The Main Features of the *e-Glava* Online Valency Dictionary

Matea Birtić, Ivana Brač, Siniša Runjaić

Institute of Croatian Language and Linguistics, Ulica Republike Austrije 16, HR-10000
Zagreb, Croatia
E.mail: mbirtic@ihjj.hr, ibrac@ihjj.hr, srunjaic@ihjj.hr

Abstract

E-Glava is an online valency dictionary of Croatian verbs. The theoretical approach to valency follows the German tradition, particularly that of the VALBU dictionary, with some minor changes and adjustments. The main principle of our valency approach is to link valency patterns to specific verb meanings. The verb list is compiled semi-automatically on the basis of the Croatian Frequency Dictionary and Croatian language textbooks. Currently, *e-Glava* contains descriptions of 57 psychological verbs with 187 meanings and 375 valency patterns. The lexicographic articles are written in Tschwanelex. A Document Type Definition editing module has been used, and the description of verbs follows a three-level linguistic schema prepared for lexicographers. Verbs are distributed throughout 34 semantic classes, and examples are extracted manually from Croatian corpora. Fully processed data for each semantic class will be publicly available in the form of a browsable HTML dictionary. The paper also presents a comparison between *e-Glava* and other cognate resources, as well as a summary of its main advantages, disadvantages, and potential applied uses.

Keywords: Croatian language; valency dictionary; e-dictionary; syntax

1. Introduction

Sentence structure and the syntactic behaviour of verbs were perhaps the most intriguing and interesting topics for early grammatical descriptions and, later, linguistic descriptions of language. Valency properties are relevant to both theoretical and applied linguistic considerations. One way to apply valency theory to real linguistic data is by processing valency e-lexicons and e-dictionaries and corresponding lexical databases intended for use by both humans and computers.

This paper will show the main features of one such e-dictionary, which was created for the Croatian language: *e-Glava*[1]. At present, *e-Glava* is a browsable HTML valency dictionary of Croatian verbs, and it represents the public results of the first phase of the *Valency Database of Croatian Verbs* project. It is accessible at http://valencije.ihjj.hr. It currently contains 57 verbs belonging to the semantic class of psychological verbs, with 187 meanings and 375 valency patterns. *E-Glava* is intended to serve as a tool for researchers interested in valency patterns of Croatian verbs, as well as a tool for teachers and students of Croatian as a second language and as an additional resource for linguistic data linking.

---

[1] In Croatian, *glava* means 'head'. It is also an abbreviation composed of *gla-* (short for *glagolska* 'verbal') and *-va* (short for *valencija* 'valency').

The first part of this paper is an introduction. The second part describes *e-Glava*'s approach to valency. The third part shows how the verb list was compiled, how the verbs were distributed throughout the semantic classes, and how corpora were used. The fourth part describes the layout of lemmas. The fifth part provides a brief description of the computational basis of e-*Glava*. In the sixth part the approaches of other online valency dictionaries are compared to e-*Glava*. The seventh and final part is a conclusion outlining the main advantages and disadvantages of *e-Glava*.

2. The approach to valency

The model of verb valency used in *e-Glava* is based on the fruitful results of German valency research and their lexicographic application in valency dictionaries (Helbig & Schenkel, 1973; Engel & Schumacher, 1978; Schumacher et al., 2004). Our direct model was the German VALBU valency dictionary (Schumacher et al., 2004), and its online version E-VALBU.

There are a number of other online dictionaries or lexical databases (also for Croatian) that process the syntactic environments of lexical units and valency in different frameworks. Differences and similarities of these databases to *e-Glava* will be described in the sixth section.

We have chosen a theoretical model based on the German valency tradition for two reasons: some previous theoretical discussions and lexicographic descriptions of verb valency in Croatian have also been written following the same tradition, such as Samardžija (1986) or Filipović (1993); and the model is simple enough that lexicographers with different backgrounds can master it. The basic assumption of VALBU's approach is the identification of valency complements at the level of sublemmas or meanings, not at the level of verb or lemma. The same principle is used in the description of verbs in *e-Glava*. A verb has one or several meanings, and each meaning or sublemma is linked with one or several valency patterns. We assume the sentence analysis used in traditional valency frameworks, whereby the verb is the center of the sentence. All syntactic (nonverbal) phrases, except for conjunctions and particles, are either complements or adjuncts. The verb selects the complement of a specific morphological form, which must have a special semantic relationship to a part of the meaning of the verb. Complements can be obligatory or optional, while adjuncts are never obligatory. Valency descriptions deal with optional and obligatory complements, while adjuncts are not part of the description. However, the practice is to record some common adjuncts as additional information belonging to the sublemma. Valency descriptions begin with the extraction of a part of a sentence that has been identified as a complement. Each complement is described as a morphological, syntactic, and semantic unit. The introduction of the morphological layer of analysis departs from the VALBU model, which describes complements only syntactically and semantically.

2.1. The syntactic level

We assume that 10 complement classes are needed at the syntactic level, i.e., for valency description in the narrow sense: Nominative Complement, Genitive Complement, Dative Complement, Accusative Complement, Instrumental Complement, Prepositional Complement, Adverbial Complement, Predicative Complement, Infinitive Complement and Sentential Complement.2 The VALBU model assumes eight classes of complements: Nominative Complement, Genitive Complement, Dative Complement, Accusative Complement, Prepositional Complement, Adverbial Complement, Predicative Complement and Sentential Complement. The Croatian model has one additional case (Instrumental Complement) due to the Croatian case system.3 Like the German model, we also use the Prepositional, Adverbial and Predicative Complements. One point at which we differ considerably from the VALBU model is in our treatment of Sentential and Infinitive Complements.

In the German model, complement sentences are viewed as a realization of either Case / Prepositional or Verbal Complement (*Verbativergänzung*). If sentences co-occur with verbs that otherwise take Case or Prepositional Complements, they are considered part of a specific Case or Prepositional Complement. If a sentence appears as a complement of a verb that does not take a Case or Prepositional Complement, the sentences together with the infinitives belong to the Verbal Complement. In a way, the VALBU model views sentences only as realizations of some other complement. In our model, all sentences as complements of verbs are regarded as a unique class of Sentential Complements, while infinitive complements belong to a separate class of Infinitive Complements. In the following passages, we will describe the 10 classes of complements in detail.

