CLITIC CLIMBING, FINITENESS AND THE RAISING-CONTROL DISTINCTION. A CORPUS–BASED STUDY[1]

EDYTA JURKIEWICZ-ROHRBACHER [1,2] – BJÖRN HANSEN[1]
– ZRINKA KOLAKOVIĆ[1]
[1] Universität Regensburg, Germany
[2] University of Helsinki, Finland

Abstract: In the paper, we discuss the phenomenon of clitic climbing out of finite da_2-complements in contemporary Serbian. Scholars' opinions on the acceptability and occurrence of this construction, based on a handful of self-made examples, vary considerably. Expanding on the assumption that the correctness of the phenomenon has often been denied due to its rareness we employ large corpora to examine the problem. We focus on possible constraints arising from the syntactic properties of clause-embedding predicates.
Keywords: srWac, constraints on clitic climbing, semifinite complements, syntax, Serbian

1   INTRODUCTION

Following [16, p. 162], under the term CLITIC CLIMBING (CC) we understand "constructions in which the clitic is associated with a verb complex in a subordinate clause but is actually pronounced in constructions with a higher predicate [...]", as illustrated in the Serbian example:

(1)   *Gde*   ***nas***$_2$ *možete$_1$ naći$_2$?*

      where   us.ACC   can.2PRS   find.INF

      'Where can you find us?'   (srWaC v1.2)

In many languages, CC is only attested in complex clauses involving infinitives; cross-linguistically, CC out of complements with inflected verbs is a rare phenomenon.

In Serbian complement clauses the infinitive competes with the so-called *da*-complement, i.e. a verb marked for person and number which is introduced by an element usually treated as a complementizer as in (2):

---

(2) (...) *na*   *celoj*   *toj*   *teritoriji*   *ne*   *možete*[1]

    on   whole.LOC   that.LOC   territory.LOC   NEG   can.2PRS

    da   *nađete*[2]   *500*   *stanovnika.*

    COMP   find.2PRS   500   inhabitants.GEN   (srWaC v1.2)

    'On that whole territory you cannot find 500 inhabitants.'

However, it remains unclear up to what extent and under what circumstances CC out of *da*-complements is possible. The present paper approaches this problem empirically. Section 2 refers to the discussion on CC out of *da*-complements in Serbian and Section 3 introduces the Raising-Control distinction. Section 4 presents the main sources of data used in this study while Section 5 explains the data collection process and the difficulties it poses. Section 6 describes the results in detail, and is followed by the final Section 7, which draws conclusions from the main results and offers a suggestion for future research.

2   THE *DA*-COMPLEMENT AND CC IN SERBIAN

As the research on the syntax of Bosnian, Croatian and Serbian is divided into descriptive empirical studies on the one hand, and works with a formal theoretical orientation on the other, it comes as no surprise that in the literature we find largely contradictory statements concerning CC out of *da*-complements.

S. Stjepanović [17, p. 174ff] argues that *da*-complements and infinitival clauses allow CC in a similar way. However, discussing examples of CC out of *da*-complements, S. Stjepanović [17, p. 201] writes imprecisely that those "are acceptable sentences, however, they are short of perfect". Similarly, according to [5, p. 243], movement out of the finite complement is only "marginally possible". In contrast, D. Ćavar and C. Wilder [2, p. 41] and W. Browne [1, p. 41] argue that CC out of finite complements is strictly impossible. Finally, Lj. Progovac [13, p. 146] admits that "some speakers of Serbian" do not accept CC in the presented contexts. All the above-mentioned authors rely exclusively on self-constructed examples.

An early empirical work concerning CC is [11][2], who assumed that the variation in clitic positioning is closely related to the (at that time) new and increasing tendency to replace the infinitive with '*da* + Present tense'. He claimed that although ekavian Serbian speakers, who had already almost completely replaced the infinitive with '*da* + Present tense', preferred keeping the pronominal clitic directly after *da* (without CC), CC was common in journalistic texts published in Sarajevo [11].

It has long been known that *da*-complements do not behave in a uniform way. M. Ivić [7] proposes to distinguish two complement types headed by *da*: 'mobile present tense' and 'immobile present tense', the previous being regularly marked for tense and the latter not. This distinction goes back to [6], and was further elaborated on by W. Browne [1] and O. Mišeska-Tomić [12] who used the terms *da*[1]- and *da*[2]-complement. Based on [18], we assume that if CC is possible, this is so in the case of *da*[2]-complements. One hypothetical reason why some scholars reject the possibility

---

[2] He does not use the term clitic climbing.

of CC out of *da*-complements is its extreme rarity in comparison to equivalent constructions without CC. In our paper, we address the following research question:

Q1: To what extent is clitic climbing out of *da*$_2$-complements in Serbian possible?

