC are semantic functors (Agent, Patient, Locative), obl indicates an obligatory argument, and typ signals a non-obligatory typical argument or modifier. The numbers indicate case morphology for the complement (1 - nominative, 2 - genitive, 3 - dative, 4 - accusative, 5 - vocative, 6 - locative, and 7 - instrumental), which can be complemented by additional morphology, such as *u* and *tijekom*. TWHEN represents a temporal-when adverbial expression which typically occurs with the verb.

The database does not have a specific search engine. However, all information is provided in organized lists, in a way that is easy to access. One can find a list of verbs alphabetically organized, a list of all classes, a list of functors, a list of forms (including information about inflexion), a list of aspectual classes, a list of complexity (verbs organized by the quantity of valence frames they can have), and other properties (which comprise derived reflexives, homographs, idiomatic frames, and homonyms). Besides, the website also dedicates a page for the explanation of the theoretical concepts used in the analysis. The database and the theoretical explanations and definitions are in English, but there are no translations for data, either verbs or examples.

### 3.6 VALLEX (Czech)

The last online open access resource for the syntactic-semantic classification of verbs to be presented here is the Valency Lexicon of Czech Verbs (VALLEX) (KETTNEROVÁ; LOPATKOVÁ; BEJČEK, 2012). VALLEX is a corpus-based lexicon, as BVI and AnCoraVerb, and it is also focused on NLP applications, similarly to VerbNet, BVI, and CROVALLEX. VALLEX lists and analyzes 2772 Czech verbs. It is available at <https://ufal.mff.cuni.cz/vallex> (accessed on: 24 Nov. 2021).

As in CROVALLEX, the Functional Generative Description (FGD) is adopted as the main theoretical background. Thus, in VALLEX, classes are also defined using functors and valence frames. Aspectual classes are also imperfective verbs and perfective verbs. VALLEX is divided into two components: the data component and the rule component. Individual verb entries can be found in the data component, while the rule component comprises information about alternations and other grammatical properties of verbs. Table 9 shows an example of verb entry, using the Czech verb *nabádat* 'encourage'<sup>12</sup>.

Table 9. Entry for *nabádat* in VALLEX.

| Verb                                   | <i>nabádat</i>  |
|--|---|
| Aspect                                 | impf  |
| Gloss                                  | vyzývat; pobízet; povzbuzovat; podněcovat; vést k určité činnosti   |
| Frame (syntactic-semantic information) | ACT <sub>1</sub> <sup>obl</sup> ADDR <sub>4</sub> <sup>obl</sup> PAT <sub>k+3</sub> <sup>obl</sup> ,aby,at,že   |
| Example                                | stále nabádala své děti, aby nechodily k rybníku / ať nechodí k rybníku / že se mají pořádně učit   |
| Class                                  | communication   |
| recip                                  | ACT-ADDR vzájemně se nabádali k opatrnosti  |
| reflex                                 | ACT-ADDR Musíš být opatrnější, nabádal jsem sám sebe a ještě jsem byl ze zážitku celý roztřesený.   |
| diat                                   | <u>deagent</u> lidé se stále nabádají, aby nechodili na slunce<br><u>passive</u> Už od mládí byl nabádán strážci královské etikety k vybranému chování. Fotbalisté ale bývají od svých klubů nabádáni, aby rozhovory do televize poskytovali. |

Source: VALLEX (<https://ufal.mff.cuni.cz/vallex/4.0/#/lexeme/nabada1/0>; accessed on: 24 Nov. 2021).

As can be seen in Table 9, VALLEX uses the same type of syntactic-semantic representation as CROVALLEX. Moreover, the VALLEX entry additionally presents a paraphrase of the verb, and more information on different types of syntactic structures, such as reciprocal (recip), reflexives (reflex),

<sup>12</sup>Data in VALLEX do not have English translations. The translation of *nabádat* is from Google Translator.

and passives. In Table 9, *impf* stands for imperfective aspect, and *ADDR* indicates the addressee argument of a communication verb. More detailed information is difficult to access in the database, because of the lack of English translations. Only a small part of the database is accessible for readers who are not fluent in Czech, as data and analyses are not translated into English.

In the same way as in CROVALLEX, in VALLEX there are only entries for verbs, but all other information is provided in organized lists. One can find a list of verbs alphabetically organized, a list of classes, a list of functors, a list of valence frames, and a list of alternations, along with other information, such as idiomatic expressions and complex predicates. The database provides also an advanced search engine, but it requires knowledge of the specific properties and annotation used in the database. On the website, a page is dedicated to the explanation of the theoretical concepts used in the analysis (only available in Czech).

#### 4 An example of theoretical research using online resources: the case of verbs of creation

The present section compares four verbs of creation in the 7 languages described in the online resources presented in Section 3: English, Basque, Spanish, Catalan, Brazilian Portuguese, Croatian, and Czech. The compared verbs are *write*, *paint*, *build*, *create* and their translations in the chosen languages. As the verb may have multiple entries in the databases, the most basic one was considered (the one signaled as default or marked as 1). The goal of this comparative analysis is to provide an example of how the online open access resources presented here can be used in linguistic research. As these resources are mostly focused on NLP applications, the main aim of this section is to highlight the value of the information they provide for Theoretical Linguistics.

