FIFTH
INTERNATIONAL
CONFERENCE

CLIB'22

COMPUTATIONAL LINGUISTICS IN BULGARIA CLIB 2022

8 – 9 September 2022

Sofia, Bulgaria

Organiser:

Department of Computational Linguistics
Institute for Bulgarian Language
Institute of Information and Communication Technologies
BULGARIAN ACADEMY OF SCIENCES

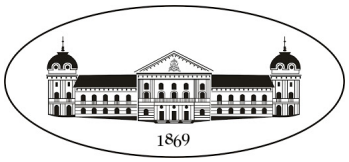
PROCEEDINGS

The Fifth International Conference *Computational Linguistics in Bulgaria* (CLIB 2022)is organised with the support of the National Science Fund of the Republic of Bulgaria under Grant Agreement No. KP-06-MNF/7 of 20.07.2022.

NATIONAL SCIENCE FUND
Ministry of Education and Science

The National Science Fund does not take responsibility for the contents of the papers presented at the Conference or for any of the Conference materials.

CLIB 2022 is organised by:



Department of Computational Linguistics
Institute for Bulgarian Language

Institute for Information and Communication Technologies

Bulgarian Academy of Sciences

PUBLICATION AND CATALOGUING INFORMATION

Title:	Proceedings of the Fifth International Conference *Computational Linguistics in Bulgaria* (CLIB 2022)
ISSN:	2367 5675 (online)
Published and distributed:	Bulgarian Academy of Sciences
Editorial address:	Institute for Bulgarian Language Bulgarian Academy of Sciences 52 Shipchenski Prohod Blvd., Bldg. 17 Sofia 1113, Bulgaria +3592/ 872 23 02

Proceedings of the

Fifth International Conference

Computational Linguistics in Bulgaria

8 – 9 September 2022
Sofia, Bulgaria

PROGRAMME COMMITTEE

Chair:

Svetla Koeva – Institute for Bulgarian Language, Bulgarian Academy of Sciences

Co-chair:

Petya Osenova – Institute of Information and Communication Technologies, Department of Linguistic Modelling and Knowledge Processing, Bulgarian Academy of Sciences / Sofia University, Faculty of Slavic Studies

Iana Atanassova – University of Burgundy, Centre for Interdisciplinary and Transcultural Research, France

Verginica Barbu Mititelu – Research Institute for Artificial Intelligence, Romanian Academy

Svetla Boytcheva – Institute of Information and Communication Technologies, Department of Linguistic Modelling and Knowledge Processing, Bulgarian Academy of Sciences

Khalid Choukri – Evaluations and Language Resources Distribution Agency, France

Ivan Derzhanski – Institute of Mathematics and Informatics, Bulgarian Academy of Sciences

Tsvetana Dimitrova – Institute for Bulgarian Language, Department of Computational Linguistics, Bulgarian Academy of Sciences

A. Seza Doğruöz – Ghent University, Belgium

Radovan Garabík – Ľudovít Štúr Institute of Linguistics, Slovak Academy of Sciences

Maria Gavrilidou – Institute for Language and Speech Processing, Natural Language Processing and Knowledge Extraction Department, Greece

Stefan Gerdjikov – Sofia University, Faculty of Mathematics and Informatics, Bulgaria

Voula Giouli – Institute for Language and Speech Processing, ATHENA Research Centre, Greece

Ivan Koychev – Sofia University, Faculty of Mathematics and Informatics, Bulgaria

Cvetana Krstev – University of Belgrade, Faculty of Philology, Serbia

Eric Laporte – University of Paris-Est Marne-la-Vallée, France

Natalia Loukachevitch – Research Computing Center of Moscow State University, Russia

John P. McCrae – National University of Ireland, Galway, Ireland

Preslav Nakov – Qatar Computing Research Institute, HBKU, Qatar

Maciej Piasecki – Wrocław University of Technology, Poland

Vito Pirrelli – Institute for Computational Linguistics, ILC-CNR, Italy

Ewa Rudnicka – Wrocław University of Technology, Poland

Ivelina Stoyanova – Institute for Bulgarian Language, Department of Computational Linguistics, Bulgarian Academy of Sciences

Stan Szpakowicz – University of Ottawa, Canada

Marko Tadić – University of Zagreb, Faculty of Humanities and Social Sciences, Department of Linguistics, Croatia

