

Siniša Runjaić

Matea Birtić

Ivana Brač

Institut za hrvatski jezik
Ulica Republike Austrije 16, HR-10000Zagreb
srunjaic@ihjj.hr
orcid.org/0000-0002-9177-150X
mbirtic@ihjj.hr
orcid.org/0000-0001-5927-782X
ibrac@ihjj.hr
orcid.org/0000-0002-3660-5285

THE SARGADA REPOSITORY FOR THE ARGUMENT/ /ADJUNCT DISTINCTION: A SYNTACTIC RESOURCE WITH A LEXICOGRAPHICAL BACKGROUND

SARGADA¹ is an online repository that contains sentences with ambiguous syntactic phrases regarding argument/adjunct status. Research on the difference between arguments and adjuncts and the creation of the repository grew out of *e-Glava*, an online valency dictionary of Croatian verbs (Birtić, Brač and Runjaić 2017). The repository is intended to include 130 verbs/lemmas that are categorized into 13 macro-groups based on a syntactic phrase with unclear syntactic status. To determine the status of a syntactic phrase, seven diagnostic tests were applied. The results of the application of the tests are quantified as percentages, following the claim that argumenthood is not a binary but a scalar category (cf. Foraker 2014). The creation of the SARGADA repository is an important step towards improving the valency dictionary as it provides valuable data that can assist in the description of verbs' valency patterns in Croatian. The objective of this paper is to outline the process of creating the repository and to explore the possibility of using the results presented in the repository for the development of the valency dictionary.

¹ This work has been fully supported by the Croatian Science Foundation under the project *Syntactic and Semantic Analysis of Arguments and Adjuncts in Croatian* (SARGADA, IP-2019-04-7896).

1. Introduction

Although creating various classic monolingual or multilingual dictionaries necessarily requires, along with lexicographic methodology, a substantial level of the authors' knowledge of certain grammatical phenomena of one or more languages,² the process of planning and organizing valency lexicons or dictionaries often places greater significance on the understanding of these grammatical phenomena, surpassing the importance of lexicographic theory and practices. Building valency lexicons or databases is an illustrative case of a closer connection between methodologies of lexicography and grammaticography (Lehmann and Maslova 2004). Valency lexicons are a type of resource that simultaneously employ lexicographic and grammaticographic methods,³ i.e. typically providing basic information on the grammatical properties of verb lemmas. More importantly, they provide information on the syntactic structure in which a verb appears as well as information on the semantic roles of its arguments.

Before any valency lexicographic project begins, extensive preparations and the formalization of a dictionary entry design are necessary. However, unexpected theoretical and practical challenges may arise during the creation of a dictionary. A notable example is *e-Glava*, an online valency dictionary (Birtić, Brač and Runjaić 2017), whose main challenge was accurately categorizing dependents as either arguments or adjuncts for certain verbs. This theoretical issue, addressed during the lexicographic work, led to the development of a research project *Syntactic and semantic analysis of arguments and adjuncts in Croatian* (SARGADA). The main objective of the SARGADA project was to establish criteria and tests (Forker 2014; Toivonen 2021) for distinguishing between arguments and adjuncts in the Croatian language and applying them to the creation of the syntactic repository SARGADA. This repository arose directly as a by-

² Most dictionaries include information on the grammatical properties of lemmas, such as their part of speech, inflectional forms, some basic syntactic patterns, etc.

³ Perhaps it would be more precise to say that it is about the relationship between metalexigraphy and metagrammaticology, as stated by Lehmann and Maslova (2004: 1857): “Metagrammaticography starts from two ends: on the one hand, it takes stock of successful grammars and distills from them the grammaticographic principles that they follow. On the other hand, it takes successful models of language description as a theoretical basis and deduces from them requirements for an adequate grammar. Grammaticography is related to investigation of grammar just as lexicography is related to lexicology (investigation of the lexicon).”

product of research on ambiguous syntactic phrases regarding their argument/adjunct status.

This paper aims to delineate the development process of the repository and emphasize how the results presented in the repository could be applied to enhance a future valency dictionary. In this paper, the introduction is followed by Section 2, which details the development of the SARGADA repository through various subsections. Subsection 2.1 covers the initial phase of developing the repository, including the planning, compiling of lemmas, and classification of lemmas. Subsection 2.2 presents the syntactic model for sentence tagging. The diagnostic tests that are included in the repository are discussed in Subsection 2.3. In Section 3, the technical details of the repository and the preliminary results of the diagnostic tests are presented. Section 4 provides a discussion of the possibility of connecting the presented records of verb lemmas with existing dictionaries, as well as the potential use of the methodology and the results of the diagnostic tests for subsequent phases in the development of the valency dictionary. Finally, the conclusion provides a summary of the paper and suggests some possible directions for future work.

2. Building the SARGADA repository

The practical aspect of the project includes data collection, corpus search, and building a database that contains ambiguous syntactic phrases regarding the argument/adjunct status. In the initial phase of designing the relational database structure for the SARGADA repository, we consulted online resources that analyse both the syntactic and semantic side of the verb's patterns, such as dependency treebanks (e.g., *Hamburg Dependency Treebank* (Foth et al. 2014)), valency lexicons (e.g., *T-PAS – Typed Predicate Argument Structure for Italian* (Jezek et al. 2014)), as well as *FrameNet* (Fillmore and Baker 2010) and *Verbnet* (Kipper Schuler 2005), etc. However, as previously mentioned, our primary focus centres on the classification of syntactic elements whose status with respect to the argument/adjunct distinction is unclear. We did not follow the solutions from previously mentioned resources directly, but it can be concluded that some solutions and knowledge are indirectly woven into the structure of the repository.

Nevertheless, certain findings from these sources are indirectly incorporated into the repository's structure. Conversely, our investigation of these resources did not reveal any comparable attempts. Therefore, the repository we present represents a potentially unique endeavour in creating a database that captures the results of testing the argument/adjunct distinction for a specific language, namely Croatian in our case.