Although the relation between event structure and argument realization holds crosslinguistically, as Levin (1993) explains, different languages do not have the same classes or the same patterns of lexicalization and argument realization. Amaral and Cançado (2020), for example, show a comparison between the verbs *drive* and *dirigir* 'drive' in English and Portuguese, arguing that, although translatable, these verbs lexicalize different event structures, occurring also in different argument structure patterns.

(18) The Porsche drove away/around/into the garage.

(19) \*O Porsche dirigiu por aí/ para dentro da garagem.

the Porsche drove for there/ to inside of.the garage.

(AMARAL; CANÇADO, 2020, p. 25)

The authors show that the structure in (18), a form of *drive* with the vehicle argument in subject position, can occur in English, but not in Portuguese, as (19) demonstrates. They argue that the different patterns in (18) and (19) arise because, despite describing the same situation in the world, *drive* lexicalizes the event structure of a motion along a path verb (similar to *throw* verbs), while *dirigir* lexicalizes the event structure of a manner of motion verb (similar to *run* verbs).

Following this line of thought, the resources presented here, other than VerbNet, should not be seen as a simple translation of Levin (1993) or of VerbNet itself, but they consist of original studies which investigate the specific classes of a language. Preradovic, Boras, and Kisicek (2009) and Kettnerová, Lopatková, and Bejček (2012) explicitly state this point in their analyses. Although CROVALLEX and VALLEX are based on VerbNet's classes, classes were adapted and reanalyzed for the classification of Croatian and Czech verbs, respectively. Cançado, Godoy, and Amaral (2013) also explain that their work, although inspired in Levin (1993), cannot be seen as a translation, and original classes based on specific properties of Brazilian Portuguese are provided in their analysis.

Hence, the comparative analysis presented here assumes that verb classes may differ from language to language, despite maintaining the relation between event structure and argument realization. The choice of the four verbs *write*, *paint*, *build* and *create* aims at representing the two main classes of verbs of creation, defined in Levin (1993): verbs of image creation (*write*, *paint*) and verbs of creation and transformation (*build*, *create*). As we will see, classes differ in the 7 languages analyzed, although they do have many similarities. For English, Levin (1993) proposes that each one of these classes is further subdivided into different subclasses. The same classification for English is also assumed in



VerbNet:

(20) Class: Image Creation Verbs

Subclasses: Verbs of Image Impression; Scribble Verbs; Illustrate Verbs; Transcribe Verbs

(21) Class: Verbs of Creation and Transformation

Subclasses: Build Verbs; Grow Verbs; Verbs of Preparing; Create Verbs; Knead Verbs; Turn Verbs; Performance Verbs

Image creation verbs describe events in which an image is created on a surface and they are all two-argument transitive verbs. Subclasses present more specific semantic properties and argument realization patterns. Levin (1993) classifies these verbs according to four properties: the possibility of accepting a PP describing the surface, the image impression alternation, the unspecified object alternation, and the types of nominalizations, exemplified in (22) to (25) for the image creation verb *inscribe* (LEVIN, 1993, p. 169):

(22) Surface PP

a) Smith inscribed his name.

b) Smith inscribed his name on the ring/ over the door/under the picture.

(23) Image Impression Alternation

a) Smith inscribed his name on the ring. (locative variant)

b) Smith inscribed the ring with his name. (with variant)

(24) Unspecified Object Alternation

a) Smith was inscribing the rings.

b) Smith was inscribing.

(25) Types of nominalizations

a) Process Nominal: the inscription of the motto

b) Result Nominal: the inscription (on the wall)

The subclasses of image creation verbs, thus, have the configuration presented in Table 10. *Write* belongs to the subclass of *scribble* verbs, and *paint* to the subclass of image impression verbs. The image impression alternation is the main difference between these subclasses. It is available for *paint*, but not for *write*.

Verbs of Creation and Transformation are two-argument transitive verbs which describe events where an agent's action brings an object into existence. The exception is the *grow* subclass, composed of intransitive verbs which describe the coming into existence of an object, without external causation. Subclasses specify different ways in which an object can come into existence, and also differ in at least five properties: the material/product alternation, the total transformation alternation, the causative alternation, the raw material alternation, and the benefactive alternation (exemplified in (26) to (30) with the verbs *carve* and *turn*).

(26) Material/product alternation

a) Martha carved a toy out of a piece of wood.

b) Martha carved the piece of wood into a toy.

(27) Total Transformation Alternation

a) The witch turned him into a frog.

b) The witch turned him from a prince into a frog.

(28) Causative Alternation

a) The witch turned him into a frog.

b) He turned into a frog.

(29) Raw Material Alternation

a) Martha carved beautiful toys out of this piece of wood.

b) This wood carves beautiful toys

(30) Benefactive Alternation

a) Martha carved a toy for the baby.

b) Martha carved the baby a toy.