Hristo Tanev – Joint Research Centre of the European Commission, Italy

Irina Temnikova – Big Data for Smart Society Institute (GATE), Bulgaria

ORGANISING COMMITTEE

Croatian repository for the argument/adjunct distinction – SARGADA

Matea Birtić
Institute of Croatian Language and Linguistics
mbirtic@ihjj.hr

Ivana Brač
Institute of Croatian Language and Linguistics
ibrac@ihjj.hr

Siniša Runjaić
Institute of Croatian Language and Linguistics
srunjaic@ihjj.hr

Abstract

The distinction between arguments and adjuncts is a relevant topic in many linguistic theories (Tesnière, 1959; Chomsky, 1981; Langacker, 1987; Van Valin, 2001; Herbst, 2014, etc.). Even though theories provide similar definitions of arguments and adjuncts, sometimes it is difficult to draw a clear line between them. In order to determine ambiguous syntactic parts as arguments or adjuncts, various tests have been proposed, but they often give contradictory results and are not fully reliable. Nevertheless, they can be used as an auxiliary tool. The project *Syntactic and Semantic Analysis of Arguments and Adjuncts in Croatian* – SARGADA was launched with the aim of thoroughly investigating the distinction between arguments and adjuncts in Croatian, and to apply the theoretical results in a syntactic repository which would be a valuable resource for improving NLP tools and for researching and teaching Croatian.

In this paper, we will present diagnostic tests chosen as a tool to distinguish between arguments and adjuncts in the Croatian language. The repository containing sentences with ambiguous syntactic phrases and our workflow will also be described.

Keywords: Croatian language, syntax, argument/adjunct distinction, diagnostic tests, digital repository.

1 Introduction

Many linguistic theories (Tesnière, 1959; Bresnan, 1982; Chomsky, 1981; Langacker, 1987, Van Valin, 2001, etc.) distinguish between arguments (complements) and adjuncts as two separate grammatical categories, defining arguments as (semantically) obligatory, selected by a specific verb, and necessary to understand the event expressed by the verb (*Peter fixes the car.*) and adjuncts as optional, not selected by a specific verb, and not necessary for understanding the event expressed by the verb (*Peter fixes the car.in the yard.*). Although the opposition is sometimes considered to be binary, most theories nowadays differentiate between obligatory and optional arguments, therefore operating with three distinct categories of non-predicate elements (obligatory arguments, optional arguments, and adjuncts).

Although a classification into arguments (complements) and adjuncts is made in almost all grammatical theories, it is rather difficult to draw a clear line between them (e.g., Vater, 1978; Schütze, 1995; Müller, 1996; Koenig, Mauner, and Bienvenue, 2003). Furthermore, quite a large number of various tests has been proposed to distinguish between arguments and adjuncts, which often yields controversial results. The project *Syntactic and Semantic Analysis of Arguments and Adjuncts in Croatian* – SARGADA, financed by the Croatian Science Foundation, was launched with the aim of clearly and precisely investigating the criteria for the definition and delimitation of arguments and adjuncts in the Croatian language, and to apply the theoretical results in a syntactic repository serving as a valuable resource for improving

natural language processing tools and for researching and teaching the Croatian language.

Since many theoretical approaches deal with distinguishing between argument and adjunct, we decided to conduct a thorough analysis of arguments and adjuncts and the criteria for their delimitation from the viewpoint of traditional Croatian grammars and three contemporary linguistic theories: valency theory and dependency grammar, generative grammar, and cognitive grammar. Combining three different linguistic theories is methodologically justified by the theoretical demands this project seeks to answer: (1) which criteria and tests are suitable to define and extract arguments and adjuncts in Croatian; (2) is the established distinction between arguments and adjuncts grammatically tenable; (3) could the distinction between arguments and adjuncts be defined independently of theory?

In this paper, in Section 3, we offer an answer to the first question by presenting diagnostic tests chosen to distinguish between argument and adjunct. In Section 4, we present the repository that contains sentences with ambiguous syntactic parts regarding the distinction of argument and adjunct. Section 5 concludes the paper.

2 Diagnostic tests

As has already been stated, there is no consensus on which tests should be used to distinguish arguments and adjuncts. In this paper, we will present tests chosen as a tool for distinguishing arguments and adjuncts in the repository. Dependency grammar uses, for example, the omission test, the implication test, the *do so* test, the paraphrase with dependent clause, and the *this happened* test. Generative grammar uses structure preservation/changeability after operation, the *do so* test, extraction from *wh*-islands, iterativity, etc. Cognitive grammar uses the methodological principle of conceptual (in)dependence. In the repository, roughly speaking, the omission test, the implication test, the *this happened* test, and the substitution test are taken from dependency grammar; the *do so* test and extraction from *wh*-islands are taken from generative grammar; and the dialogue and iterativity test come from functional generative description. A few other tests were considered, but it was decided not to include them because they are not applicable to Croatian or not relevant (the dialogue test, paraphrase with a dependent clause, etc.).