2.1.1. Nominative Complement

The Nominative Complement corresponds to the traditional concept of the subject. The majority of verbs in Croatian have a Nominative Complement and only a few do not. The Nominative Complement is always obligatory. Verbs which lack any complements are avalent verbs. There are also verbs which have one or two complements, neither of which belong to the Nominative Complement. Also, not all noun phrases with the nominative case belong to the Nominative Complement.

Nominative nouns or pronouns in copular sentences (*Žena je profesorica* (nom[4]) 'The woman is a teacher') or in secondary predication (*Marko je postao bogataš* (nom)

---

[2] Samardžija (1986) also assumes that for description of Croatian valency patterns ten complement classes are needed.

[3] The Croatian case system has seven cases, but there are five Case complement classes. The vocative is, today, never a case on an argument, and the locative case is always found within a prepositional phrase.

[4] Abbreviation used in this paper: nom = nominative; gen = genitive; dat = dative; acc = accusative; inst = instrumental; pl = plural; sg = singular; fem = feminine gender; pres = present tense; past = past tense.

'Marko became a rich man') belong to the Predicative Complement. In Croatian, which is a null pronoun language, pronominal subjects do not need to be expressed in the first and second person ('*Došla sam.*' Came – 1sg past fem 'I came', '*Jedeš.* (Eat – 2sg pres 'You are eating.'). We assume the Nominative Complement is also present in these sentences, though not realized. In such cases, a personal pronoun is added in parentheses following the verb. The way these examples are processed is illustrated in Figure 1.

2 doživljavati *vidjeti koga/što kakvim ; imati kakav dojam o kome/čemu*

NomD, AkD, PredikD

◊ U tom pogledu našu budućnost doživljavam (ja) vrlo svijetlom.

   ja - NomD: nominativ [onaj tko što vidi kakvim: živo, osoba, skupina ljudi]

   našu budućnost - AkD: akuzativ [ono što tko vidi kakvim: bez ograničenja]

   vrlo svijetlom - PredikD: instrumental + pridjev [onakvo kakvim tko što vidi: stanje, svojstvo]

Figure 1: The layout of the verb with the unexpressed Nominative Complement

2.1.2. Genitive Complement

The Genitive Complement mostly corresponds to the genitive object (e.g., *Svijet se sjeća pape Wojtyle* 'The world remembers Pope Wojtyla'). Also, in processing psychological verbs, we decided to define the complements of some existential verbs as the Genitive Complement (*Ovdje nedostaje etike* (gen pl) *i morala* (gen pl) 'Ethics and morality are lacking here'). Genitive noun phrases with existential verbs are considered partitive genitive. Prototypical instances of partitive genitives are found in the object position where the genitive form replaces the accusative. Despite the similarities, we do not consider the partitive genitive in the object position as a separate (Genitive) complement, but rather a realization of the Accusative Complement. In the case of existential verbs, we find the partitive genitive only in some verb meanings, while other meanings use the nominative case. Thus, the nominative and genitive are not interchangeable in some verb meanings. This is why we have introduced a separate complement in the case of several existential verbs.

2.1.3. Dative Complement

The Dative Complement includes indirect objects and logical subjects marked with the dative case (*Oko se divi ljepoti* (dat) 'The eye admires beauty'; *Vrti mi* (dat) *se* 'I am dizzy'). The Dative Complement can be both obligatory and optional. Apart from being complements, nouns marked with the dative case are frequently adjuncts as well. Logical subjects in the dative case, dative experiencers, or dative stimuli with psychological verbs (*Blanki* (dat) *je dosadila duga kosa* 'Blanka is bored with long hair'); equational datives (*Lijeva strana odgovara desnoj* (dat) 'The left side

corresponds to the right'); predicative datives (*Maslina pripada voću* (dat) 'The olive belongs to [the category of] fruit'); and some directional datives (*Prišao je djevojci* (dat) *na šanku* 'He approached the girl at the bar') are considered obligatory Dative Complements. Dative nouns with a thematic role of recipient frequently belong to the optional category of Dative Complements (*Ona mi* (dat) *se žalila na bolove* 'She complained (to me) of her pain'). The ethical dative is an adjunct (*Ona mi* (dat) *se danas dobro osjeća* 'She (to/for me) feels well today').

2.1.4. Accusative Complement

The Accusative Complement corresponds to the direct object. Not all noun phrases marked with the accusative case are part of the Accusative Complement. Some belong to the Adverbial Complement, also known as 'measure accusatives' (*Kaput je stajao hrpu novaca* (acc) 'The coat cost a pile of money'); or to adjuncts – very often to manner adjuncts (*Hodali su ruku pod ruku* 'They walked arm in arm'). However, cognate objects belong to the Accusative Complement due to their argumental properties (see Birtić & Matas Ivanković, 2009). As stated above, the partitive genitive and the genitive of negation in the object position are considered Accusative Complements.

2.1.5. Instrumental Complement

The Instrumental Complement comprises indirect objects in the instrumental case (*Ronaldo se ponosi sinom* (inst) 'Ronaldo is proud of his son') and of nominal phrases with the semantic role of instrument, which traditional grammars consider adjuncts (*Razveseljavali su nas svojim pričama* (inst) 'They cheered us up with their stories'; *Marko se oženio Ivanom* (inst) 'Marko married Ivana'). Some nouns in the instrumental case are part of a Predicative Complement (*Svi ga doživljavaju svecem* (inst) 'They all consider him a saint'). Also, many nouns in the instrumental case belong to adjuncts (*Hodao je ulicom* 'He walked down the street'). Instrumental Complements with divalent verbs are mostly obligatory, while Instrumental Complements with trivalent verbs are mostly optional.

2.1.6. Prepositional Complement

The Prepositional Complement is a complement described by traditional grammars as a prepositional object (*Zaljubila sam se u tebe* 'I fell in love with you'; *Ne ljute se svi roditelji na svoju djecu* 'Not all parents get angry at their children'). Prepositional phrases also belong to the category of Predicative Complements (*Smatrali su ga za prijatelja* 'They consider him a friend'); Adverbial Complements (*Ona živi u Londonu* 'She lives in London'); or frequently to the category of adjuncts (*Više se ne uzrujavam zbog sitnih pogrešaka* 'I do not get upset about minor errors anymore').