Our approach was primarily grounded on dependency grammar(s) and the non-binary difference between arguments and adjuncts. We attempted to define criteria for distinguishing arguments and adjuncts that would be as theoretically neutral as possible, so we used tests from generative grammar in addition to tests from dependency grammars. Alongside these linguistic theories, we incorporated insights from Croatian (e.g., Barić et al. 1997; Silić and Pranjković 2007; Raguž 2010; Belaj and Tanacković Faletar 2017), Serbian (e.g., Piper et al. 2005, Mrazović 2009) and Bosnian (Jahić, Halilović and Palić 2004) grammar books. This approach made it possible to develop a comprehensive model that considers various linguistic perspectives and provides a more nuanced understanding of the topic under investigation.

2.1. List of verbs and macro-groups

Building upon the findings of the aforementioned linguistic theories, the initial phase in building the repository involved compiling a list of 130 Croatian verbs, which are accompanied by syntactically ambiguous phrases that can be either arguments or adjuncts. Further analysis revealed that certain verbs have distinct meanings that encompass different valency patterns. This led to the identification of 130 lemmas.⁴ This part of the preparatory work for the repository is comparable to the lemma selection method as one of the basic lexicographic methods.⁵ Consequently, the lemmas were classified into separate groups based

⁴ We continuously update the verb list as our research uncovers new verbs with ambiguous syntactic elements. In Birtić, Brač, and Runjaić (2022), we initially identified 111 lemmas and categorized them into 12 macro-groups. However, we have since expanded the verb list to include a new macro-group specifically for verbs with passive agents.

⁵ "Before one starts writing dictionary articles, a lemma candidate list must be drawn up... This list should be an open list so that during the process of dictionary production new lemmata can be entered..." (Schierholz 2015: 331).

on the ambiguous syntactic phrases present in their valency patterns. To facilitate the creation of the repository, macro-groups were established for sets of syntactically ambiguous phrases that occur with specific verbs. The verbs in the repository are classified into 13 macro-groups, which share a particular ambiguous phrase, as presented in Table 1. The macro-groups of verbs with place, goal, source, time, quantity, manner, cause and purpose adverbials and the verbs that these macro-groups contain were mainly extracted from Palić (2011) and Belaj and Tanacković Faletar (2017), and their classification is also motivated by classes of complements proposed by various German linguists (see for example Engel (2009)). The verbs with inner objects and sports verbs are from Birtić and Matas Ivanković (2009, 2010). The macro-groups of verbs with beneficiary and passive agents were inspired by Needham and Toivonen (2011). The macro-group of verbs with the dependents in the instrumental case was listed according to Brač (2018).

Table 1: Classification of the lemmas into the macro-groups

Macro-group	Verb
Verbs with place adverbials	<i>živjeti</i> ‘live’, <i>stanovati</i> ‘live’
Verbs with goal adverbials	<i>baciti</i> ‘throw’, <i>dosegnuti</i> ‘reach’
Verbs with source adverbials	<i>dolaziti</i> ‘come from’
Verbs with time adverbials	<i>trajati</i> ‘last’
Verbs with quantity adverbials	<i>stajati</i> ‘cost’, <i>prodati</i> ‘sell’
Verbs with manner adverbials	<i>ponašati se</i> ‘behave’
Verbs with cause adverbials	<i>proizlaziti</i> ‘result’
Verbs with purpose adverbials	<i>koristiti se</i> ‘use’, <i>služiti se</i> ‘use’
Verbs with NPs in the instrumental case	<i>mirisati</i> ‘smell’, <i>zračiti</i> ‘radiate’
Verbs with beneficiary	<i>ispeći</i> ‘bake’, <i>sašiti</i> ‘sew’
Verbs with inner objects	<i>sanjati</i> ‘dream’, <i>živjeti</i> ‘live’
Sport verbs	<i>trčati</i> ‘run’, <i>plivati</i> ‘swim’
Verbs with passive agents	<i>čitati</i> ‘read’, <i>graditi</i> ‘build’

2.2. Syntactic model for sentence tagging

To accurately tag the sentence elements and analyse the sentences in accordance with our research objectives, we established a simple syntactic model of 12 tags. These tags include 11 definite tags for various sentence elements: ‘*Argument_S*’ (argument; subject), ‘*Verb*’ (verb), ‘*Argument_DO*’ (argument; direct object), ‘*Argument_IO*’ (argument; indirect object), ‘*Argument_PP*’ (argument; prepo-

sitional phrase), ‘*Adverb*’ (adverb), ‘*Adjunct*’ (adjunct), ‘*Aux*’ (auxiliary verb), ‘*Reflex*’ (reflexive pronoun), ‘*Conj*’ (conjunction), and ‘*Neg*’ (negation), as well as a twelfth tag specifically designed for a syntactically ambiguous phrase that we are testing, labelled as ‘*Test*’. We decided to tag unquestionable arguments and adjuncts during the parsing process and test only ambiguous phrases. Due to the need for a simple application in the database, we had to adopt the most general approach and use a few operational definitions of the argument to tag those sentence elements.⁶

2.3. Testing and diagnostic tests

Although the various syntactic theories disagree on many things, there is one domain in which they agree – they all distinguish arguments and adjuncts. Definitions of arguments and adjuncts⁷ differ with regard to their formal apparatus, but they also agree on which phrases are problematic from the viewpoint of the argument/adjunct distinction. As for English, authors often cite several typical phrases that are problematic: beneficiary NPs, resultatives, passive by-phrase, etc. (see more in Toivonen 2016). Some of these constituents are problematic in Croatian too: beneficiary NPs, prepositional phrases for expressing the passive agents (*od*-phrases), prepositional phrases with the semantic role of the goal, and some NPs with the instrumental case. Our repository included all these groups.

⁶ For example, the sentence in which we examine the quantity adverbial with the verb *prodati* ‘sell’ is syntactically parsed in the repository as shown in (i) (repository tags highlighted):

(i) Dorian	je	prodao	dušu	vragu	za vječnu mladost.
Dorian _{NOM.SG}	AUX	sell	soul _{ACC.SG}	devil _{DAT.SG}	for eternal youth
Argument_S	Aux	Verb	Argument_DO	Argument_IO	Test

‘Dorian sold his soul to the devil for eternal youth.’

In other words, *Dorian* (a seller) is an unquestionable argument (subject), *dušu* ‘soul’ (selling object) is an unquestionable argument (direct object), *vragu* ‘devil’ (a buyer) is an unquestionable argument (indirect object), and we test the prepositional phrase *za vječnu mladost* ‘for eternal youth’ (quantity/price) to find whether it is an argument or an adjunct in this example.