(LEVIN, 1993, p. 173, 177)

The subclasses of verbs of creation and transformation, therefore, have the configuration presented

Table 10. Properties of subclasses of image creation verbs.

| Image creation verbs   |            |                              |                                |     |  |  |
|--|------------|------------------------------|--------------------------------|-----|--|--|
| Subclasses   | Properties |                              |                                |     |  |  |
|  | Surface PP | Image Impression Alternation | Unspecified Object Alternation | Ob- | Types of nominalizations   |  |
| Verbs of Image Impression ( <i>embroider, mark, tattoo</i> ) | Yes        | Yes                          | Yes                            |     | Process Nominal; Result Nominal  |  |
| Scribble Verbs ( <i>carve, doodle, scribble</i> )            | Yes        | No                           | Some verbs                     |     | Zero-related nominal   |  |
| Illustrate Verbs ( <i>adorn, illustrate, decorate</i> )      | No         | No                           | No                             |     | Process Nominal (some verbs); Result Nominal (some verbs); Zero-related nominal (some verbs) |  |
| Transcribe Verbs ( <i>copy, film, transcribe</i> )           | No         | No                           | No                             |     | Zero-related nominal   |  |

Source: Levin (1993).

in Table 11. *Build* belongs to the subclass of *build* verbs, and *create* belongs to the subclass of *create* verbs. The main differences between these subclasses are the material/product alternation and the raw material alternation, available for *build*, but not for *create*.

Levin's (1993) analysis shows that verbs of creation are an extremely complex and diverse group in English. Thus, for the comparison of this type of verbs in different languages, we hypothesized the existence of the two classes of verbs of creation proposed in Levin (1993), and analyzed two members of each of those classes and their translations into Basque, Spanish, Catalan, Brazilian Portuguese, Croatian and Czech.<sup>13</sup> We investigated in the object languages if there are the two distinct types of verbs of creation: verbs of image creation and verbs of creation and transformation. We also checked if they differ in the patterns of lexicalization and argument structure. Starting from English data, *write* and *paint* describe the creation of an image on a surface; they are two-argument transitive verbs, but accept a third PP argument which describes the surface. These verbs differ in the possibility of accepting the image impression alternation. *Build* and *create* are two-argument transitive verbs which describe the action carried out by an agent resulting in the creation of an object. These verbs differ in the possibility of accepting the material/product alternation and the raw material alternation.

In Basque, as the data in BVI show, there are three types of verbs of creation within the four verbs under scrutiny, differently from the English bipartite classification. *Idatzi* 'write' takes three arguments, an Agent, a Theme, and a Destination. The analysis shows that *idatzi* describes not only an image creation event, but an image creation with the purpose of communicating a message. In Levin's (1993) analysis of English, this property is also described for *write*, as the verb is also listed in the class of verbs of transfer of a message.

- (31) *Idatzi* 'write'  
 arg0: agent (erg)  
 arg1: theme (abs)

<sup>13</sup> Translations for these verbs were taken from the original databases in BVI, VerboWeb and AnCoraVerb; Google Translator was used to translate verbs from Croatian and Czech.

**Table 11.** Properties of subclasses of verbs of creation and transformation.

| Verbs of creation and transformation                 |                                |                                  |                       |                          |                         |
|--|--------------------------------|----------------------------------|-----------------------|--------------------------|-------------------------|
| Subclasses   | Properties                     |                                  |                       |                          |                         |
|  | Material / product alternation | Total Transformation Alternation | Causative Alternation | Raw Material Alternation | Benefactive Alternation |
| Build Verbs ( <i>build, carve, crochet</i> )         | Yes                            | No                               | No                    | Yes                      | Yes                     |
| Grow Verbs ( <i>evolve, grow, mature</i> )           | Yes (intransitive)             | No                               | Yes                   | No                       | No                      |
| Verbs of Preparing ( <i>boil, prepare, roast</i> )   | No                             | No                               | No                    | No                       | Yes                     |
| Create Verbs ( <i>coin, compose, create</i> )        | No                             | No                               | No                    | No                       | No                      |
| Knead Verbs ( <i>blend, knead, squash</i> )          | No                             | No                               | Yes (some verbs)      | No                       | No                      |
| Turn Verbs ( <i>alter, change, turn</i> )            | No                             | Yes                              | Yes                   | Yes                      | No                      |
| Performance Verbs ( <i>chant, compose, perform</i> ) | No                             | No                               | No                    | No                       | Yes (some verbs)        |

Source: Levin (1993).

arg2: destination (dat)

The Basque verb *margotu* 'paint' has a single Theme absolutive argument (the object painted). As this verb is analyzed automatically in BVI, and the corpus examples presented are translatable into passive structures in English, we believe that it could be in fact a two-argument transitive verb. In any case, it differs substantially from the structure in (31) for *idatzi* 'write'.

(32) *Margotu* 'paint'

arg1: theme, - (abs)

This example indicates a setback in automatic corpus analyses, since the verb may be analyzed as intransitive if the corpus only includes its passive use.

The Basque verbs *eraiki* 'build' and *sortu* 'create' are analyzed in BVI as three-argument verbs, bearing Agent, Product and Material roles. The syntactic expression of these arguments is the same for both verbs, as the case morphology indicates in the structure. The only distinction is that *sortu* 'create' also has a Benefactive argument, absent in the representation of *eraiki* 'build'. Interestingly, as shown in Table 11, in English, the benefactive alternation occurs with *build*, but not with *create*.