2.1.7. Adverbial Complements

Although most of adverbial phrases are optional adjuncts, it has been observed that some adverbials cannot be omitted, and their presence is decisive for the grammaticality of a sentence (Samardžija, 1986; Silić & Pranjković, 2005, Palić, 2011, Belaj & Tanacković Faletar, 2017). Such adverbials express location (*Ona živi u Londonu* 'She lives in London'; *Bacili su knjigu na stol* 'They threw the book on the table'); manner (*Ponašaju se nepristojno* 'They behave rudely'); cause (*Ta prava ne proistječu iz Ustava* 'These rights do not arise from the constitution'); measures of time and quantity (*Sjednica je trajala tri sata* 'The session lasted three hours'); and results (*Dijete na mlijeko reagira proljevom* 'The child reacts to milk with diarrhea'). The Adverbial Complement is obligatory or optional, but the separation between the optional Adverbial Complement and the adjunct is very complex, and depends mostly on the researcher's intuition and the chosen theory.

2.1.8. Predicative Complement

The Predicative Complement includes syntactic phrases considered part of the predicate, e.g. nouns and adjectives in copular sentences (*Profesor je šutljiv/budala* 'The professor is quiet / a fool') or part of secondary predications (*Oni svi su ga smatrali glupim / budalom / za budalu* 'They all consider him stupid / a fool / as a fool'). The Predicative Complement is realized by noun or adjective phrases in the nominative or instrumental case, by *kao*-phrase '*as*-phrase', prepositional phrase, or adverb. The Predicative Complement is always obligatory.

2.1.9. Infinitive Complement

In our approach, the Infinitive Complement represents a separate class of complements, although it is part of other complements in some models (e.g., in VALBU). Infinitives are often complements of modals and verbs that express phases of an action. Some verbs are not strictly modal, but they attain a modal component of meaning when used with an infinitive (*Bojim se ući* 'I am afraid to enter').

2.1.10. Sentential Complement

The Sentential Complement includes all sentences as complements of verbs. As mentioned above, the VALBU model considers some sentences as part of case and Prepositional Complements, while others (with verbs that do not take case or Prepositional Complements) belong to the Verbal Complement. We decided to keep all sentential complements in a separate complement class regardless of their co-occurrence with verbs which do or do not take case complements for two reasons. Firstly, sometimes it is difficult to decide whether a sentential complement actually substitutes another case complement. Hence, it is easier for a lexicographer to describe a syntactic environment of a verb. Secondly, from the viewpoint of the user,

it is easier to notice that a verb can take sentential complement instead of case complement if the information is conceptually and visually separated.

2.2.   The morphological level

In addition to a syntactic description through 10 classes of complements, each complement is also described morphologically. *E-Glava* regards morphology as the realization of syntax. It is defined that syntactic (valency) complements are realized by four major morphological categories and a number of subcategories. The major morphological categories needed to morphologically describe syntactic complements in Croatian are (1) prepositions, (2) cases, (3) sentential realizations and (4) other. Prepositions include all Croatian prepositions, which amounts to 199. Cases include all Croatian morphological cases (nominative, genitive, dative, accusative, instrumental, and locative[5]) except for the vocative case, which is never realized on verbal arguments; it is always an independent phrase. Sentential realizations include the Croatian conjunctions (*da*, *što*, *kako*, *gdje*, *li*, WH-word, *neka*, *kao+*) and other elements by which a sentence can be introduced next to a verb (quotes and the zero conjunction). Quotes (marked with the word Navod[6]) and the zero conjunction (0) are listed alongside conjunctions. The fourth morphological category (other) includes (4.a.) adverbs and adverbial phrases, (4.b.) the infinitive, (4.c.) *kao*-phrase, (4.d.) quantificational phrases, and (4.e.) adjectives. As is apparent from the list above, morphological categories are not distributed in any meaningful way, but by functional principle. Some morphological realizations are mainly typical for some complements: *kao*-phrase and adjectives are frequently realizations of predicative complements.

2.3.   Semantic level

Complements are semantically described in two layers: the verb-specific description of a participant, i.e. an individual semantic role, and the assignment of a semantic category to a specific complement.[7] For each complement, the individual semantic role is defined on the basis of the definition of the verb's meaning. Semantic categories can be chosen from a list amounting to 34 categories, most of which have been adopted from the VALBU dictionary. Categories such as animate, person, animal, plant, etc. are not organized hierarchically, so both animate and person must be chosen for each complement which can refer to a person. A more developed approach to semantic categories would be a hierarchically ordered tagset of semantic labels, which will be considered for introduction in the next phase of the project. The semantic category is not recorded if any noun can qualify as a realization of a specific

---

[5] Locative does not refer to a complement, but to a morphological subcategory, because for the description of locative prepositional phrases, the locative case must be chosen together with a specific preposition. The locative case never appears outside prepositional phrases in the Croatian language.

[6] Navod 'quote, quotation'.

[7] A similar kind of verb-specific description is also provided by VDE (2004), and some similar features can be found in FrameNet's descriptions of participants (Herbst, 2007: 25–26).

complement. In such cases, the complement is described as 'without restrictions'. Figure 2 below shows the semantic description of the nominative complement that appears with the verb *bojati se* 'fear' (*Marko se boji neprijatelja* 'Marko fears the enemy').

Figure 2: Two-layered semantic description of the Nominative Complement of the verb *bojati se* 'fear'

In addition, every semantic category can be preceded by the label *pren.*, which means 'figurative'. In cases where words are used metaphorically or metonymically, the figurative label is used.

3. Verb list, semantic classes and the usage of corpora

A verb list of approximately 900 of the most frequent verbs necessary for mastering Croatian at the B1 level according to the Common European Framework of Reference for Languages was extracted. The final list of verbs was compiled semi-automatically by comparing a verb list extracted manually from an older Croatian language resource, *Hrvatski čestotni rječnik* (Croatian Frequency Dictionary, Moguš, Bratanić & Tadić, 1999), and a verb list from more comprehensive textbooks of Croatian as a second language (e.g. Čilaš Mikulić et al., 2011; 2012; 2013).[8]

This list of 900 verbs intended for processing in e-*Glava* is distributed among 34[9] semantic classes and 91 subclasses. It is a well-known fact that verbs have several

---

[8] The lists of verbs used in texts are compiled at the end of the textbooks.