⁷ For example, definitions such as “The arguments are the participants minimally involved in the activity or state expressed by the predicate” (Haegeman 1994: 44). According to McInerney (2022: 1) “... dependents of a given head H can be partitioned into two classes (...) ‘core elements’, the arguments of H, and (...) ‘peripheral’ elements, the adjuncts of H”. In the classical valency theory, the arguments and adjuncts are differentiated by their specific relation with a verb, arguments are dependent on the verb and are also required by the verb’s valency, while adjuncts are only dependent on the verb (see more in Birtić and Brač 2023).

This paper will introduce a set of diagnostic tests chosen as a tool for distinguishing between arguments and adjuncts in the repository. It is important to note that there is no agreement on which tests are the most useful for distinguishing between them (see, for example, Przepiórkowski 2016; Haspelmath 2014). Since different authors and different theories employ different tests, some tests are the same across different frameworks (e.g., the optionality test, the verb specificity test, the iterativity test), some are similar in different frameworks (the core participant test and the implication test; paraphrase with a proverb and the *do so* test), some are designed in specific theories (the dialogue test in functional generative description; the adjunct island test in generative grammar), and some tests are valid only for certain languages. In the repository, we use the omission, implication, substitution and replacement test from dependency grammar, the *do so* test and *this happened* test from both dependency and generative grammar, and the dialogue test from functional generative description. Some tests applied in the first phase of the research were later removed, as we concluded that they did not yield relevant results when testing Croatian phrases with ambiguous status. For every verb with an ambiguous phrase that we test, there are two examples from the *Croatian Web Corpora – HrWac* (Ljubešić and Klubička 2014) or *Croatian Language Repository – Riznica* (Ćavar and Brozović Rončević 2012). CQL queries are used to identify suitable examples, with minor adaptations to tailor the sentences for our specific purposes while retaining their essential meaning. Two researchers independently analyse the sentences, doing manual parsing and conducting the diagnostic tests. Each researcher analyses one sentence example for each verb, and, subsequently, researchers exchange complete descriptions, collaboratively annotating and agreeing upon the results for both sentences per verb meaning. The project’s principal investigator then reviews both sets of annotations. Each test can result in ‘*Arg*’ (argument), ‘*Adj*’ (adjunct), or ‘*Not*’ (‘test not used’),⁸ aiming to provide results that are not simply binary but scalar (Forker 2014).

⁸ The tag ‘*Not*’ (‘test not used’) is mostly chosen for the *do so* test and *this happened* test with stative verbs, which is explained in Section 2.3.3 and Section 2.3.4.

2.3.1. Omission test (*OptT*)

The omission test, also known as the optionality test (Schütze 1995; Needham and Toivonen 2011), is a widely used test to distinguish between obligatory and non-obligatory elements in a sentence. If a syntactic phrase is essential for the sentence to be grammatical, the omitted part is considered an obligatory argument (1). On the other hand, if the syntactic phrase can be omitted without affecting the grammaticality of the sentence, it is either an optional argument or an adjunct (2). This test allows us to identify accurately obligatory arguments; however, it cannot distinguish between optional arguments and adjuncts.

(1)

*Moj rođak boravi *(u Chicagu).*
 my cousin is-staying in Chicago_{LOC.SG}
 ‘My cousin is staying *(in Chicago).’

(2)

Dječak baca kamenje (u vodu).
 boy throws stones into water_{ACC.SG}
 ‘A boy throws stones (into the water).’

In example (1), the prepositional phrase *u Chicagu* ‘in Chicago’ cannot be omitted and so it is an obligatory argument, while in example (2) we can omit *u vodu* ‘into the water’, and thus this prepositional phrase is either an optional argument or an adjunct. To define the status of the prepositional phrase in (2), the next test is used.

2.3.2. Implication test (*ImplT*)

The implication test, also referred to as the core participant test (Needham and Toivonen 2011), is a test used to identify arguments in a sentence based on verb semantics. If a verb presupposes the existence of an entity, then that entity is classified as an argument. An entity’s presence in the semantic structure of a verb can be indicated by a pronoun or an adverb (3). The implication relation is also tested by the procedure in which the examined argument has to be denied and the sentence must bring a contradiction (4).

(3)

Dječak baca kamenje nekamo.
 boy throws stones somewhere
 ‘A boy throws stones somewhere.’

(4)

#Dječak baca kamenje nekamo, ali ne baca nikamo.
 boy throw stones somewhere but_{NEG} throw nowhere
 ‘#A boy throws stones somewhere, but does not throw anywhere.’

The implication test has been found to be an effective method for identifying optional arguments, but its use is not necessary for determining obligatory arguments, as they can be identified through the omission test. Nonetheless, in the SARGADA repository, both the implication test and the omission test are utilized since we do not follow the dependency grammar model completely.

2.3.3. *Do so test (DoSoT)*

The *do so* test substitutes a non-stative verb and its arguments with the phrase *do so*, while elements that appear outside the *do so* replacement are considered adjuncts since they do not belong to the nuclear VP (Lakoff and Ross 1976; Ackema 2015). As a result of the test’s application, direct and indirect objects, directional adverbs, and affected locations are considered part of the verb phrase, while other adverbials, as well as instruments, are outside the nuclear verb phrase. There are several problems with the application of this test (cf. Przepiórkowski 2016). Firstly, it can only be applied to agentive verbs. Secondly, it is challenging to choose the appropriate Croatian equivalent for the English verb *to do*. In Croatian, *do so* can be translated as *(u)činiti isto*, *postupiti/postupati isto*, *postupiti/postupati tako* and other similar phrases. The verb *(u)činiti* requires the direct object in the accusative case and the indirect object in the dative case, while the verb *postupiti/postupati* requires the prepositional phrase *s* ‘with’ + instrumental, i.e., it does not take a direct object. Therefore, the results of the test may vary depending on the similar argument structure between the chosen pro-verb and the tested verb. Despite these limitations, the *do so* test is still used in the repository as a tool for identifying arguments.

(5)

Dječak baca kamenje u vodu, a djevojčica čini isto u zrak.
 boy throws stones into water_{ACC.SG} and girl does so into air_{ACC.SG}
 ‘A boy throws stones into the water, and a girl does so into the air.’