(33) *Eraiki* 'build' / *sortu* 'create'

arg0: agent, - (erg)

arg1: product, - (abs)

arg2: material, - (ins/soz)

The Material argument in the structure of *eraiki* 'build' and *sortu* 'create' in Basque resembles the material/product alternation, which occurs with *build* in English (recall example (26)).

The data from Spanish and Catalan, differently, seems to motivate a unified class of verbs of creation in those languages. In AnCoraVerb, all four verbs have the same event structure properties, and the same transitive argument structure.

(34) Spanish: *escribir* 'write', *pintar* 'paint', *construir* 'build', *crear* 'create'

Catalan: *escriure* 'write', *pintar* 'paint', *construir* 'build', *crear* 'create'

lss: A21.transitive-agentive-patient

| Function | Argument | Theme |
|----------|----------|-------|
| subj     | arg0     | agt   |
| cd       | arg1     | pat   |

However, these verbs differ in some other properties. Of all four verbs, Spanish *escribir* is the only one which accepts a directional phrase, similar to the Destination argument in Basque *idatzi* 'write' and also to the analysis of *write* as a verb of transfer of a message. This possibility does not appear for *escriure* 'write' in Catalan. In both languages, *construir* 'build' and *crear* 'create' accept a locative phrase, indicating the place of creation; but *pintar* 'paint' does not. English and Basque analyses do not show Locative arguments for verbs of creation and transformation.

In Brazilian Portuguese, *escrever* 'write' and *pintar* 'paint' are three-argument verbs, with Agent, Theme, and Location arguments, similarly to the English verbs of image creation, which can occur with a surface PP argument. They belong to the class of causation verbs: change of locative state. Change of locative state verbs describe a change of state of an entity in relation to a place.

(35) *escrever* 'write' and *pintar* 'paint' (Causation: change of locative state verbs)

- Common semantic content in the class: y comes to be in a specific state in or out of a place
- Basic syntactic structure: [NP V NP (PP)] (bitransitive verb)
- Thematic roles structure: Agent, Patient, (Locative)
- Predicate decomposition structure:  
[[X ACT volition] CAUSE [BECOME [[Y<RESULT-STATE>] LOC Z]]]
- Basic lexical aspect: accomplishment

However, *pintar* 'paint' is placed in the subclass of image creation verbs, while *escrever* 'write' is not. This happens because only *pintar* 'paint' in Portuguese occurs in the image impression alternation, as in the English example in (23). So, in a more fine-grained analysis, these verbs have distinct properties. This analysis is very similar to Levin's (1993), as in English *paint* and *write* are also analyzed as belonging to different subclasses according to the image impression alternation. *Paint* allows the alternation, as *pintar*, and *write* does not, as *escrever*.

*Construir* 'build' and *criar* 'create' are not found in VerboWeb. But in Amaral and Caçado (2014) these verbs are described as two-argument verbs, similar to their translations in Spanish and Catalan and clearly distinct from *escrever* 'write' and *pintar* 'paint'. Amaral and Caçado (2014) observe that *escrever* 'write' and *pintar* 'paint' can occur in intransitive sentences, they occur in the unspecified object alternation, as the English *paint* and *write*, while *construir* 'build' and *criar* 'create' do not.

In Croatian, as the data in CROVALLEX show, *pisati* (*pisati*) 'write' is a two-argument verb, but *slikati* (*slikati*) 'paint' is monovalent. However, both verbs can be used with Locative arguments (in the register class), which indicates that both are similar to English *write* and *paint* in allowing the expression of the surface in which the image is created. *Pisati* (*pisati*) 'write' is also similar to *write*, to Basque *idatzi* 'write', and to Spanish *escribir* 'write', in the sense that it describes not only an image creation, but the transfer of a message.

(36) *pisati* (*pisati*) 'write'

frame: AGT<sub>0\_or\_1</sub><sup>obl</sup> PAT<sub>o+6</sub><sup>obl</sup>

(37) *slikati* (*slikati*) 'paint'

frame: AGT<sub>0\_or\_1</sub><sup>obl</sup>

Differently from Basque *margotu* 'paint', the only argument which appears with *slikati* (*slikati*) 'paint' is the Agent subject, resembling the unspecified object alternation which occurs with English *paint* and Brazilian Portuguese *pintar*.

*Stvarati* (*stvarati*) 'create' and *graditi* (*graditi*) 'build' in Croatian are both two-argument verbs, and they differ from *pisati* (*pisati*) 'write' and *slikati* (*slikati*) 'paint' in not having a Locative argument indicating a surface, and also in the case markings of the Patient argument. They seem to have a similar pattern of lexicalization in comparison to *build* and *create* in the other languages.

(38) *graditi* (*graditi*) 'build' and *stvarati* (*stvarati*) 'create'

frame: AGT<sub>0\_or\_1</sub><sup>obl</sup> PAT<sub>4</sub><sup>obl</sup>

Finally, in VALLEX, the Czech verb *psát*, *psávat* 'write' does not appear with a Locative argument,

but similarly to English, Basque and Spanish, it has an optional Addressee argument, indicating that this verb describes the transmission of a message. In fact, in VALLEX, it is placed in the class of communication verbs.