2.1 Omission test

The omission test, also called the optionality test (Needham and Toivonen, 2011), the *Eliminierungs* test (Helbig and Schenkel, 1978), *Reduktionstest* (Engel, 2009[4]), etc., is a standard test to separate obligatory elements in a sentence from non-obligatory elements, i.e., optional arguments and adjuncts. If a syntactic phrase can be omitted, and the sentence remains grammatical, the omitted part is not an obligatory argument, but either an optional argument (1) or an adjunct (2). The problem is that some arguments can be omitted (e.g., with the verbs *eat*, *read*, *sing*) and some adjuncts are obligatory (e.g., some phrases in passive constructions). According to dependency grammar models, every obligatory phrase co-occurring with a specific verb is an argument.

(1) *Ivan jede pizzu.*
Ivan is-eating pizza.ACC.SG.
'Ivan is eating (pizza).'

(2) *On ide u crkvu (nedjeljom).*
he goes to church Sunday.INST.SG.
'He goes to church (on Sunday).'

2.2 Implication test

The implication test or *Folgerungs* test (Engel, 2009[4]) is also known as the *Core Participant Test* (Needham and Toivonen, 2011). The test relies on the semantics of verbs. According to this test, if a verb presupposes the appearance of an entity, then we are dealing with an argument.[1] The presence of a participant in the semantic structure of a verb can be signaled by a pronoun or an adverb (3) and the pronoun or adverb cannot be negated (4). The Croatian verb *boraviti* 'stay' always presupposes that there is a place where someone is staying. The verb's meaning cannot be realized without a "place".

(3) *On boravi negdje.*
he is-staying somewhere
'He is staying somewhere.'

(4) **On boravi negdje, ali*

[1] One of the reviewers observed that by implication test adjuncts would qualify as arguments since most concrete acts would imply a place which is commonly assumed to be an adjunct. What matters here is that we are talking about what the verb presupposes, not the action in general.

he is_staying somewhere but
negdje *ne* *postoji.*
somewhere NEG exists
'*He is staying somewhere, but somewhere does not exist.'

In dependency grammars, this procedure is called anaphorisation. The application of this test makes sense for the optional arguments, while it is not needed for the obligatory arguments since they are already indicated by the omission test.

2.3 *Do so* test

In order to prove that Chomsky's claim (1965: 95–106) that place and time adverbials are sister constituents of VP and can occur freely with any VP, while direction, duration, place, frequency, and some manner adverbials subcategorize the verb, Lakoff and Ross (1976) introduced the *do so* test. According to the *do so* test, a non-stative verb and its arguments may be substituted with *do so*, while elements that occur after *do so* are outside the nuclear VP and are adjuncts.[2] Thus, the direct object, indirect object, directional adverbs, and affected locations are inside the verb phrase, while other adverbials are outside the nuclear verb phrase. In example (5), a *trip* is an argument and *last Tuesday* is an adjunct.

(5) John took a trip last Tuesday, and I'm going to do so tomorrow.

In many studies (e.g. Przepiórkowski, 2016), it is shown that the test is not reliable, especially for instruments and some *with* phrases that are, according to this test, always adjuncts. The problem that we would like to point out lies in the translation, i.e., choosing the Croatian equivalent of the verb *do*. *Do so* can be translated into Croatian as '*činiti isto*', '*postupiti isto*', etc. If we apply this test to three-place verbs that originally take accusative and dative arguments, such as the verb *pružati* 'bring, give', and we replace it with the verb *činiti* that has the same valency pattern as the original verb *pružati* 'bring, give', it follows that the dative complement is an adjunct since it occurs after the pro-verb (6). But if we replace the verb *pružati* 'give' with the verb *postupiti*, which in this case has the prepositional phrase *s* 'with' + the instrumental as its argument, it follows that the dative is an argument (7). So, the results depend on the distributional properties of a pro-verb or its subcategorization.

(6) *Djeca* *pružaju* *utjehu*
children give comfort.ACC.SG
odraslima, *a* *odrasli*
adults.DAT.PL and adults.NOM.SG
to čine *djeci.*
it do children.DAT.SG
'Children give comfort to adults, and adults do so to children.'