[9] Semantic classes in *e-Glava*: verbs of thinking, knowledge and learning; verbs of motion; verbs of communication; verbs of creation and transformation; verbs of positional change and placement; psych verbs; verbs involving the body (somatic verbs); verbs of social interaction; verbs of possession, taking and giving; verbs of change in possession; verbs of change in state; verbs of removing, separating and disassembling; verbs of ruling, control and influence; verbs of perception; verbs of effort and intention; verbs of emission; verbs of killing and hurting; verbs of placement in space; verbs of ingesting; aspectual verbs; verbs of carrying and sending; verbs related to money; general actions; verbs of combining and attaching; verbs of keeping and caring; verbs of inhabiting and staying; verbs of fighting; verbs of usage; verbs of happening; verbs of lingering and rushing; existential verbs; verbs of relations; verbs of judgment and success; weather verbs; and verbs of sounds made by animals.

meanings, and that the most frequent meaning does not always correspond to the prototypical one, so it is important to choose which criteria are to be used for classification. We decided to classify the verbs according to the first meaning written in two monolingual Croatian language dictionaries: *Školski rječnik hrvatskoga jezika* (Croatian School Dictionary) (Birtić et al., 2012), and the *Hrvatski jezični portal* (Croatian Language Portal) online dictionary (http://hjp.znanje.hr/). If these dictionaries did not have the same meaning written in the first place, we followed *Školski rječnik*, because it is a corpus-based dictionary (Birtić et al., 2012: xii). Our general classification is inspired by Levin (1993), but it relies more on verb semantics than syntax as compared to Levin's approach, which classifies verbs mainly on the basis of syntactic alternations. As will become clear below, each verb belongs to one prototypical semantic class, but their different meanings also allow them to belong to other semantic classes. This multiple categorisation is enabled through the ability to choose a semantic class at different levels in the description. The prototypical semantic class is written next to the lemma, and possible changes in semantic class are recorded next to the sublemma, i.e., a specific meaning of the verb.

As the verbs are processed according to their semantic classification, not according to alphabetical order, semantic class is considered a module (Klosa, 2013) or a phase in the lexicographic process. The advantage of this approach is that it enables the observation of syntactic and semantic differences between similar verbs, or of syntactic alternations in the same semantic class, such as the well-known syntactic alternations in psychological verbs (psychological verbs can express an experiencer either as subject or object, and in Croatian, a language with morphological cases, the experiencer can be realized as a noun in the nominative, accusative, or dative case). An additional advantage is that the combined processing of verbs of the same semantic class enables non-native speakers to learn how to presuppose valency patterns according to the semantic group the verb belongs to.

The processing of verbs in *e-Glava* is based on two Croatian corpora: *Hrvatska jezična riznica* (Croatian Language Repository) and *Hrvatski mrežni korpus* – hrWaC (Croatian web corpus – hrWaC), but is not directly linked to any (annotated) corpora. The Croatian Language Repository, which is also compiled at the Institute of Croatian Language and Linguistics, did not comprise annotated corpora when the project *e-Glava* begun, but its annotation has recently started. Manual corpora research is relevant at the three stages of verb processing. Firstly, corpus is a tool which enables us to check definitions of verb meanings already noted in existing dictionaries. It helps us to find the meanings of the verbs that have not yet been recorded. Secondly, after all the meanings of a verb have been identified, the corpus is searched to find valency patterns which belong to each meaning. Finally, the corpus examples are selected manually and entered into a database.

4. The three-level description of verbs in e-Glava

E-Glava describes verbs on three levels. The first level provides information regarding the verb overall, the second level introduces different meanings of the verb, and the third level is a valency description.

4.1. The first level

The first level consists of a verb lemma in the infinitive, except for inherent reflexive verbs, which are entered with the reflexive particle *se*. Each lemma or verb is connected with four sections: a grammatical block, the prototypical semantic class of the verb and its subclass, idioms and collocations, and notes. The grammatical block encompasses verb inflections (first person singular present, third person plural present, masculine perfect participle, feminine perfect participle and masculine passive participle), and an aspect label. The aspect label includes abbreviations for imperfective, perfective and biaspectual values. In *e-Glava*, the semantic class of a verb is visualized directly below the lemma and above the verb inflections. The idiom and collocation block is placed at the end of the lemma visualization. It consists of a collocation or an idiom (e.g. *mrziti iz dna duše* 'to hate from the depths of one's soul'); its definition (*jako mrziti koga ili što* 'to strongly hate someone or something'); and a usage example (*Ako idete na posao, mrzit ćete budilicu iz dna duše* 'If you go to work, you will hate your alarm clock from the depths of your soul'). The note block contains information that applies to the verb overall, not to one of its meanings or a separate valency pattern (for example, the remark that a specific verb is non-standard or is used only in a specific style).

4.2. The second level

The second description level consists of different meanings of verbs, which are introduced by numbered sublemmas (e.g., *1 mrziti, 2 mrziti* 'hate'). Each sublemma is connected with a reflexive label, a definition, a possibility of changing a verb's semantic class, and additional information. The reflexive label has two values: reflexive and zero. The reflexive value mostly serves to mark the reflexivity of reflexive verbs which are not *reflexiva tantum* or inherently reflexive, i.e., those entered with particle *se*. All reflexive verbs that are not inherently reflexive are treated as sublemmas, i.e., as one of the meanings of the verb. Definitions consist of three parts: a stylistic label, paraphrase definitions (two can be entered) and synonyms. The stylistic label (e.g., historical, poetic) precedes the definition.

An illustration of the first and second levels of the description of the verb *vrijeđati* 'offend, insult, irritate', with an introduction of the separate sublemmas for particular meanings, is provided in Figure 3.

Figure 3: An illustration of the first and second level of the description of the verb *vrijeđati* 'offend, insult, irritate'

4.3. The third level

Clicking on a sublemma brings the user to the third level, which contains the valency analysis. The valency analysis consists of an example sentence and parts of sentences recognized as valency complements. Valency analyses contain a morphological, syntactic and semantic description of a complement (in square brackets). Above the detailed valency analysis, valency patterns are written as abbreviations of complements (e.g., NomD, InfD).[10] Each meaning can be associated with several valency patterns, and each valency pattern can be linked to several examples.