(39) *psát, psávat* 'write'

frame: ACT<sub>1</sub><sup>obl</sup> ADDR<sub>3</sub><sup>opt</sup> PAT<sub>4, o+6, aby, at, zda, že, cont</sub><sup>obl</sup> DIR<sub>3</sub><sup>typ</sup> MANN<sup>typ</sup> MEANS<sub>7</sub><sup>typ</sup>

*Namalovat* 'paint' also does not appear in VALLEX with the expression of the surface in which the image is created. Differently from *psát, psávat* 'write', however, it is a two-argument verb, similar to *pintar* 'paint' in Spanish and Catalan.

(40) *Namalovat* 'paint'

frame: ACT<sub>1</sub><sup>obl</sup> PAT<sub>4</sub><sup>obl</sup>

*Stavět, stavívat* 'build' and *tvořit, tvořívát* 'create' have four arguments. Besides Agent and Patient, these verbs also describe the origin (optional) of the creation (analogously to the Material in Basque and in English), and also the Beneficiary (typical). They belong to the change class in VALLEX.

(41) *Stavět, stavívat* 'build' and *tvořit, tvořívát* 'create'

frame: ACT<sub>1</sub><sup>obl</sup> PAT<sub>4</sub><sup>obl</sup> ORIG<sub>Z+2</sub><sup>opt</sup> BEN<sub>3, pro+4</sub><sup>typ</sup>

Although it is not our aim here to provide an in-depth analysis of verbs of creation across languages, some important aspects of these verbs have emerged: for instance, it seems that the distinction between image creation and creation and transformation is relevant not only in English, but also in the other languages analyzed. In English, Basque, Spanish, Croatian, and Czech, the verbs equivalent to *write* describe not only an image creation, but an image creation with the purpose of communicating a message. Although this information is not present in VerboWeb and in AnCoraVerb\_CA, we believe that in Brazilian Portuguese and in Catalan this verb may also be used as a communication verb. In this sense, *write* is distinguished from *paint*, although both are classified as image creation verbs. An interesting characteristic of *paint* is that its translations occur as intransitive in English, Basque, Brazilian Portuguese, and Croatian. In Basque and Croatian, the intransitive form of this verb is considered to be its more basic use. In the other languages, it results from the unspecified object alternation. A surface PP appears for *write, paint* and their equivalents in English, Brazilian Portuguese and Croatian. *Build* and *create* and its translations specify a material from which a new object is created in English, Basque, and Croatian. In all languages, these verbs are transitive and distinct from *paint* and *write*.

These few examples of verbs of creation demonstrate that the resources presented in Section 3 provide valuable information about verbs in the analyzed languages, including a great amount of data and theoretical analyses. The brief comparison between verbs of creation illustrates the type of theoretical work that can be done using those resources, but it does not do justice to the full range of possibilities they allow. More complex analyses, as well as comparisons between larger parts of the lexicons, are possible via the use of computational tools.

## 5 Final remarks

The study of verb classes has been central in lexical-semantics literature, since Levin's (1993) foundational work for English. The types of correlations between lexical semantics and sentence form, used to define classes, gave rise to important generalizations about the syntax-semantics interface. One of the most important of those generalizations is the widely accepted idea that event structure and argument realization are strictly tied. Assuming the centrality of verb classes for linguistic analysis, several authors have developed syntactic-semantic classifications of verbs in different languages. The purpose of this paper was to provide an overview of online open access resources for the syntactic-semantic classification of verbs (namely VerbNet, BVI, AnCoraVerb, VerboWeb, CROVALLEX, and VALLEX). We have shown that these resources, following Levin's (1993) model of verb classification, share many theoretical assumptions, specifically regarding which criteria to use in verb classification. Despite adopting more specific theoretical claims, all resources show a strong link between event structure and argument structure, as well as the definition of a verb class as a group of verbs which is semantically defined and syntactically coherent.

All the resources presented here offer a substantial body of data and rich theoretical analyses which are of value and importance for theoretical linguists. Although not flawless, these resources present valuable information about verbs and verb classes. We hope to have demonstrated their theoretical value with the brief comparison of verbs of creation in 7 different languages, based on information gathered from the resources presented here. VerbNet, BVI, AnCoraVerb, VerboWeb, CROVALLEX, and VALLEX provide a bird's eye view of the verbal lexicons of English, Basque, Spanish and Catalan, Brazilian Portuguese, Croatian, and Czech, respectively.

Finally, these resources are great examples of Open Science practices, as they provide free access to all data and analyses for whoever has an internet connection, as well as recognize the authorship of data gathering. The scientific community has acknowledged the importance of open access to data, as well as the recognition of authorship in data work, in linguistics and other fields. The reader is encouraged to access the websites of each resource presented here (and of others listed in Appendix A) and to find more information on specific topics. Although these resources assume a specific lexical-semantic approach for verb classification, they provide valuable data and analyses that can be useful for research in different theoretical perspectives.

## 6 Acknowledgments

The author thanks the financial support of Fulbright Brazil (Fulbright Junior Faculty Member Award 2020/2021 - Department of Linguistics of the University of New Mexico/Fall semester 2021).