3   CC AND THE RAISING-CONTROL DISTINCTION

If CC out of *da*$_2$-complements is possible, the question arises which syntactic features enable or block climbing. To start with, we investigate the potential link between CC and the Raising-Control Distinction, usually held to be crucial to categorizing different types of sentences with complement clauses.

Due to lack of space, we will confine ourselves to some basic empirical observations discussed in various theoretical frameworks. Roughly speaking, in raising constructions the subject does not receive its thematic role directly from the matrix predicate but from the embedded predicate. In a control construction, in contrast, the matrix verb and the embedded verb each assign a subject thematic role; therefore, there are two syntactic arguments present: the surface subject and the non-overt infinitival null subject PRO. W. Davies and S. Dubinsky [4, pp. 4–8] propose relatively robust, cross-linguistically applicable tests to distinguish raising from control constructions: i) the argument of the matrix predicate takes over the theta role of the argument of the embedded predicate, ii) the argument of the matrix predicate takes over the selectional restrictions:

(3) *Veliki*   *vizionar*   *može*   *da*   *donese*
big.NOM   visionaire.NOM   Can.3PRS   COMP   bring.3PRS
najvredniji   *trofej*   *u*   *Ukrajinu.*
most valuable.ACC   trophy.ACC   in   Ukraine.ACC   (srWaC v1.2)
'The great visionary can bring the most valuable trophy to Ukraine.'   Raising

(4) *Velike*   *sile*   *pokušale*   *su*   *da* PRO
big.NOM.F   forces.NOM.F   try.PTCP.PL.F   be.3PL   COMP
spreče   *taj*   *ustanak.*
stop.3PRS   that.ACC   rebellion.ACC   (srWaC v1.2)
'Big forces tried to stop that rebellion.'   Control

and iii) passivization does not change the meaning of the sentences:

(3') *Najvredniji*   *trofej*   *može*   *da*   *bude*
most valuable.NOM   trophy.NOM   can.3PRS   COMP   be.3SG
donesen   *u*   *Ukrajinu.*
bring.PASS.NOM.M   in   Ukraine.ACC
'The most valuable trophy can be brought to Ukraine (by the great visionary).'

(4') **Taj*   *ustanak*   *pokušao*   *je*   *da*
that.NOM.M   rebellion.NOM.M   try.PTCP.SG.M   be3SG   COMP

bude		sprečen.			
be.3SG		stop.PASS.NOM.M			

'Attempts were made to have that rebellion stopped (by big forces)'

A distinction should be made between subject and object control constructions. Whereas predicates that have only one individual argument besides the clausal argument are always subject control predicates, polyvalent predicates may show either subject or object control. According to [15, p. 412], verbs denoting directive speech acts (e.g. *zamoliti* 'to request') belong to the canonical class of object control predicates, and predicates that refer to commissive speech acts (e.g. *obećati* 'promise') are typical subject control predicates.

(5)	*Dekan*$_X$	*je*	*sve*	*prisutne*$_Y$	*zamolio*	*da* PRO$_Y$
	dean.NOM	be.3SG	all.ACC	present.ACC	ask.PTCP.SG.M	COMP
	kažu	*svoje*	*utiske (...)*			
	say.3PRS	their.ACC	impressions.ACC		(srWaC v1.2)	

'The Dean kindly asked the attending members to share their impressions (...).'

The point of departure of our study is divergent statements on the link between CC and the Raising-Control dichotomy in Czech. According to M. Rezac [14], CC is allowed out of infinitival complements of raising, subject-control, and object-control verbs alike. U. Junghanns [8] basically agrees with M. Rezac [14], but raises doubt as to the acceptability of CC with object control. Based on this debate, we formulate our second research question:

Q2: Does clitic climbing out of *da*-complements in Serbian depend on verb type with respect to the Raising-Control Distinction?

In order to approach the two research questions we examine the behaviour of CLs in relation to the type of clause embedding predicate (CEP). The most suitable method of exploring our research questions appears to be a corpus-based approach.

4　SOURCE OF DATA

The lack of resources for most of the South Slavic languages was recognized by the group of linguists behind the Regional Linguistic Data Initiative. Namely, only few electronically stored corpora of contemporary, original Serbian texts are easily accessible for research.

Since other relevant criteria for our study are size, morphosyntactic annotation, type and variety of texts, the number of relevant sources drops drastically. srWaC [9] seems currently to be the most suitable source for studies on rarely occurring phenomena[3].