2.3.4. *This happened test (ThisHT)*

The *this happened* test (Brown and Miller 1991: 90) states that if a sentence can be paraphrased as two separate sentences, whereby one contains a verb and its arguments and the other contains the tested constituent then the syntactic phrase in the second sentence is an adjunct. With the *this happened* test, the tested syntactic phrase is shifted to the second sentence and introduced with the phrase *this happened*. If the second sentence is well formed, the tested phrase is an adjunct. If the sentence is not well formed, the tested phrase is an argument. Example (6) cannot be paraphrased in two sentences; therefore, *u vodu* ‘into the water’ is an argument. The *this happened* test cannot be applied to stative verbs.⁹

(6)

*Dječak baca kamenje. *To se dogodilo u vodu.*
 boy throws stones this_{REFL} happened into water
 ‘A boy throws stones. *This happened into the water.’

2.3.5. *Replacement test (ReplT)*

The replacement test is based on the assumption that the verb governs the morphological form of an argument, i.e., the precise choice of the case and the preposition, and therefore the argument cannot be expressed by a different morphological form, whereas an adjunct can. For instance, in example (7), the phrase with the preposition *u* ‘into’ and the noun in the accusative can be replaced with the prepositional phrase with the preposition *na* ‘onto’ and the noun in the accusative case, as well as the prepositional phrase with the preposition *preko* ‘over’

⁹ The problem with this test is that the verb *dogoditi se / događati se* in example (6) requires the prepositional phrase with a noun expressed in the locative case, while the verb *bacati* ‘throw’ requires the prepositional phrase with a noun in the accusative case.

and the noun in the genitive case. Therefore, the prepositional phrase *u vodu* ‘into the water’ is an adjunct according to this test.

(7)

Dječak baca kamenje u vodu / na krov / preko kuće.
 boy throws stones into water_{ACC.SG} / onto roof_{ACC.SG} / over house_{GEN.SG}
 ‘A boy throws stones into the water / onto the roof / over the house.’

The efficacy of the replacement test is limited when it comes to adverbial arguments, as they do not follow a consistent morphological pattern, yet they are obligatory elements.

2.3.6. Substitution test (*SubT*)

The substitution test refers to a verb’s specificity. This test determines whether a syntactic phrase is an adjunct by examining if a different verb can be substituted for the original verb without altering the grammaticality of the sentence (Ágel 2000; Šojat 2008). It is important to note that the same syntactic phrase can be an argument in one context but not in another. In the example (8), it can be seen that the verb can be substituted with other verbs within the same semantic class, while it cannot be substituted with verbs from a different semantic class. Furthermore, when dealing with verbs that have three complements, it is challenging to determine the extent to which the acceptability of the sentence with different verbs depends on the compatibility between the verbs and the second complement (*kamen* ‘stone’ in (8)) versus the compatibility between the verb and the third complement (*u vodu* ‘into the water’ in (8)).

(8)

*Dječak je bacio / gurnuo / zavitlao / *razveselio se / *pojeo kamen u vodu.*
 boy_{AUX} threw / pushed / swirled / cheered_{REFL} / ate stone into water
 ‘A boy threw / pushed / swirled / *cheered / *ate a stone into the water.’

2.3.7. *Dialogue test (DialT)*

The dialogue test originated from the framework of the functional generative description (Sgall, Hajičová and Panevová 1986). This test helps to determine if certain participants in an event are semantic arguments by asking a question about them. In (9), the answer cannot be *I don't know* since the verb *ležati* 'lie' requires an obligatory argument. Even if an optional argument is not explicitly stated in a sentence, the information about it should still be present in the speaker's mind (if it is an argument). However, the answer *I don't know* is more acceptable with the three-place verbs, such as the verb *bacati/baciti* 'throw' (10).

(9)

A: *Dinara leži.

*'Dinara lies.'

B: Gdje Dinara leži?

'Where does Dinara lie?'

A: #Ne znam.

#I don't know.'

(10)

A: Dječak baca kamenje.

'A boy throws stones.'

B: Kamo ih baca?

'Where does he throw them?'

A: ?Ne znam.

'?I don't know.'

According to researchers in German dependency models, such as Engel (2009), the most dependable tests for argument identification are the omission and implication tests. While the omission test certainly extracts obligatory arguments, the implication test further singles out optional arguments. As Engel (2009) has stated, the element which does not pass the implication test is never an argument, and the element which is negative in the omission test is always an argument. The omission test, *do so* test, replacement test, and substitution test mostly test syntactic arguments, while the implication test, *this happened* test, and dialogue

test are claimed to test semantic arguments. The argument that can be omitted (passes the omission test) is an optional argument. Therefore, there is a clear correlation between the semantic/syntactic identification of arguments and their obligatory vs. optional status. However, the determination of whether an argument can be omitted or not may pose challenges, as judgments may vary among native speakers. Consequently, the distinction between obligatory and optional arguments has not been included in our repository. While the replacement test does not work with many adverbial complements due to their morphological variability, their non-omissionability indicates their argument status.

3. Technical details and preliminary results

The repository was grounded on the database whose building encompasses several steps during the four years of the project:

- designing a specific relational database in an SQL environment on a server infrastructure (year 1);
- creating a personalized input interface for the project members (year 2);
- setting up a stable working version with sentence descriptions, along with the relevant central content management system (CMS) (year 3);
- reviewing and organizing the data for public access, as well as adding additional user functions (ongoing, year 4).

The Ubuntu 18 server operating system with LAMP architecture (Linux, Apache, MySQL and PHP) was successfully installed and configured for the server infrastructure. Additionally, a subdomain <http://ihjj.hr/sargada> was created. The initial version of the database needed for the SARGADA repository was established and organized with all the necessary programs for the creation of basic models. To translate the presented linguistic model into a graphical interface, a Vue.js framework was utilized with *Javascript*. This enables convenient editing of the logical structure. The data were structured and displayed based on the design with the use of HTML markup language, while the visual user interface was set up using *Cascading Style Sheets* (CSS) as per the project member's instructions (as shown in Figure 1).