## References

- AMARAL, L. A polysemy account of alternating verbs of creation in Brazilian Portuguese. *Studies in Hispanic and Lusophone Linguistics*, v. 9, n. 1, p. 3–28, May 2016. DOI: 10.1515/shll-2016-0001. Available from: <https://www.degruyter.com/document/doi/10.1515/shll-2016-0001/html>. Visited on: 8 June 2022.
- AMARAL, L.; CANÇADO, M. Verbos de criação do português brasileiro: classificação e representação lexical. *Linguística*, v. 10, p. 51–73, 2014. DOI: 10.31513/linguistica.2014.v10n1a4580.
- AMARAL, L.; CANÇADO, M. Metonymy triggers syntactic argument alternation: vehicle for conductor metonymy as a constraint on lexical-constructural integration. *Cognitive Linguistics*, v. 31, n. 1, p. 113–148, Feb. 2020. DOI: 10.1515/cog-2018-0098. Available from: <https://www.degruyter.com/document/doi/10.1515/cog-2018-0098/html>. Visited on: 8 June 2022.
- APARICIO, J.; TAULÉ, M.; MARTÍ, M.A. AnCora-Verb: two large-scale lexicons for Catalan and Spanish. In: XIII Euralex International Congress. Barcelona: [s.n.], 2008. Available from: [https://euralex.org/elx\\_proceedings/Euralex2008/00printedbook.pdf](https://euralex.org/elx_proceedings/Euralex2008/00printedbook.pdf). Visited on: 10 Mar. 2022.
- BEAVERS, J.; KOONTZ-GARBODEN, A. *The Roots of Verbal Meaning*. 1. ed. [S.l.]: Oxford University Press, Mar. 2020. DOI: 10.1093/oso/9780198855781.001.0001. Available from: <https://oxford.universitypressscholarship.com/view/10.1093/oso/9780198855781.001.0001/oso-9780198855781>. Visited on: 8 June 2022.
- BONIAL, C. et al. PropBank: semantics of new predicate types. In: THE 9th edition of the Language Resources and Evaluation Conference. Reykjavik, Iceland: [s.n.], 2014.
- BORBA, F. S. *Dicionário gramatical de verbos do português contemporâneo do Brasil*. 1a. ed. São Paulo: Editora UNESP, 1990.
- BROWN, S.; BONN, J., et al. VerbNet representations: subevent semantics for transfer verbs. In: FIRST International Workshop on Designing Meaning Representations. Florence, Italy: [s.n.], 2019.
- BROWN, S.; PUSTEJOVSKY, J., et al. Integrating Generative Lexicon Event Structures into VerbNet. In: PROCEEDINGS of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). Miyazaki, Japan: European Language Resources Association (ELRA), May 2018. Available from: <https://aclanthology.org/L18-1009>. Visited on: 8 June 2022.
- CANÇADO, M.; AMARAL, L. *Introdução à semântica lexical: papéis temáticos, aspecto lexical e decomposição de predicados*. Petrópolis: Editora Vozes, 2016.
- CANÇADO, M.; AMARAL, L.; MEIRELLES, L. *VerboWeb: syntactic-semantic classification of Brazilian Portuguese verbs*. [S.l.: s.n.], 2017. Available from: <http://www.lettras.ufmg.br/verboweb/>. Visited on: 8 June 2022.