(7) *Djeca* *pružaju* *utjehu*
children give comfort.ACC.SG
odraslima, *a* *odrasli*
adults.DAT.PL and adults.NOM.SG
**tako* *postupaju* *djeci.*
so do children.DAT.SG
'Children give comfort to adults, and adults do so to children.'

2.4 *This happened* test

According to the *this happened* test (Brown and Miller, 1991: 90), if a sentence can be paraphrased by two sentences, one contains a nuclear predication and the other an adverbial. Example (8) can be paraphrased by two sentences; therefore, *in the kitchen* is an adjunct, while *on the table* in (9) is an argument.

(8) Ivan se popeo na stol. To se dogodilo u kuhinji.
'John stood on the table. This happened in the kitchen.'
(9) *Ivan se popeo. To se dogodilo na stol.
'*John stood. This happened on the table.'

2.5 Replacement test

The replacement test, as we call it in our repository, or *Ersatzprobe* (Ágel, 2000: 180), targets the syntactic level and should differentiate arguments from adjuncts. It is connected with the assumption that the morphological form of an argument is dictated by a verb (10), while the morphological form of an adjunct is not (11).

(10) *On piše* *zadaću /* **zadaći*
he is-writing homework.ACC homework.Dat
/ *na zadaći.
on homework.LOC
'He is writing homework / *to homework / *on homework.'

(11) *On piše* *zadaću* *na stolu*
he is_writing homework.ACC on table.LOC
/ u kuhinji */ jučer.*
in kitchen.LOC yesterday
'He is writing homework on the table / in the kitchen / yesterday.'

[2] Although adjuncts can be included in *do so* repetition.

2.6 Substitution test

The substitution test or *Supklassentest* (Engel, 2009) examines verb specificity. If the verb can be replaced with another verb or verb form in the environment of the same syntactic phrase, then the phrase next to it is an adjunct (Ágel, 2000; Šojat, 2008). Authors have noted that the same syntactic phrases can be arguments in one case, but not in another. The given example shows that the examined verbs can be replaced by verbs from the same or related semantic class and they require the same arguments (12).

(12)
Brat je bacio / gurnuo / zavitlao
brother AUX threw pushed swirled
*/*razveselio se / *pojeo kamen u*
cheered REFL ate stone into
vodu.
water
'The brother threw / pushed / swirled /*cheered / *ate a stone into the water.'

This is closely connected with the notion of subcategorization in generative grammar or what is called *Subklassenspezifik* in the German dependency tradition, but is also widely used in traditional grammars. In dependency grammar, it is said that arguments are specific for a subclass of verbs, and therefore they are subclass specific. In the generative tradition, it is said that the verb is subcategorized for its arguments. The test is not reliable for adverbial arguments since they are not uniform in their morphological form, but are still obligatory.

2.7 Extraction from *wh*-islands

According to generative grammar, islands are parts of sentences from which it is difficult to extract phrases. There are strong islands, from which nothing can be extracted, and weak islands, from which some phrases can be extracted. The traditional assumption, going back to Huang (1982) and Chomsky (1986), is that arguments can be extracted from weak islands, but adjuncts and subjects cannot. The extraction of arguments from weak islands is better than the extraction of adjuncts and subjects, but is still not considered completely acceptable.

(13a) Marko piše zadaću na stolu.
'Marko is writing homework on the table.'
(13b) *Gdje se Marija pita piše li Marko zadaću?
'Where does Mary wonder if he writes homework?

(14a) Marko popravlja auto.
'Marko fixes the car.'
(14b) ?Što se Marija pita poravlja li Marko?
'What does Mary wonder if Marko has fixed?'

According to Chomsky (1986) and Huang (1982), (14b) is supposed to be better than the example in (13b), which is true according to our intuition. In (13b) the adjunct phrase is extracted from a *wh*-island (indirect question with *li*-particle), and in (14b) the argument phrase is extracted from a weak island.

A legitimate question in this context, which has to be further investigated, is what counts as a weak island in Croatian. For instance, are indirect questions with the particle *li* really weak islands in Croatian, or do we have to find another context that will be a better context for sorting out arguments? There is also a long-standing question in linguistic literature about whether extraction from a weak island is truly sensitive to the argument/adjunct distinction or to some other linguistic property (Miliorini 2019).