This is illustrated in Figure 4, which provides the complete processing of the verb *živcirati* 'to upset someone/to become irritated'. This illustration shows the sentence examples, which are introduced with a diamond. The example section, shown below, is connected to the syntactic, morphological and semantic descriptions with a hyphen.

---

[10] NomD is an abbreviation for *Nominativna dopuna* 'Nominative Complement', InfD is an abbreviation for *Infinitivna dopuna* 'Infinitive Complement'.

živcirati *nesvr.*

prez. 1. l. jd. živciram, *3. l. mn.* živciraju, *prid. r. m.* živcirao, *prid. r. ž.* živcirala, *gl. prid. trp.* živciran

1 živcirati *izazivati u kome živčanost*

NomD, AkD

◊ Mene zaista živciraju takve špekulacije.

    takve špekulacije - NomD: nominativ [ono što u kome izaziva živčanost: bez ograničenja]

    mene - AkD: akuzativ [onaj u kome što izaziva živčanost: živo, osoba, skupina ljudi]

◊ Našu javnost najviše intrigiraju i živciraju Balkan i balkanizacija.

    Balkan i balkanizacija - NomD: nominativ [ono što u kome izaziva živčanost: bez ograničenja]

    našu javnost - AkD: akuzativ [onaj u kome što izaziva živčanost: živo, osoba, skupina ljudi]

RečD, AkD

◊ Douglasa posebno živcira što se u svakoj prigodi nježno drže za ruke.

    što se u svakoj prigodi nježno drže za ruke - RečD: što [ono što u kome izaziva živčanost: propozicija]

    Douglasa - AkD: akuzativ [onaj u kome što izaziva živčanost: živo, osoba, skupina ljudi]

NomD, AkD (InstD)

◊ Ona me je živcirala neobičnošću i ekscentričnošću.

    ona - NomD: nominativ [ono što u kome izaziva živčanost: bez ograničenja]

    me - AkD: akuzativ [onaj u kome što izaziva živčanost: živo, osoba, skupina ljudi]

    neobičnošću i ekscentričnošću - (InstD): instrumental [ono čime što u kome izaziva živčanost: bez ograničenja]

◊ Populistička lokalna stranka živcirala je ostatak nacije svojim istrijanstvom.

    populistička lokalna stranka - NomD: nominativ [ono što u kome izaziva živčanost: bez ograničenja]

    ostatak nacije - AkD: akuzativ [onaj u kome što izaziva živčanost: živo, osoba, skupina ljudi]

    svojim istrijanstvom - (InstD): instrumental [ono čime što u kome izaziva živčanost: bez ograničenja]

2 živcirati se *biti živčan, katkad izazvan čime ; osjećati se živčan*

NomD

◊ Ja se uvijek živciram kad kasnim.

    ja - NomD: nominativ [onaj tko je živčan: živo, osoba, skupina ljudi]

◊ Hrvatska će se još živcirati u žici stvorenih obveza.

    Hrvatska - NomD: nominativ [onaj tko je živčan: pren. geografsko mjesto]

NomD (PrijeD)

◊ Ministrica pravosuđa nije se živcirala oko rasprave o reformi pravosuđa.

    ministrica pravosuđa - NomD: nominativ [onaj tko je živčan: živo, osoba, skupina ljudi]

    oko rasprave o reformi pravosuđa - (PrijeD): oko + genitiv [ono čime je izazvano da je tko živčan: bez ograničenja]

Obavijest: U ovome značenju (živcirati se 2) uz glagol se često pojavljuje dodatak uzroka ostvaren prijedložnom skupinom s prijedlogom *zbog* i imenicom u genitivu (*zbog* + gen.), npr. *Više se ne živciram zbog sitnih pogrešaka kao prije.*

Čvrste sveze

Figure 4: The complete layout of the verb *živcirati* 'to upset someone/to become irritated'

5. The computational basis of e-Glava

In 2013, a newly formed team of researchers initiated the *Valency Database of Croatian Verbs* project at the *Institute of Croatian Language and Linguistics*, and a linguistic model had been chosen by the end of 2014. Valency had been researched at the Institute prior to this, but the outcomes of these descriptions were compiled as non-structured or linear data. As a part of preparation[11] we had to re-evaluate the entire concept, and the team had to decide whether to develop its own customized Content Management System (CMS) or to use an existing lexicographic package.[12] Considering the fact that there was no funding for the project, and that the team members had previous experience in compiling dictionaries using TshwaneLex, we began to develop a three-level linguistic schema for a valency dictionary in TshwaneLex (see Section 4), which we considered a computerisation phase of our lexicographic process. Accordingly, we began writing new lexicographic entries in the prepared TschwaneLex schema for 57 psychological verbs. The I.T. department attempted to make the dictionary entry writing process as precise and user-friendly as possible for researchers and lexicographers, mostly through the implementation of drop-down menus and controlled multiple choice options for all linguistic features.

After this small dictionary of psychological verbs was compiled, it was made publicly available in order to receive initial feedback from fellow researchers and other interested parties. Although the dictionary grammar was developed using a Document Type Definition (DTD) editing module of TshwaneLex and an ODBC connection, and the DTD was automatically transcribed into a PostgreSQL database environment, the project team still had to make some adjustments before the data could be presented on an internet platform. We decided to export the native XML file for all verbs within the semantic class that were marked "completed" to an easily-accessible SQL database. This process made the part of the dictionary that we consider completed, automatically browsable through a web-based search engine using PHP and HTML5. This gave researchers the ability to make verbs currently being described (the semantic class of verbs of moving and putting) available by

---

[11] Klosa (2013) has defined six phases in computer lexicographic process for online dictionaries under construction: the phase of preparation, the phase of data acquisition, the phase of computerization, the phase of data processing, the phase of data analysis, and the phase of preparation for online release. The phase of preparation is partly described in this section and in Section 3 (criteria for choosing verbs for a verb list). In the phase of data acquisition we decided to use the *Croatian Language Repository*, the *Croatian web corpus – hrWaC*, and the *Croatian Frequency Dictionary* as primary sources. Our secondary sources were textbooks of Croatian as a second language, *Školski rječnik hrvatskog jezika* and *Hrvatski jezični portal* (Section 3). The corpus designed especially for the purpose of *e-Glava* was omitted from this project. The phase of computerization and data processing is described in this section (5): the choice of dictionary writing system and the specification of database system. The phase of data analysis is presented to a lesser extent in last part of Section 3 (the usage of corpora) and mostly in Section 4. Finally, the phase of preparation for online release is described at the end of this section (5). As Klosa (2013) states, following Klein (2004): "all phases of the computer-lexicographical process merge giving yet unknown flexibility to the lexicographer."