- CANÇADO, M.; AMARAL, L.; MEIRELLES, L. Verboweb: Uma proposta de classificação verbal. *Revista da Anpoll*, v. 1, n. 46, p. 123–141, Oct. 2018. DOI: 10.18309/anp.v1i46.1077. Available from: <https://revistadaanpoll.emnuvens.com.br/revista/article/view/1077>. Visited on: 8 June 2022.
- CANÇADO, M.; GODOY, L.; AMARAL, L. *Catálogo de verbos do português brasileiro: classificação verbal segundo a decomposição de predicados*. Belo Horizonte: Editora da UFMG, 2013.
- CANÇADO, M.; GONÇALVES, A. Lexical Semantics: Verb Classes and Alternations. In: WETZELS, L.; COSTA, J.; MENUZZI, S. (Eds.). *The Handbook of Portuguese Linguistics*. 1. ed. [S.l.]: Wiley, May 2016. P. 374–391. DOI: 10.1002/9781118791844.ch20. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/9781118791844.ch20>. Visited on: 8 June 2022.
- CROFT, W. *Verbs: Aspect and Causal Structure*. [S.l.]: Oxford University Press, Mar. 2012. DOI: 10.1093/acprof:oso/9780199248582.001.0001. Available from: <https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199248582.001.0001/acprof-9780199248582>. Visited on: 8 June 2022.
- CROFT, W.; VIGUS, M. Constructions, frames and event structure. In: THE 2017 AAAI Spring Symposium. Stanford: [s.n.], 2017.
- DE CLERCK, B.; COLLEMAN, T.; WILLEMS, D. Introduction: A multifaceted approach to verb classes. *Linguistics*, v. 51, n. 4, Jan. 2013. DOI: 10.1515/ling-2013-0030. Available from: <https://www.degruyter.com/document/doi/10.1515/ling-2013-0030/html>. Visited on: 8 June 2022.
- DOWTY, D. *Word Meaning and Montague Grammar*. Dordrecht: Springer Netherlands, 1979.
- DOWTY, D. Thematic proto-roles and argument selection. *Language*, v. 67, n. 3, p. 547–619, 1991. DOI: 10.1353/lan.1991.0021. Available from: <https://muse.jhu.edu/content/crossref/journals/language/v067/67.3.dowty.html>. Visited on: 8 June 2022.
- ESTARRONA, A. et al. A methodology for the semiautomatic annotation of EPEC-RoSem, a Basque corpus labeled at predicate level following the PropBank-VerbNet model. *Digital Scholarship in the Humanities*, v. 31, n. 3, p. 470–492, Sept. 2015. DOI: 10.1093/llc/fqv001. Available from: <https://academic.oup.com/dsh/article-lookup/doi/10.1093/llc/fqv001>. Visited on: 8 June 2022.
- FELLBAUM, C. (Ed.). *WordNet: An Electronic Lexical Database*. [S.l.]: The MIT Press, 1998. DOI: 10.7551/mitpress/7287.001.0001. Available from: <https://direct.mit.edu/books/book/1928/wordnetan-electronic-lexical-database>. Visited on: 8 June 2022.
- FILLMORE, C. The case for case. In: BACH, E.; HARMS, R. T. (Eds.). *Universals in linguistic theory*. New York, NY: Holt, Rinehart & Winston, 1968. P. 1–90.
- FILLMORE, C. The grammar of hitting and breaking. In: FILLMORE, C. (Ed.). *Form and meaning in language: papers on semantics roles*. Stanford, CA: CSLI Publications, 2003. P. 123–139.
- GODOY, L. *Os verbos recíprocos no PB: interface sintaxe-semântica lexical*. 2008. Master's thesis (MA in Linguistic Studies) – Universidade Federal de Minas Gerais, Belo Horizonte.
- GOLDBERG, A. E. *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press, 1995.
- HAJIČOVÁ, E.; PANEVOVÁ, J.; SGALL, P. *A manual for tectogrammatical tagging of the Prague dependency treebank*. [S.l.], 2002.
- HARTMANN, L.; HASPELMATH, M.; TAYLOR, B. *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. 2013. Available from: <https://valpal.info/>. Visited on: 8 June 2022.
- JACKENDOFF, R. *Semantic structures*. Cambridge, Mass: MIT Press, 1990.
- KALM, P.; REGAN, M.; CROFT, W. Event structure representation: between verbs and argument structure constructions. In: THE first international workshop on designing meaning representations. Stroudsburg: [s.n.], 2019.
- KETTNEROVÁ, V.; LOPATKOVÁ, M.; BEJČEK, E. The syntax-semantics interface of Czech verbs in the valency lexicon. In: THE 15th EURALEX international congress. Oslo: [s.n.], 2012.
- KIPPER, K. et al. A large-scale classification of English verbs. *Language Resources and Evaluation*, v. 42, n. 1, p. 21–40, Mar. 2008. DOI: 10.1007/s10579-007-9048-2. Available from: <http://link.springer.com/10.1007/s10579-007-9048-2>. Visited on: 8 June 2022.
- LEVIN, B. *English verb classes and alternations: a preliminary investigation*. Chicago: University of Chicago Press, 1993.