Izaberi glagol

Odaberi makroskupinu

Primjer ✕

Odaberi dio Argument_S ✕ Verb ✕ Test ✕

Moj rođak

boravi

u Chicagu.

Dodaj
Izmjeni primjer

Figure 1: Interface for editing the sentence and testing the ambiguous element in the stable working version of the repository.¹⁰

As already stated, in the final project year, intensive processing, revision and organization of data in the database is planned. So far, we have entered some prototypical examples of verbs with adverbial complements. Specifically, these examples include verbs with place, source, manner, purpose, cause, and goal adverbials, as shown in Figure 2.

Macro-group	No	Example	Tests						
			OptT	ImplT	DoSoT	ThisHT	RepIT	SubT	DialT
Verbs with place adverbial	11	Moj rođak boravi u Chicagu. 'My cousin is staying in Chicago.'	Arg	Arg	Not	Not	Adj	Adj	Arg
Verbs with place adverbial	12	Dinara leži na granici između Hrvatske i Bosne i Hercegovine. 'Dinara is lying on the border between Croatia and Bosnia and Herzegovina.'	Arg	Arg	Not	Not	Adj	Adj	Arg
Verbs with source adverbial	13	Moj pas dolazi iz njemačke uzgajivačnice. 'My dog comes from a German kennel house.'	Arg	Arg	Not	Arg	Adj	Arg	Arg
Verbs with manner adverbial	14	On djeluje ozbiljno. 'He seems serious.'	Arg	Arg	Not	Not	Adj	Arg	Arg
Verbs with manner adverbial	15	Dže se (oni) vrlo dobro. '(They) are holding up very well.'	Arg	Arg	Not	Not	Adj	Adj	Arg
Verbs with manner adverbial	16	Odnose se (oni) prema nama s prijezrom. '(They) treat us with contempt.'	Arg	Arg	Not	Not	Adj	Arg	Arg
Verbs with manner adverbial	17	On se ponaša nepristojno. 'He is behaving rudely.'	Arg	Arg	Not	Not	Adj	Arg	Arg
Verbs with manner adverbial	18	Ubojca je prošao nekažnjeno. 'The killer went unpunished.'	Adj	Arg	Not	Not	Adj	Arg	Adj
Verbs with purpose adverbial	19	Plan smo namijenili (mi) za cijelu Gajevu ulicu. 'The plan was intended (by us) for the entire Gajeva Street.'	Arg	Arg	Not	Not	Adj	Arg	Arg
Verbs with purpose adverbial	20	Politeretno vozilo služi za prijevoz stvari. 'The semi-truck vehicle is used for transportation of goods.'	Arg	Arg	Not	Not	Adj	Arg	Arg
Verbs with cause adverbial	21	Ljudsko obilježje proisteklo je iz pokreta. 'The human form arose from the body movements.'	Arg	Arg	Arg	Arg	Arg	Arg	Arg
Verbs with cause adverbial	22	Neuspjeh proizlazi iz nerada. 'Failure comes from inaction.'	Arg	Arg	Not	Not	Adj	Arg	Arg
Verbs with goal adverbial	23	Brat je bacio kamen u vodu. 'The brother threw a stone into the water.'	Adj	Arg	Adj	Arg	Adj	Arg	Adj

Figure 2: Extracted examples of the tests' results with the verbs with place, source, manner, purpose, cause, and goal adverbials.

¹⁰ Figure 1 shows the following database entry fields:

- a) *Izaberi glagol* 'Select a verb' in which a verb is selected from a pre-prepared drop-down list with 130 lemmas (see section 2.1);
- b) *Odaberi makroskupinu* 'Select a macro-group', in which a macro-group is selected from a pre-prepared drop-down list with 13 macro-groups (see section 2.1);
- c) *Primjer* 'Example', in which sentence elements are entered element by element according to the chosen syntactic model for sentence tagging (see section 2.2).

Although the majority of investigated syntactic phrases with verbs from Figure 2 are undoubtedly arguments, we used them to “test” the tests. Consequently, the results of the application of the diagnostic tests in the repository confirmed that all the examined syntactic phrases are arguments but with varying percentages. The omission test demonstrated that only the third argument of the verb *bacati* ‘throw’ can be omitted (in Figure 2, example 23), and therefore it is not an obligatory argument.¹¹ The implication test indicated that syntactic phrases with all verbs are arguments. The *do so* test and *this happened* test are not applied to stative verbs, and the *do so* test only to agentive verbs as explained in Subsection 2.3.3 and 2.3.4 (in Figure 2, examples 11 – 20 and 22). If the verb *bacati* ‘throw’ is replaced with *učiniti isto* ‘do so’, its third complement can be repeated in the second clause, and therefore it is an adjunct according to this test (in Figure 2, example 23). The *this happened* test showed that the prepositional complements with the verbs *dolaziti* ‘come from’ (in Figure 2, example 13), *proisteći* ‘arise’ (in Figure 2, example 21) and *bacati* ‘throw’ (in Figure 2, example 23) are arguments, which is consistent with findings in other languages as they are directional adverbials. According to the replacement test, all prepositional phrases are adjuncts since directional or place adverbials can be expressed with different prepositional phrases. The result of the application of the dialogue test demonstrated that the syntactic phrases with all analysed verbs, except the verb *bacati* ‘throw’, are arguments since the information about the participant that they denote has to be present in the speaker’s mind. Regarding the verbs with manner, purpose and cause adverbial phrases, the results are quite similar. According to the omission test, the implication test and the dialogue test, the investigated adverbial syntactic phrases are arguments. However, according to the replacement test, it turns out that they are adjuncts, except for the verb *proisteći* ‘arise’. Furthermore, according to the substitution test, the tested syntactic phrase with the verb *držati se* ‘behave’ is an adjunct, whereas with other verbs it is an argument.

In summary, the SARGADA repository provides valuable insights into the valency description of various verbs and their syntactic phrases, helping to deter-

¹¹ The problem is that the application of tests in the repository allows for a binary choice between arguments and adjuncts, and there is no option for an optional argument, which is the prepositional phrase with the verb *bacati/baciti* ‘throw’.

mine the argument status of these phrases through the application of diagnostic tests.