2.8 Iterativity test

According to the iterativity test, adjuncts can be iterated freely, while arguments cannot (15) (Bresnan, 1982; Forker, 2014). However, on closer inspection, adjuncts can be iterated only if they refer to the same phenomenon or entity with a different degree of precision (Verspoor, 1997: 66; Brunson, 1993: 14), as shown in the example from Verspoor (1997: 66) (16) and its translation into Croatian (16b). The problem is that iteration is often possible for arguments as well (17).

(15) *John escaped from prison with dynamite with a machine gun.
(16a) Sam kicked a ball in the morning at 10 o'clock.
(16b) Sam je udario loptu ujutro u 10 sati.
(17) On se žalio na susjedu, na njezino ponašanje.
'He complained about the neighbor, about her behavior.'

According to Przepiorkowski (2016) and Bresnan (1982), time, location, and manner can occur with any verb and can be iterated, but instruments cannot.

3 Repository

The applied part of the project includes the gathering of data, a corpus search, and creating a database for the description of ambiguous phrases regarding argument/adjunct status. In this paper, the current state of the repository's development

after two years of the four-year project is presented.

During the planning of the development of the relational database structure for the SARGADA repository, we consulted online resources in which conceptual solutions for the repository could be found. Linguistic information resources, in which the syntactic and semantic level of sentence parts are processed, can be roughly divided into several categories based on selected linguistic methodologies and schools:

a) Syntactically parsed and morphosyntactically marked parts of general or specialized corpora of texts; e.g. numerous corpora via the *Sketch engine* platform (Kilgarriff et al., 2014).
b) Dependency treebanks, as exclusively syntactic resources in the narrowest sense; e.g. *The Hamburg Dependency Treebank* (Foth et al., 2014), *Dependency Treebank for Czech* (Hajič et al., 2018).
c) Valency lexicons, i.e. syntactic resources in a broader sense, created as the result of general linguistic or national projects; e.g. *ValPal – Leipzig Valency Classes Project* (Hartmann, Haspelmath, and Taylor, 2013), *T-PAS – Typed Predicate Argument Structure for Italian* (Jezek et al., 2014).
d) Lexical databases with elaborated systems for marking semantic frames; e.g. *Framenet* (Fillmore and Baker, 2010), *Verbnet* (Kipper Schuler, 2005).

The SARGADA repository with its conceptual basis and as a digital resource of a specific part that directly arises as a by-product of syntactic research of ambiguous syntactic parts does not belong to those categories and therefore does not have a specific model. Another important distinguishing feature of the SARGADA repository concerning the studied resources is that the goal of its development is not to include already prepared linguistic data according to an unambiguous theoretical idea, but quite the reverse. This repository should examine new linguistic data about less researched syntactic categories of arguments and adjuncts for the Croatian language.

When compiling the model, we mostly followed dependency grammar due to the notion of the non-binary determination of the distinction between arguments and adjuncts. Notions about arguments and adjuncts from generative grammar will serve as an additional control during the process of examining individual examples. In parallel with the study of these linguistic theories, the traditional grammar of Croatian, Serbian and Bosnian was consulted, as well as the works of prominent South Slavic syntacticians who, directly or indirectly, touch on the topic of arguments and adjuncts.

3.1 Workflow

Following the previously mentioned theories and analyzed data in the literature, in the first phase of preparation for the repository, a list of verbs was compiled. The list includes 111 Croatian verbs which are accompanied by ambiguous sentence parts that can be either arguments or adjuncts. After deeper analysis, we found that some of these verbs have different meanings that involve various valency patterns, so we are actually operating with 111 lemmas. Therefore, we decided to classify the lemmas into separate groups according to the ambiguous sentence part that appears in their valency patterns. For the purpose of creating the repository, these groups of syntactically ambiguous parts that occur with certain verbs have defined so-called "macrogroups" (groups of verbs that co-occur with the same ambiguous part). The verbs in the repository are classified according to these macrogroups, and we have singled out 12 groups.[3]

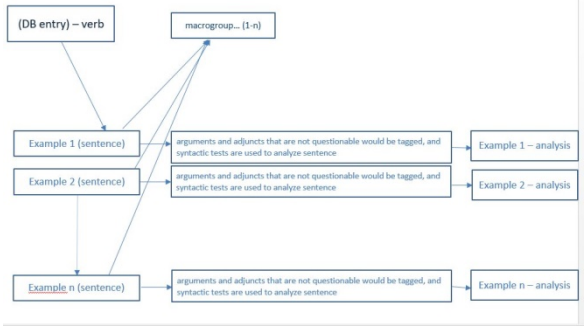


Figure 1. Schematic representation of the database organization.