[12] For more details, see Birtić & Nahod (2016: 103–105).

exporting an updated XML file, which then goes "live" on the website. In addition to this first version, which is browsable by lemma, an advanced search function is being developed which will enable users to search by specific categories, such as valency complements, morphological forms, or semantic features.

6. A comparison of e- Glava and other online dictionaries and lexicons

This section compares the main features of *e-Glava* to those of some other well-known online dictionaries (FrameNet, FrameBank, VALLEX, Crovallex, VALBU).

One of the most well-known online dictionaries is UC Berkeley's FrameNet, which is based on the theory of frame semantics (Fillmore & Baker, 2010). The most notable difference between descriptions of verbs in *e-Glava* and descriptions of nouns, adjectives, and verbs in FrameNet is their ordering and the hierarchy of their syntactic, morphological and syntactic descriptions. While *e-Glava* begins its valency description with the syntactic level, followed by morphological and semantic layers, FrameNet begins from the semantic layer in accordance with the theory of frame semantics (Fillmore & Baker, 2010). FrameNet derives grammatical function (external argument, object and dependent) and phrase type algorithmically (Ruppenhofer et al., 2016: 41) based on frame element label (semantic role), position in the sentence, and part of speech. Deriving grammatical functions from the position of phrases in sentences is not possible for Slavic languages with free word order. We believe that detailed descriptions of both morphology and syntax are essential for languages with rich morphological systems. For example, the Russian FrameBank also employs morphological descriptions. As can be concluded, *e-Glava* differs considerably from FrameNet in several respects: it deals only with verbs; its starting point is syntax; examples are extracted manually (FrameNet automatically extracts examples from the British National Corpus); and word order is not taken in account.

Semantic and syntactic verb descriptions are a part of the Russian FrameBank (Lyashevskaya & Kashkin, 2011; Lyashevskaya, 2012) and the Czech VALLEX (Kettnerová, Lopatková & Bejček, 2012; Lopatková et al., 2006). The differences between e-*Glava* and FrameBank or VALLEX are less significant than the differences between e-*Glava* and FrameNet. Unlike FrameNet, FrameBank and e-*Glava* take morphology into account. FrameBank and *e-Glava* share some units of description: e.g., the morphosyntactic features of elements in FrameBank and morphological descriptions in e-*Glava*; the lexical-semantic class of elements in FrameBank and semantic categories in e-*Glava* (e.g., human, animate); and the division of complements into optional and obligatory. FrameBank also includes the syntactic rank of elements / grammatical functions (e.g., subject, object, predicate, peripheral and clause) and the semantic roles of arguments (Agent, Patient and Instrument). FrameBank consists of examples taken randomly from the annotated Russian National Corpus (Lyashevskaya & Kashkin, 2011), while *e-Glava* is not linked to any

annotated corpora. In *e-Glava,* examples are chosen intentionally as the best fit for meanings and valency descriptions. FrameBank is based on Construction Grammar (Goldberg, 1995) and the Moscow Semantic School (e.g., Apresjan, 1995).

E-Glava also shares similarities with the Valency Lexicon of Czech Verbs VALLEX which also focuses on the most frequent verbs and their meanings. VALLEX and *e-Glava* share the same general approach to valency: valency patterns are identified at the level of particular verb meanings, not at the level of the verb. VALLEX also provides information on the number of complements, functors, or semantic roles, their morphological realizations, and the obligatoriness of complements (Kettnerová, Lopatková & Bejček, 2012). The same information is provided by *e-Glava*, except that semantic descriptions in *e-Glava* use an individual semantic role and semantic category, not general semantic roles (functors). Both e-dictionaries provide some additional information about idioms, reflexivity, reciprocity and aspect. Reflexivity and aspect values are approached differently in VALLEX and *e-Glava*. Imperfective and perfective verbs are considered the same entry in VALLEX, whereas the perfective and imperfective variants of verbs are considered two separate entries in *e-Glava*. Each imperfective verb in e-*Glava* does not need to have its perfective pair entered by default: each verb lemma is entered independently depending on its frequency of appearance. *E-Glava* enters *reflexiva tantum* or inherent reflexive verbs as separate lemmas, whereas all other reflexive verbs are considered sublemmas of lemma. VALLEX also records *reflexiva tantum* as separate lemmas, but in addition to this, it treats derived reflexives[13] as separate lemmas as well (for more on this, see Oraić Rabušić & Bošnjak Botica, 2016; Kettnerová & Lopatková, 2014). VALLEX entries are also manually taken from the Czech National Corpus. VALLEX divides verbs into 22 semantic classes according to their prototypical meaning, which is based on intuition (Lopatková et al., 2006: xxiii). As we have already stated, verb classification in *e-Glava* is performed in a more systematic and precise manner than in VALLEX. Verbs belong to a prototypical semantic class and can be linked to one or more additional semantic class. In contrast, VALLEX associates each verb only with one semantic class. For some other Slavic languages electronic valency dictionaries or dictionaries including verb descriptions are available, e.g., the Polish Valency Dictionary (Walenty) (Przepiórkowski et al., 2014), Slovene Lexical Database (Gantar & Krek, 2011).