4. Can the analysis of data from the repository help in future work?

By testing the behaviour of prototypical adverbial complements with specific verbs regarding their semantic obligatoriness or morphological interchangeability, we can compare those valency patterns with previous dictionary entries using the analysis results. In this way, the SARGADA repository may serve as a tool for improving our comprehension of Croatian syntactic phenomena in both theoretical and applied research.

As an example of a possible improvement for future linguistic description, the verb *služiti* ‘serve’ can be used. This verb is polysemous in general lexicographic descriptions and is used as a transitive, intransitive and reflexive verb. Example (11) (also example 20 in Figure 2) in the repository serves to examine the status of the prepositional phrase *za prijevoz stvari* ‘for transportation of goods’ in the verb’s meaning ‘to be usable for what’ in an intransitive construction.

(11a)

<i>Poluteretno vozilo</i>	<i>služi</i>	<i>za</i>	<i>prijevoz</i>	<i>stvari.</i>
semi-truck vehicle _{NOM.SG}	serve _{3SG}	for	transportation _{ACC.SG}	things _{GEN.PL}

‘The semi-truck vehicle is used for the transportation of goods.’

When conducting the seven diagnostic tests (as shown in Figure 3), only the replacement test indicated that the tested prepositional phrase could be an adjunct, while the other four (applicable)¹² tests yielded results indicating it to be an argument.

¹² The *do so* test and *this happened* test were not applicable because the tested sentence element follows the stative verb *služiti* ‘serve’.

OptT:

Arg Adj Not

ImpT:

Arg Adj Not

DoSoT:

Arg Adj Not

ThishT:

Arg Adj Not

ReplT:

Arg Adj Not

SubT:

Arg Adj Not

DialT:

Arg Adj Not

Rezultati:

Arg = 4/5 (80%)

Adj = 1/5 (20%)

Figure 3: Screenshot of the diagnostic test results for the prepositional phrase with the verb *služiti* ‘serve’ in the SARGADA repository

The outcome of the replacement test is due to the morphological interchangeability of the tested prepositional phrase *za* ‘for’ + *prijevoz*_{ACC.SG} ‘transportation’ with the noun phrase in the dative case (*prijevozu*_{DAT.SG}); therefore, the sentence remains grammatical, as demonstrated in (11b).

(11b)

Poluteretno vozilo *služi* *prijevozu* *stvari.*
 semi-truck vehicle_{NOM.SG} serve_{3SG} transportation_{DAT.SG} thing_{GEN.PL}
 ‘The semi-truck vehicle is used for the transportation of goods.’

During the lexicographic work on the valency dictionary *e-Glava*, in addition to morphological and syntactic realizations, a semantic description was also pro-

vided. Accordingly, we tried to perform a semantic analysis in this case too. Although, usually, a beneficiary semantic role is linked to the dative case, in this case the semantic role of purpose can be realized either by the dative case or by a prepositional phrase which consists of *za* ‘for’ and the accusative NP. It is important to note that the appearance of the dative does not change the semantic role. However, the beneficiary in the refined realization of the same sentence would also be expressed by the dative NP *čovjeku* ‘person_{DAT.SG}’ as in (11c). Therefore, in sentence (11c) we have both semantic roles realized: the purpose is expressed as the prepositional phrase with the preposition *za* ‘for’ and the accusative noun phrase, while the beneficiary is expressed as the dative NP.

(11c)

Poluteretno vozilo služi čovjeku za prijevoz stvari.
 semi-truck vehicle_{NOM.SG} serve_{3SG} man_{DAT.SG} for transport_{ACC.SG} thing_{GEN.PL}

‘The semi-truck vehicle serves a person to transport the goods.’

If the beneficiary in the dative case is added, it is not possible to replace the prepositional phrase *za* ‘for’ + *prijevoz* ‘transport_{ACC.SG}’ with the dative NP *prijevozu* ‘transport_{DAT.SG}’ (11d). Hence, if both roles are realized, the purpose can only be expressed with the prepositional phrase. While one might initially attribute the ungrammaticality of this sentence to the presence of two successive dative arguments (one with a benefactive and the other with a purpose role) we suppose that a more in-depth investigation of this issue is necessary. This is because instances of two successive arguments marked by the same case but with different semantic roles are not unrecorded in the Croatian language.¹³

(11d)

**Poluteretno vozilo služi čovjeku prijevozu stvari.*
 semi-truck vehicle_{NOM.SG} serve_{3SG} man_{DAT.SG} transport_{DAT.SG} thing_{GEN.PL}

After the analysis, we can conclude that for this intransitive realization of the verb *služiti* ‘serve’ with this specific meaning, we expect a syntactic realization with semantic arguments of theme, beneficiary and purpose. In Croatian, the argument with the beneficiary may be syntactically omitted, but the sentence is still grammatical if the adverbial complement realizes the semantic role of the

¹³ The well-known case is two successive accusative objects of the verb *poučiti/poučavati* ‘to teach’ (*poučavati učenika matematiku* ‘to teach the student_{ACC.SG} mathematics_{ACC.SG}’).

purpose with the prepositional phrase *za* ‘for’ + accusative case (11a). Example (11b) is also grammatical if the adverbial prepositional phrase is replaced with a dative NP which also expresses the purpose. The beneficiary role is inherently predictable from the very meaning of the verb and the comparison to the syntactic realization in (11c).

5. Conclusion

This paper presents the SARGADA repository, an online resource specifically designed for distinguishing between arguments and adjuncts. As has been stated, the need for the repository arose from work on the online Croatian verb valency dictionary *e-Glava* and it aims to enhance work on *e-Glava*. Seven tests were applied to sentence examples in the repository to distinguish arguments from adjuncts. These examples were categorized into 13 macro-groups based on their syntactically ambiguous elements. Since the syntactic elements tested in the repository can be classified as “in-between arguments and adjuncts”, a scalar approach was employed to differentiate between them. In this paper, the preliminary results of the application of the diagnostic tests to the macro-groups with place, source, goal, manner, purpose, and cause adverbials are presented. The results of the diagnostic tests show the potential of the repository to improve valency description, as demonstrated through the detailed analysis of the verb *služiti* ‘serve’. Furthermore, the repository plays a significant role in Croatian syntax studies, as it provides a comprehensive description of syntactic phrases that had not been previously explored in such depth. Although the tests may yield contradictory results, which is expected in a scalar approach, and the acceptability judgments of sentences may vary among native speakers, our goal was to establish relatively applicable criteria and tests for distinguishing between arguments and adjuncts. This project has the potential to advance our understanding of the argument/adjunct distinction not only in the Croatian language but also cross-linguistically. The freely accessible online sentence repository will prove valuable in various domains, including in the teaching and learning of Croatian, as well as in the enhancement of natural language processing tools. Future work will focus more on incorporating semantic roles and expanding the number of

verbs/lemmas included in the repository, which will further improve its applicability and potential uses.