[3] 1. verbs with place adverbials (e.g. *živjeti* 'live'); 2. verbs with goal adverbials (e.g. *baciti* 'throw'); 3. verbs with source adverbials (e.g. *dolaziti* 'come from'); 4. verbs with time adverbials (e.g. *trajati* 'last'); 5. verbs with quantity adverbials (verbs of exchange of goods and money, e.g. *stajati* 'cost'); 6. verbs with manner adverbials (e.g. *ponašati se* 'behave'); 7. verbs with cause adverbials (e.g. *proizlaziti* 'result'); 8. verbs with purpose adverbials (e.g. *koristiti se* 'use'); 9. verbs with instrumental case (e.g. *mirisati* 'smell'); 10. verbs with benefactive dative case (e.g. *ispeći* 'bake'), 11. verbs with inner objects (e.g. *sanjati* 'dream'); 12. sport verbs (e.g. *trčati* 'run').

The workflow can be divided into a few steps, also shown in Figure 1:

1. Lemma input.
2. Selected sentence as an example for syntactic testing.[4]
3. Linking the selected example to a macrogroup.[5]
4. Tagging sentence parts in the example that are not ambiguous in terms of distinguishing arguments and adjuncts[6] (manual parsing).
5. Determining the sentence part that will be analyzed in the example by syntactic tests for argument/adjunct distinction.

When the sentence part for which the argument/adjunct distinction has to be tested is determined, the selected tests are performed outside the information system, and the outcomes of the tests are recorded in the database for each of them. Every test can give three possible results: '*Arg*' (argument), '*Adj*' (adjunct), or '*Not*' ('test not used'). In this way, we seek to present results that are not binary but scalar.[7] In theory, it is possible that, for a particular example, all tests will give the answer '*Arg*', and then the system will show that the particular ambiguous sentence part is, without any doubt, an argument. The same, of course, applies to the adjunct. However, for most examples, different tests are expected to show different results that can be expressed as a scalar value and then graphically displayed after final processing. This would fulfil the applied part of the project in accordance with the work plan, and we believe that in this way the completed analysis would be more potent for the further development of research on the distinction between arguments and adjuncts in the Croatian language.

3.2 Technical information on the current stage of the development of the repository

The server infrastructure has been set up, i.e., an *Ubuntu 18* server operating system with LAMP architecture (Linux, Apache, MySQL and PHP) has been configured and installed, and a subdomain `http://sargada.jezik.hr` has been opened. The first (changeable) version of the database for the needs of the SARGADA repository was created and structured, and all the necessary programs for the development of basic models were installed. The presented linguistic model has been translated into a graphical interface using the *Javascript* language, i.e., the *Vue.js framework*, which enables flexible editing of the logical structure. The mark-up language HTML was used to structure and display the data according to the design, and the visual user interface was described and set up using *Cascading Style Sheets* (CSS) according to the instructions of the project members (as shown in Figure 2).

[4] Examples are collected from Croatian linguistic literature, by translating cognate examples from international linguistic literature, and some of the examples were gathered by researchers during their investigations.

[5] In theory, the number of sentences or examples for one meaning of a verb and for one macrogroup is unlimited (*n*), but it was decided in advance that one or two illustrative examples would be processed for each macrogroup for one meaning of a verb. On the other hand, we have already emphasized that due to polysemy, several macrogroups can be processed for each lemma.

[6] These are the following sentence parts: *Argument_S* (argument_subject), *Verb* (verb), *Argument_DO* (argument_direct object), *Argument_IO* (argument_indirect object), *Argument_PP* (argument_prepositional phrase), *Adjunct* (adjunct), *Aux* (auxiliary verb), *Reflex* (reflexive), *Conj* (conjunction) and *TEST* (ambiguous part for testing). We decided to tag unquestionable arguments and adjuncts during the parsing process and test only ambiguous sentence parts.

[7] One of the reviewers has brought to our attention that the scalar approach is not actually useful for the organization of the lexical or grammatical database. In a lexical or grammatical database, it is essential to define each element by some binary feature. Thinking exhaustively about the problem, we understand the reviewer's point of view, and we will reconsider our approach by giving a definite opinion on the status of some complements as arguments or adjuncts for further manipulation of the data. However, we think that the scalar approach is adequate as an illustration of our research, and it will give appropriate insights for teachers, students and researchers of the Croatian language.