- LEVIN, B. Argument Structure. In: ARONOFF, M. (Ed.). *Oxford Bibliographies in Linguistics*. New York: Oxford University Press, May 2018. DOI: 10.1093/obo/9780199772810-0099. Available from: <https://oxfordbibliographies.com/view/document/obo-9780199772810/obo-9780199772810-0099.xml>. Visited on: 8 June 2022.
- LEVIN, B.; RAPPAPORT HOVAV, M. *Argument Realization*. 1. ed. Cambridge: Cambridge University Press, June 2005. DOI: 10.1017/CBO9780511610479. Available from: <https://www.cambridge.org/core/product/identifier/9780511610479/type/book>. Visited on: 8 June 2022.
- LEVIN, B.; RAPPAPORT HOVAV, M. Lexicalized meaning and manner/result complementarity. In: ARSENIJEVIC, B.; GEHRKE, B.; MARÍN, R. (Eds.). *Studies in the composition and decomposition of event predicates*. Dordrecht: Springer, 2013. P. 49–70.
- MALCHUKOV, A.; COMRIE, B. (Eds.). *Introducing the Framework, and Case Studies from Africa and Eurasia*. [S.l.]: De Gruyter, Aug. 2015. DOI: 10.1515/9783110338812. Available from: <https://www.degruyter.com/document/doi/10.1515/9783110338812/html>. Visited on: 13 June 2022.
- MEIRELLES, L. L.; CANÇADO, M. A propriedade semântica movimento na representação lexical dos verbos do português brasileiro. *Alfa : Revista de Linguística (São José do Rio Preto)*, v. 61, n. 2, p. 425–450, Aug. 2017. DOI: 10.1590/1981-5794-1709-8. Available from: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1981-57942017000200425&lng=pt&tlng=pt](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1981-57942017000200425&lng=pt&tlng=pt). Visited on: 8 June 2022.
- MOGUŠ, M.; BRATANIĆ, M.; TADIĆ, M. *Hrvatski pestotni rječni*. Zavod za lingvistiku i školska knjiga: Zagreb, 1999.
- PRERADOVIC, N.M. CROVALLEX lexicon improvements: subcategorization and semantic constraints. *WSEAS Transactions on Computers*, v. 3, p. 288–298, 2010.
- PRERADOVIC, N.M; BORAS, D.; KISICEK, S. CROVALLEX: Croatian verb valence lexicon. In: THE ITI 2009 31st International Conference on Information Technology Interfaces. Zagreb: [s.n.], 2009. Available from: <https://core.ac.uk/download/pdf/299371065.pdf>.
- RAMCHAND, G. *Situations and syntactic structures: rethinking auxiliaries and order in English*. Cambridge, Massachusetts: The MIT Press, 2018.
- RAPPAPORT HOVAV, M.; LEVIN, B. Building verb meanings. In: BUTT, M.; GEUDER, W. (Eds.). *The projection of arguments: lexical and syntactic constraints*. Stanford: CSLI Publications, 1998. P. 97–134.
- ROTHSTEIN, S. *Structuring events: a study in the semantics of lexical aspect*. Malden, MA: Blackwell, 2004.
- RUPPENHOFER, J. et al. *FrameNet II: extended theory and practice*. [S.l.: s.n.], 2016. Available from: <https://framenet2.icsi.berkeley.edu/docs/r1.7/book.pdf>. Visited on: 25 Nov. 2021.
- SCHULER, K.K. *VerbNet: a broad-coverage, comprehensive verb lexicon*. 2005. Phd dissertation (Doctorate in Computational Linguistics) – University of Pennsylvania, Philadelphia.
- SERRAPILHEIRA. *Guia de boas práticas em ciência aberta e reprodutível*. 2020. Available from: [https://olz34z4bb51rsojq274o1g19-wpengine.netdna-ssl.com/wp-content/uploads/2019/11/serrapilheira-guia\\_ciencia\\_aberta-e-reprodutivel.pdf](https://olz34z4bb51rsojq274o1g19-wpengine.netdna-ssl.com/wp-content/uploads/2019/11/serrapilheira-guia_ciencia_aberta-e-reprodutivel.pdf). Visited on: 25 Nov. 2021.
- SMITH, C. *The Parameter of Aspect*. Dordrecht: Springer Netherlands, 1991. V. 43. DOI: 10.1007/978-94-015-7911-7. Available from: <http://link.springer.com/10.1007/978-94-015-7911-7>. Visited on: 8 June 2022.
- TAULÉ, M.; MARTÍ, M. A.; BORREGA, O. AnCora-Net: Mapping the Spanish AnCora-Verb lexicon to Verb-Net. *Interdisciplinary workshop on verbs the identification and representation of verb features*, Jan. 2010. Available from: [https://www.academia.edu/871272/AnCora\\_Net\\_Mapping\\_the\\_Spanish\\_AnCora\\_Verb\\_lexicon\\_to\\_Verb\\_Net](https://www.academia.edu/871272/AnCora_Net_Mapping_the_Spanish_AnCora_Verb_lexicon_to_Verb_Net). Visited on: 8 June 2022.
- TESNIÈRE, L. *Éléments de syntaxe structurale*. Paris: Klincksieck, 1959.
- VAN VALIN, R. *Exploring the Syntax-Semantics Interface*. 1. ed. Cambridge: Cambridge University Press, July 2005. DOI: 10.1017/CBO9780511610578. Available from: <https://www.cambridge.org/core/product/identifier/9780511610578/type/book>. Visited on: 8 June 2022.
- VENDLER, Z. *Linguistics in philosophy*. Ithaca: Cornell Univ. Press, 1967.



## **A APPENDIX A: online resources for the study of verbs**

AnCora Project: <http://clitc.uib.edu/corpus/lexicons/>  
BVI/E-Rolda: <http://ixa2.si.ehu.es/e-rola/>  
CROVALLEX: <http://theta.ffzg.hr/crovallex/>  
Leipzig Valency Classes Project (ValPal): <https://valpal.info/>  
Multilingual Central Repository: <https://adimen.si.ehu.es/web/MCR>  
Unified Verb Index (FrameNet, VerbNet, PropBank, and others): <https://uvi.colorado.edu/>  
VALLEX: <https://ufal.mff.cuni.cz/vallex>  
VerboWeb: [www.letras.ufmg.br/verboweb](http://www.letras.ufmg.br/verboweb)  
WordNet: <https://wordnet.princeton.edu/>

## **B APPENDIX B: Abbreviations used in the text**

abl: ablative case  
abs: absolutive case  
ACT: actor/agent (functor)  
ADDR: addressee (functor)  
AGT: agent (functor)  
arg0: subject argument  
arg1: object argument  
arg2: oblique argument  
argM: adjunct  
BEN: benefactive (functor)  
cau: cause  
cc: oblique argument  
cd: object argument  
dat: dative case  
diat: diathesis  
DIR: directional (functor)  
erg: ergative case  
Impf: imperfective aspect  
ins/soz: instrumental case  
loc/LOC: locative role/functor  
MANN: manner (functor)  
obl: obligatory  
opt: optional  
ORIG: origin (functor)  
PAT: patient (functor)  
REC: recipient (functor)  
recipr: reciprocal  
reflex: reflexive  
subj: subject argument  
tem: theme  
TWHEN: temporal “when”  
typ: typical argument or modifier