Abbreviations

ACC	accusative
Adj	adjunct
Arg	argument
AU	auxiliary
DA	dative
DialT	dialogue test
DoSoT	<i>do so</i> test
GEN	genitive
ImplT	implication test
INS	instrumental
LOC	locative
NEG	negation
OptT	omission/optimality test
PL	plural
REFL	reflexive
ReplT	replacement test
SG	singular
SubT	substitution test
ThisHT	<i>this happened</i> test

References

- ACKEMA, PETER. 2015. Arguments and Adjuncts. *Syntax – Theory and Analysis: An International Handbook: vol. 1*. Eds. Kiss, Tibor; Alexiadou, Artemis. De Gruyter Mouton. Berlin. 246–273.
- BARIC, EUGENIJA ET AL. 1997. *Hrvatska gramatika*. Školska knjiga. Zagreb.
- BELAJ, BRANIMIR; TANACKOVIĆ FALETAR, GORAN. 2017. *Kognitivna gramatika hrvatskoga jezika. Knjiga druga. Sintaksa jednostavne rečenice*. Disput. Zagreb.
- BIRTIĆ, MATEA; BRAČ, IVANA; RUNJAIĆ, SINIŠA. 2017. The Main Features of the e-Glava Online Valency Dictionary. *Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference*. Eds. Kosem, Iztok; Tiberius, Carole; Jakubiček, Miloš et al. Lexical Computing CZ s.r.o. Brno. 43–62.
- BIRTIĆ, MATEA; BRAČ, IVANA; RUNJAIĆ, SINIŠA. 2022. Croatian Repository for the Argument/Adjunct Distinction – SARGADA. *Proceedings of the Fifth International Conference Computational Linguistics in Bulgaria (CLIB 2022)*. Ed. Koeva, Svetla. Bulgarian Academy of Sciences. Sofia. 225–234.
- BIRTIĆ, MATEA; BRAČ, IVANA. 2023. Tests for distinguishing arguments and adjuncts and their application to Croatian locative phrases. *The Verbal Kaleidoscope: Perspectives on the Syntax and Semantics of Verbs*. Eds. Batinić Angster, Mia; Angster, Marco. Sveučilište u Zadru – Morepress. Zadar. Forthcoming.
- BIRTIĆ, MATEA; MATAS IVANKOVIĆ, IVANA. 2009. Akuzativne dopune uz neprijelazne glagole: što su unutrašnji objekti? *Rasprave Instituta za hrvatski jezik i jezikoslovlje* 35. 1–19.
- BIRTIĆ, MATEA; MATAS IVANKOVIĆ, IVANA. 2010. Jesu li akuzativne dopune uz sportske glagole objekti? *Sintaksa padeža*. Eds. Birtić, Matea; Brozović Rončević, Dunja. Institut za hrvatski jezik i jezikoslovlje – Filozofski fakultet Sveučilišta Josipa Jurja Strossmayera u Osijeku. Zagreb. 35–55.
- BRAČ, IVANA. 2018. *Instrumental u hrvatskom jeziku*. Institut za hrvatski jezik i jezikoslovlje. Zagreb.
- BROWN, KEITH; MILLER, JIM. 1991. *Syntax: A Linguistic Introduction to Sentence Structure*. Routledge. London.
- ČAVAR, DAMIR; BROZOVIĆ, RONČEVIĆ, DUNJA. 2012. Riznica: The Croatian Language Corpus. *Prace filologične* 63: 51–65.
- ENGEL, ULRICH. 2009. *Syntax der deutschen Gegenwartssprache*. Erich Schmidt Verlag. Berlin.
- FILLMORE, CHARLES J.; BAKER, COLLIN F. 2010. A Frames Approach to Semantic Analysis. *The Oxford Handbook of Linguistic Analysis*. Eds. Heine, Bernd; Narrog, Heiko. Oxford University Press. Oxford, UK – New York, NY. 313–340.

- FORKER, DIANE. 2014. A Canonical Approach to the Argument/Adjunct Distinction. *Linguistic Discovery* 12/2. 27–40.
- FOTH, KILIAN A. ET AL. 2014. Because Size Does Matter: The Hamburg Dependency Treebank. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Eds. Calzolari, Nicoletta; Choukri, Khalid; Declerck, Thierry et al. European Language Resources Association (ELRA). Reykjavik. 2326–2333.
- HAEGEMAN, LILIANE. 1994. *Introduction to Government and Binding Theory*. Blackwell. Oxford.
- HASPELMATH, MARTIN. 2014. Arguments and Adjuncts as Language-Particular Syntactic Categories and as Comparative Concepts. *Linguistic Discovery* 12/2. 3–11.
- JAHIĆ, DŽEVAD; HALILOVIĆ, SENAHID; PALIĆ, ISMAIL. 2004. *Gramatika bosanskoga jezika za srednje škole*. Dom štampe. Zenica.
- JEZEK, ELISABETTA ET AL. 2014. T-PAS; A Resource of Typed Predicate Argument Structures for Linguistic Analysis and Semantic Processing. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Eds. Calzolari, Nicoletta; Choukri, Khalid; Declerck, Thierry et al. European Language Resources Association (ELRA). Reykjavik. 890–895.
- KIPPER SCHULER, KARIN. 2005. *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. Philadelphia: University of Pennsylvania.
- LAKOFF, GEORGE; ROSS, JOHN ROBERT. 1976. Why You Can't Do So into the Sink. *Syntax and Semantics, Volume 7: Notes from the Linguistic Underground*. Ed. McCawley, James D. Academic Press. New York. 101–131.
- LEHMANN, CHRISTIAN; MASLOVA, ELENA. 2004. Grammaticography. *Morphologie. Ein Handbuch zur Flexion und Wortbildung. 2. Halbband*. Eds. Booij, Geert E.; Lehmann, Christian; Mugdan, Joachim et al. De Gruyter. Berlin – New York. 1857–1882.
- LJUBEŠIĆ, NIKOLA; KLUBIČKA, FILIP. 2014. {bs,hr,sr}WaC — Web corpora of Bosnian, Croatian and Serbian. *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*. Eds. Bildhauer, Felix; Schäfer, Roland. Association for Computational Linguistics. Gothenburg. 29–35.
- MCINNERNEY, ANDREW. 2022. *The Argument/Adjunct Distinction and the Structure of Prepositional Phrases*. Dissertation. The University of Michigan.
- MRAZOVIĆ, PAVICA. 2009. *Gramatika srpskog jezika za strance*. Izdavačka knjižarnica Zorana Stojanovića. Sremski Karlovci – Novi Sad.
- NEEDHAM, STEPHANIE; TOIVONEN, IDA. 2011. Derived Arguments. *Proceedings of the LFG11 Conference*. Eds. Butt, Miriam; Holloway King, Tracy. CSLI Publications. Stanford, CA. <http://web.stanford.edu/group/cslipublications/cslipublications/LFG/16/papers/lfg11needhamtoivonen.pdf> (pristupljeno 14. travnja 2022.).