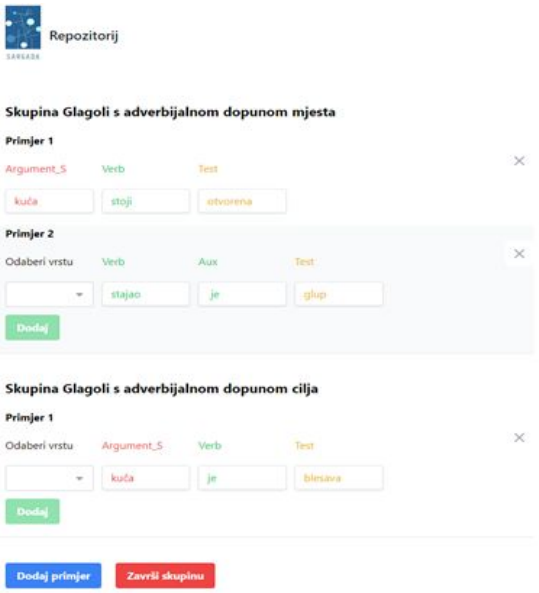


Figure 2. The current version of the SARGADA repository user interface.

The development of a central data management system (CMS) for users (project members) continues, so they will soon enter, edit and control linguistic data through this user interface. Currently, the PHP code is being developed and, through it, this input system will communicate with the configured database and save the structured data according to linguistic settings. When this code is completed, a stable (full-length version) will be prepared for entering data. Simultaneously, the database, back-end system and central management system will be tested based on these user actions. After all the data has been entered and harmonized, a graphic template will be designed for interaction with external users. This will allow for the creation of a visible system (front end) for online publishing and searching on the Internet, which would fulfil the work plan on the applied part of the SARGADA project.[8]

4 Conclusion

The paper presents the theoretical and applied part of the SARGADA project. The approach to distinguishing between argument and adjunct is presented in the first part of the paper. Arguments are separated from adjuncts based on eight tests mostly taken over from dependency grammar and to a lesser degree from generative grammar. The tests are applied to sentence examples in the repository. The sentence examples are sorted according to their characteristic ambiguous part into 12 macrogroups. Since the ambiguous sentence parts examined in our project are "in-between arguments and adjuncts", we decided to employ a gradual approach to distinguishing between argument and adjunct and to present scalar data.[9] The current state of the infrastructure of the digital repository SARGADA, which emerges as a product of work on the distinctions between arguments and adjuncts in these sentences, is also presented. The biggest gain of the parallel working process is that the need to create an applied digital resource prompted the creation of a methodology by which the tested results of theoretical research should be expressed at a scalar rather than a binary level. However, even greater added value is the fact that the process of transposing the linguistic model into the structure of the database and user interface spurred additional project tasks and produced results that were not even conceived at the initial stage of the project.

This project is important for a better understanding of the argument/adjunct distinction both cross-linguistically and with regard to Croatian and cognate languages. In addition, our research is also important for Croatian studies since the examined syntactic phrases had not previously been exhaustively described and their status was not unambiguously solved within Croatian linguistic literature. The repository of sentences that is freely available online will be of use in several segments of society (a tool for teaching and studying Croatian, or for improving natural language processing tools).

Acknowledgments

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 (2019–04–7896).

[8] Online publishing on the Internet would be the minimum goal of creating a repository, and the added value would be, for example, the development of an application programming interface (API) of the SARGADA repository with other linguistic resources of the Institute of Croatian Language and Linguistics (or other research groups).

[9] See footnote 7.

References

Vilmos Àgel 2000. *Valenztheorie*. Gunther Narr Verlag, Tübingen.

Joan Bresnan. 1982. Polyadicity. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*. MIT Press, Cambridge, Massachusetts, pages 149–172.

Keith Brown and Jim Miller. 1991. *Syntax*: *A Linguistic Introduction to Sentence Structure*. Routledge, London.

Noam Chomsky. 1965. *Aspects of the Theory of Syntax*. The M.I.T. Press, Cambridge, Massachusetts.

Noam Chomsky. 1981. *Lectures on Government and Binding*: *The Pisa Lectures*. Foris Publications, Dordrecht.

Noam Chomsky. 1986. *Barriers*. The MIT Press, Cambridge – London.

Ulrich Engel. 2009. *Syntax der deutschen Gegenwartssprache*. Erich Schmidt Verlag, Berlin.

Charles J. Fillmore and Collin F. Baker. 2010. A Frames Approach to Semantic Analysis. In *The Oxford Handbook of Linguistic Analysis*. Oxford University Press, Oxford, UK/New York, New York.