Needless to say, it is very important to mention another online Croatian valency dictionary, CROVALLEX, developed by Mikelić Preradović (2008; 2010). It describes 1,739 verbs with 5,118 valence frames classified into 72 semantic classes and subclasses (173 in total). The number of verb lemmas exceeds the designated number of verbs in *e-Glava*. Just like VALLEX, CROVALLEX also enters only *reflexiva*

---

[13] Derived reflexives are verbs derived from a corresponding non-reflexive verb, but their meaning is so distant from their non-reflexive counterpart that they must be viewed as a separate verb.

tantum and derived reflexives as separate lemmas. As in *e-Glava*, valency is related to meaning, and a valency frame example and class is defined for each meaning. Slots in valence frames are filled with functors, which can be inner participants and free modifications. Functors roughly correspond to deep cases (Agent, Patient, Recipient, Result and Origin) and can appear in a sentence only once. There are about 30 free modifications, and they can appear in a valence frame more than once. According to descriptions in CROVALLEX, inner participants and free modifications can be optional or obligatory (despite the term free). The valence frame is notated with abbreviations of functors. Obligatory vs. optional status is marked in superscript, while morphological form is marked in subscript with the abbreviation of a functor. If the approaches to valency used in both Croatian dictionaries are compared, it can be said that *e-Glava* is more syntax-oriented than CROVALLEX, in which semantic description prevails despite the presence of both syntactic and morphological descriptions. Both share the principle of defining complements on the level of meaning. Verb meanings are finer-grained in *e-Glava*, as they are defined and divided on the basis of Croatian corpora, and do not rely only on available dictionaries. At the level of sentence periphery, CROVALLEX provides more phrases which are considered adjuncts in *e-Glava*. In CROVALLEX, idioms and collocations are listed as a part of verb meaning, while in *e-Glava* they form a separate unit. In terms of semantic classes, CROVALLEX defines a new semantic class for each (different) meaning, but, as opposed to *e-Glava*, it does not specify the prototypical semantic class of a verb.

Finally, although we have followed VALBU quite consistently, there are some points in our treatment of valency in which we depart from our model, as has been mentioned in several parts of this paper. Specifically, we treat reflexive verbs differently: VALBU enters each reflexive verb as a separate lemma entry; we have added morphological descriptions, which are justified for languages with rich morphology; we have introduced semantic verb classes; and we treat Sentential and Infinitive Complements quite differently from VALBU.

7. Conclusion

In conclusion, we would like to outline what has been done so far and set out the main advantages and disadvantages of *e-Glava*. The first version of *e-Glava* is available online and is accessible for free. It offers a detailed description of the syntactic and semantic interface of one semantic class of verbs. Additionally, many verb meanings that are not found in dictionaries of the contemporary Croatian language are described in *e-Glava* thanks to its corpus-based analysis. Consequently, semantic switches and new uses are described. Since it is sometimes an intricate task to properly assign a semantic role to a specific participant, we decided to use semantic (conceptual) categories, e.g., person, animal, place, etc. We also believe these categories to be more intuitively recognizable for dictionary users without formal linguistic expertise. The main disadvantage of *e-Glava* is its manual extraction

of examples and descriptions, which is time-consuming, resulting in slow project progress. On the other hand, this kind of lexicographic work guarantees better and more reliable descriptions.

When we think about other possible usage advantages, it occurs that mastering verb valency is a very important part of language learning, in particular when it comes to learning Slavic languages. *E-Glava* allows non-native speakers to check verb meanings, syntactic patterns, and their morphological realizations. Consequently, *e-Glava* might become a useful tool for learning Croatian as a second language. However, learners should possess a basic understanding of Croatian, as all definitions with simple metalanguage are written in the Croatian language. To master a second language at a higher level, an understanding of idiomatic phrases is also important. Idioms are included and visually represented in a special field separate from the syntactic patterns, and so their meanings can be easily explained to learners.

E-Glava's data could also become an additional resource for linguistic data linking in comprehensive research on the Croatian language. Its detailed descriptions can be used as the starting point for various lexical resources, as the syntactic, semantic and morphological levels are represented as structured data. Related ongoing projects at the Institute of Croatian Language and Linguistics, such as the *Croatian e-Dictionary* (*MREŽNIK*), the *Croatian Collocation Database*, and the *Croatian Metaphor Repository*, could certainly benefit from it. Moreover, *e-Glava*'s research team is open to providing all project data in open format to the greater NLP community in Croatia if they consider it usable for the morphosyntactic and semantic tagging and parsing of corpora or for other processes.

8. References

Apresjan, J.D. (1995). *Izbrannye trudy, tom I. Leksicheskaja semantika.* Moskva: Jazyki Russkoj Kul'tury, Vostochnaja Literatura.

Belaj, B. & Tanacković Faletar, G. (2017). *Kognitivna gramatika hrvatskoga jezika: sintaksa jednostavne rečenice.* Zagreb: Disput.

Birtić, M. & Matas Ivanković, I. (2009). Akuzativne dopune uz neprijelazne glagole: što su unutrašnji objekti?. In *Rasprave Instituta za hrvatski jezik i jezikoslovlje*, 35(1), pp. 1–19.

Birtić, M. & Nahod, B. (2016). An outline of the online valency dictionary of Croatian verbs. In Karolína Skwarska & Elżbieta Kaczmarska (eds.) *Research of Verbal Valency in Slavic Languages in the Past and Present = Výzkum slovesné valence ve slovanských zemích včera a dnes.* Praha: Slovanský ústav AV ČR, v.v.i., pp. 103–116.

Birtić, M. et al. (2012). *Školski rječnik hrvatskoga jezika.* Zagreb: Institut za hrvatski jezik i jezikoslovlje & Školska knjiga.

CROVALLEX 2.0008: *The Croatian Valency Lexicon of Verbs.* (2008.) at: http://theta.ffzg.hr/crovallex/ (1-10 May 2017).

Čilaš Mikulić, M. et al. (2011). *Hrvatski za početnike 1: vježbenica i gramatički pregled hrvatskoga kao drugog i stranog jezika.* Zagreb: Hrvatska sveučilišna naklada.

Čilaš Mikulić, M. et al. (2012). *Razgovarajte s nama! B1-B2: vježbenica, gramatika i fonetika hrvatskog jezika za niži srednji stupanj.* Zagreb: FF press.

Čilaš Mikulić, M. et al. (2013). *Razgovarajte s nama! A2-B1: vježbenica, gramatika i fonetika hrvatskog jezika za niži srednji stupanj.* Zagreb: FF press.

E-VALBU: *das elektronische Valenzwörterbuch deutscher Verben.* Accessed at: http://hypermedia.ids-mannheim.de/evalbu/index.html. (10 May 2017)

Engel, U. & Schumacher, H. (1978). Kleines Valenzlexikon deutscher Verben. Manheim: Institut für deutsche Sprache.