- PALIĆ, ISMAIL. 2011. O glagolima koji vežu obvezatne adverbijalne dopune u bosansko-mrežnom jeziku. *Suvremena lingvistika* 37/72. 201–217.
- PIPER, PREDRAG ET AL. 2005. *Sintaksa savremenoga srpskog jezika. Prosta rečenica*. Institut za srpski jezik SANU – Beogradska knjiga – Matica srpska. Beograd.
- PRZEPIÓRKOWSKI, ADAM. 2016. How Not to Distinguish Arguments from Adjuncts in LFG. *Proceedings of the Joint Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*. Eds. Arnold, Doug; Butt, Miriam Jessica; Crysmann, Berthold et al. CSLI Publications. Stanford. 560–580.
- RAGUŽ, DRAGUTIN. 2010. *Gramatika hrvatskoga jezika*. Vlastito izdanje. Zagreb.
- SCHIERHOLZ, STEFAN J. 2015. Methods in Lexicography and Dictionary Research. *Lexikos* 25. 323–352.
- SCHÜTZE, CARSON T. 1995. PP Attachment and Argumenthood. *Papers on Language Processing and Acquisition. MIT Working Papers in Linguistics 26*. Eds. Schütze, Carson T.; Ganger, Jennifer B.; Broihier, Kevin. MIT. Cambridge. 95–151.
- SGALL, PETR; HAJIČOVÁ, EVA; PANEVOVÁ, JARMILA. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Reidel Publishing Company. Dordrecht – Boston – Lancaster – Tokyo.
- SILIĆ, JOSIP; PRANJKOVIĆ, IVO. 2007. *Gramatika hrvatskoga jezika za gimnazije i visoka učilišta*. Školska knjiga. Zagreb.
- ŠOJAT, KREŠIMIR. 2008. *Sintaktički i semantički opis glagolskih valencija u hrvatskom*. Doktorski rad. Filozofski fakultet Sveučilišta u Zagrebu. Zagreb. 277 pp.
- TOIVONEN, IDA. 2016. *Revisiting the argument-adjunct distinction*, Handout from talk at Stanford University.
- TOIVONEN, IDA. 2021. Arguments and adjuncts across levels. *Proceedings of the LFG'21 Conference. On-Line*. Eds. Butt, Miriam; Findlay, Jamie Y.; Toivonen, Ida. CSLI Publications. Stanford, CA. 306–331.

Repozitorij SARGADA za razlikovanje dopuna i dodataka: sintaktički resurs s leksikografskom podlogom

Sažetak

Rad na repozitoriju SARGADA uključuje opis rečenica koje sadrže dvojbene sintaktičke skupine s obzirom na status dopune ili dodatka. Stvaranje repozitorija potaknuto je radom na mrežnom valencijskom rječniku *e-Glavi* (Birtić, Brač i Runjaić 2017). Repozitorij sadržava 130 glagola koji su podijeljeni u 13 makrogrupina na temelju sintaktičke skupine s nejasnim sintaktičkim statusom (glagoli s priložnim skupinama mjesta, cilja, izvora, vremena, količine, načina, uzroka, svrhe, sredstva, glagoli s benefaktivom, glagoli s

unutarnjim objektom, sportski glagoli te glagoli s pasivnim agensom). Da bi se odredilo pripada li sintaktička skupina kategoriji dopuna ili dodataka, primijenjeno je sedam testova koji su najvećim dijelom preuzeti iz gramatike zavisnosti, zatim iz generativne gramatike te funkcijskoga generativnog opisa. U radu su opisani sljedeći testovi: 1. test ispuštanja, eliminacijski test ili test redukcije, 2. implikacijski test, 3. test *učiniti isto*, 4. test *to se dogodilo*, 5. test zamjenjivosti ili supstitucije drugim morfološkim oblikom, 6. test supstitucije glagola u okolini istoga rečeničnog dijela i 7. dijaloški test. Rezultati primjene testova kvantificirani su kao postoci, slijedeći tvrdnju da je argumentnost skalarna, a ne binarna kategorija (usp. Forker 2014). U radu su prikazani preliminarni rezultati primjene testova na makroskupine glagola s priložnim skupinama mjesta, cilja, izvora, vremena, količine, načina, uzroka i svrhe. Stvaranje repozitorija SARGADA poslužit će poboljšanju valencijskoga opisa jer pruža vrijedne podatke koji pomažu u identifikaciji valencijskih obrazaca glagola u hrvatskom jeziku. Cilj je rada opisati proces nastajanja repozitorija te opisati mogućnosti primjene rezultata nastalih tijekom izgradnje repozitorija na stvaranje valencijskoga opisa i valencijskoga rječnika hrvatskoga jezika.

Keywords: Croatian language; valency dictionary; syntax; argument/adjunct distinction; diagnostic tests; digital repository

Ključne riječi: valencijski rječnik; sintaksa; dopuna; dodatak; dijagnostički tekstovi; digitalni repozitorij