Diana Forker. 2014. A Canonical Approach to the Argument/Adjunct Distinction. *Linguistic Discovery*, 12: 27–40.

Kilian Foth, Arne Köhn, Niels Beuck, and Wolfgang Menzel. 2014. Because Size Does Matter: The Hamburg Dependency Treebank. In *Proceedings of the Language Resources and Evaluation Conference 2014 / European Language Resources Association* (ELRA) (2014), eBook.

Jan Hajič et al. 2018. *Prague Dependency Treebank* 3.5. Institute of Formal and Applied Linguistics, LINDAT/CLARIN, Charles University, LINDAT/CLARIN. PID: http://hdl.handle.net/11234/1-2621.

Iren Hartmann, Martin Haspelmath, and Bradley Taylor, editors. 2013. *Valency Patterns Leipzig*. Max Planck Institute for Evolutionary Anthropology, Leipzig. http://valpal.info.

Gerhard Helbig and Wolfgang Schenkel. 1983. *Wörterbuch zur Valenz und Distribution deutscher Verben*. 7. Aufl. Niemeyer, Tübingen.

Thomas Herbst. 2014. The Valency Approach to Argument Constructions. In Thomas Herbst, Hans-Jörg, and Susen Falhaber, editors, *Constructions*. *Collocations*. *Patterns*. De Gruyter Mouton, Berlin – Boston, pages 167–216.

Cheng-Teh James Huang. 1982. *Logical Relations in Chinese and the Theory of Grammar*. Doctoral dissertation, MIT.

Elisabetta Jezek, Bernardo Magnini, Anna Feltracco, Alessia Bianchini, and Octavian Popescu. 2014. T-PAS; A Resource of Typed Predicate Argument Structures for Linguistic Analysis and Semantic Processing. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation* (LREC'14), pages 890–895, Reykjavik, Iceland. European Language Resources Association (ELRA).

Adam Kilgarriff, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, and Vít Suchomel. 2014. The Sketch Engine: Ten Years On. *Lexicography*, 1: 7–36.

Karin Kipper Schuler. 2005. *VerbNet*: *A Broad-Coverage, Comprehensive Verb Lexicon*. Dissertations available from ProQuest. AAI3179808.

Jean-Pierre Koenig, Gail Mauner, and Breton Bienvenue. 2003. Arguments for Adjuncts. *Cognition* 89, 67–103.

George Lakoff and John Robert Ross. 1976. Why You Can't Do So into the Sink. In James D. McCawley, editor, *Syntax and Semantics*, *Volume 7*: *Notes from the Linguistic Underground*. Academic Press, New York, 101–131.

Ronald W. Langacker. 1987. *Foundations of Cognitive Grammar*. *Volume I*: *Theoretical Prerequisites*. Stanford University Press, Stanford.

Rafaela Miliorini. 2019. Extraction from Weak Islands: Alternative to the Argument/Adjunct Distinction. *ReVEL* 17/16: 37–58.

Stefan Müller. 1996. *Complement Extraction Lexical Rules and Argument Attraction*. https://hpsg.fu-berlin.de/~stefan/Pub/case_celr.html.

Stephanie Needham and Ida Toivonen. 2011. *Derived Arguments*. web.stanford.edu/group/cslipublications/cslipublications/LFG/16/papers/lfg11needhamtoivonen.pdf.

Adam Przepiórkowski. 2016. How Not to Distinguish Arguments from Adjuncts in LFG. In *Proceedings of the Joint Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*, pages 560–580.

Carson T. Schütze. 1995. PP Attachment and Argumenthood. In Carson T. Schütze, Jennifer Ganger, and Kevin Broihier, editors, *Papers on Language Processing and Acquisition*. MIT Working Papers in Linguistics 26. MIT, Cambridge, pages 95–151.

Krešimir Šojat. 2008. *Sintaktički i semantički opis glagolskih valencija u hrvatskom*. PhD dissertation, University of Zagreb, Zagreb.

Robert D. Jr. Van Valin. 2001. *An Introduction to Syntax*. Cambridge University Press, Cambridge.

Heinz Vater. 1978. On the Possibility of Distinguishing between Complements and Adjuncts. In Werner Abraham, editor, *Valence, Semantic Case and Grammatical Relations*. John Benjamins B. V., Amsterdam, pages 21–45.

Cornelia Maria Verspoor. 1997. *Contextually-Dependent Lexical Semantics*. PhD dissertation, University of Edinburg, Edinburg.