Filipović, R. (ed.) (1993) *Teorija valentnosti i rječnik valentnosti hrvatskih glagola, Kontrastivna analiza engleskog i hrvatskog jezika IV.* Zagreb: Zavod za lingvistiku Filozofskoga fakulteta u Zagrebu.

Fillmore, C.J. & Baker, C.F. (2010). A frames approach to semantic analysis. In Heine, B. & Narrog, H. (eds.) *Oxford Handbook of Linguistic Analysis.* New York: Oxford University Press, pp. 313–341.

Gantar, P. & Krek, S. (2011). Slovene lexical database. In D. Majchraková & R. Garabík (eds.) *Natural language processing, multilinguality: sixth international conference.* Modra, Slovakia, pp. 72–80.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure.* Chicago & London: The University of Chicago Press.

Herbst, T. (2007). Valency complements or valency patterns? In T. Herbst & K. Götz-Votteler (eds.) *Valency: Theoretical, Descriptive and Cognitive Issues.* Berlin: Walter de Gruyter, pp. 15–35.

Helbig, G. & Schenkel, W. (1973). Wörterbuch zur Valenz und Distribution deutsher Verben, Leipzig VEB Bibliographisches Institut.

Hrvatska jezična riznica = Croatian Language Repository. Accessed at: http://riznica.ihjj.hr/philologic/. (10 May 2017)

Hrvatski mrežni korpus: hrWac = Croatian web corpus. Accessed at: http://nl.ijs.si/noske/all.cgi/first_form?corpname=hrwac. (10 May 2017)

Kettnerová, V., Lopatková, M. & Bejček, E. (2012). The Syntax-Semantics Interface of Czech Verbs in the Valency Lexicon. In R. Vatvedt Fjeld & J.-M. Torjusen (eds.) *Proceedings of the 15th EURALEX International Congress.* Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo, pp. 434–443.

Kettnerová, V. & Lopatková, M. (2014). Reflexive Verbs in a Valency Lexicon: The Case of Czech Reflexive Morphemes. In A. Abel, Ch. Vettori, & N. Ralli (eds.) *Proceedings of the XVI EURALEX International Congress: The User in Focus, EURALEX 2014.* Bolzano: Institute for Specialised Communication and Multilingualism, pp. 1007–1022.

Kipper Schuler, K. (2005). *Verbnet: A Broad-Coverage, Comprehensive Verb Lexicon.* Doctoral dissertation, University of Pennsylvania.

Levin, B. (1993). *English Verb Classes and Alternations*. The University of Chicago Press, Chicago – London.

Lopatková, M. et al. (2006). *Valency Lexicon of Czech Verbs VALLEX 2.0*. ÚFAL Technical Report TR-2006-34. Available at: https://ufal.mff.cuni.cz/~lopatkova/literatura/06-TR-vallex-2.0.pdf.

Lyashevskaya, O. & Kashkin, E. (2011). FrameBank: a database of Russian lexical constructions. In *Communications in Computer and Information Science Springer*, Vol. 542. Springer Verlag: Berlin & Heidelberg, pp. 337–348.

Lyashevskaya, O. (2012). Dictionary of Valencies Meets Corpus Annotation: A Case of Russian FrameBank. In R. Vatvedt Fjeld & J.-M. Torjusen (eds.) *Proceedings of the 15th EURALEX International Congress*. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo, pp. 1023–1030.

Klosa, A. (2013). The lexicographic process (with special focus on online dictionaries). In R. F. Gouws et al. (eds.) *Dictionaries. An International Encyclopedia of Lexicography*. Supplementary Volume: *Recent Developments with Focus on Electronic and Computational Lexicography*. De Gruyter.

Mikelić Preradović, N. (2008). *Pristupi izradi strojnog tezaurusa za hrvatski jezik*, doctoral thesis, Faculty of Humanities and Social Sciences, University of Zagreb.

Mikelić Preradović, N. (2010). Semantic classification of verbs in CROVALLEX. In S. Lagakos at al. (eds.) *Recent Advances in Computer Engineering and Applications* 1. Harvard University, Cambridge, USA, pp. 53–59.

Moguš, M., Bratanić, M. & Tadić, M. (1999). *Hrvatski čestotni rječnik*. Zagreb: Zavod za lingvistiku Filozofskog fakulteta & Školska knjiga.

Oraić Rabušić, I. & Bošnjak Botica, T. (2016). Chorvatský vs český model valenčního popisu. In K. Skwarska & E. Kaczmarska (eds.) *Research of Verbal Valency in Slavic Languages in the Past and Present = Výzkum slovesné valence ve slovanských zemích včera a dnes*. Praha: Slovanský ústav AV ČR, v.v.i., pp. 305–318.

Palić, I. (2011). O glagolima koji vežu obvezatne adverbijalne dopune u bosanskome jeziku. *Suvremena lingvistika,* 37(72), pp. 201–217.

Przepiórkowski, A. et al. (2014). Extended phraseological information in a valence dictionary for NLP applications. In *Proceedings of the Workshop on Lexical and Grammatical Resources for Language Processing (LG-LP 2014)*. Dublin, Ireland, pp. 83–91.

Ruppenhofer, J. et al. (2016). FrameNet II: Extended Theory and Practice. https://framenet2.icsi.berkeley.edu/docs/r1.7/book.pdf (30 April 2017)

Samardžija, M. (1986). *Valentnost glagola u suvremenom hrvatskom književnom jeziku*. Doctoral thesis, Faculty of Humanities and Social Sciences, University of Zagreb.

Schumacher, H. et al. (2004). *VALBU – Valenzwörterbuch deutscher Verben*. Tübingen: Gunter Narr Verlag.

Silić, J. & Pranjković, I. (2005). *Gramatika hrvatskoga jezika za gimnazije i visoka učilišta*. Zagreb: Školska knjiga.

VDE (2004): *A Valency Dictionary of English*: *A Corpus-Based Analysis of the Complementation Patterns of English Verbs, Nouns and Adjectives.* Berlin–New York: Mouton de Gruyter.

VALLEX 3.0: *Valency Lexicon of Czech Verbs.* Accessed at: http://ufal.mff.cuni.cz/vallex. (30 April 2017).