---

[3] The only alternative is offered by the Corpus of Contemporary Serbian Langauge (version SrpKor2013) developed by Miloš Utvić and Duško Vitas, but it is five times smaller. The quality of metainformation and availability of search options are additional reasons why we excluded this corpus from the current study.

srWaC is the biggest corpus of Serbian. The current version 1.2 is a web corpus collected from .rs top-level domains containing nearly 555 million tokens. The corpus is automatically annotated with diacritic restoration, morphosyntax and lemma layers. The accuracy of morphosyntactic tagger performance has been evaluated at 92.33%, while part-of-speech tagger accuracy reached 97.86% in the tests [10]. Some imperfections of taggers can be identified through frequency list analysis and compensated for in the query formula.

The corpus is available for download, but it is also accessible via an on-line interface, NoSketchEngine, which offers a more convenient way for the linguistic community to search the corpus structures in comparison to self-written scripts. The on-line version provides a Corpus Query Language-based concordancer as well as many useful tools such as filtering or frequency lists.

Next to size, available meta-information and accessibility, a great advantage of srWaC is data variety. Analysis of url domain lists shows that not only does srWac cover texts typically included in corpora of standard language such as literary, journalistic and administrative texts, and academic and popular scientific texts: it is also a valuable source of less formal language appearing in user-generated content such as comments and fora.

Although the Internet is often criticized for poor quality of texts, which covers numerous spelling errors, omission of diacritic signs and non-standard use of upper and lower case, it is also a source of authentic, spontaneously produced written texts. Spatially unrestricted access to the Internet additionally gives some prospects for the study of regional differences, which otherwise might remain undiscovered.

The main drawback of srWaC, that is the lack of control for text-types and authorship, has not yet been solved, so some caution must be applied with regard to linguistic variation.

Another problem arises from the heterogeneity of text-types represented by a single domain, which is particularly important for separating narrative texts. Complementary literary texts, however, can easily be obtained from InterCorp [3]. The most recent, ninth version contains a section with original Serbian texts. Although it contains only 563 782 words, it comprises eight contemporary literary positions (all written after 1960), all in the core part of InterCorp. This implies that the annotation process has been manually revised. The corpus is accessible through the Kontext interface, which uses the same search engine as NoSketchEngine. The tagsets are identical in both corpora. Therefore, the same queries can be applied to both corpora, and InterCorp can be treated as a complementary source of literary data for our study.

5   QUERIES FOR THE DA$_2$-COMPLEMENT

5.1   Query Design

The corpus queries have to take into account four word-order patterns (CL—clitic, CEP—clause embedding predicate, *DA—da* complementizer, COMP—embedded finite complement):

1. CL CEP DA COMP
2. CEP CL DA COMP

3. CEP DA CL COMP
4. CEP DA COMP CL

1 & 2 are clear cases of CC, while in 3 & 4 no CC takes place. Each of the four elements in the pattern was defined in the query (see example query for pattern 3 in Figure 1) as a tag, word form or lemma, or a combination of these. This made it possible to exclude many ambiguities, such as the most often inaccurately lemmatized words. Instead of single CLs, we allowed clitic clusters comprising maximally four CLs.

In order to improve our recall, we allowed up to five empty positions between the core elements of the query. Thus, we were able to eliminate from those positions the core elements, as well as those expressions that refer to sentence clause crossing, i.e.: punctuation signs, accidentally attached dots, other verbs, conjunctions, participles and complementizers. We are, of course, aware that some markers can function within a sentence clause e.g. as connectors within a noun phrase (e.g. *Marko i Ana*), but due to a shortage in human resources, we could not afford excessive manual filtering.

[lemma="moći"][!(word="da"|tag="V.n"){tag="(Va[aefpm].*)|(Vc[fm].*)|(Vm.*)"&word!="ćeš")|tag="C.*"|tag="Z"|tag="P[iq].*"|word="(šta|sta|što|sto|kada?|gdj?e|kako|koliko|t?ko|onda|kada?|otkako|otkada?|odakle|dokle|pošto)"|tag="Rr"|word=".*\..*"|lemma="što"|word="(me)|([mj]u)|(joj)|(i[hm])|(ga)|([nv]as)"|(word="je"&tag!="V.*")|(word="[mt]i"&tag!="Pp[12]-pn")|(word="te"&tag!="(Pd-[fm][sp][ga])|(Cc)")]{0,5}[word="da"][!(word="(me)|([mj]u)|(joj)|(i[hm])|(ga)|([nv]as)")|(word="je"&tag!="V.*")|(word="[mt]i"&tag!="Pp[12]-pn")|(word="te"&tag!="(Pd-[fm][sp][ga])|(Cc)"|word="da"|tag="V.*"|tag="C.*"|tag="Z"|tag="P[iq].*"|word="(šta|sta|što|sto|kada?|gdj?e|kako|koliko|t?ko|onda|kada?|otkako|otkada?|odakle|dokle|pošto)"|tag="Rr"|word=".*\..*"|lemma="što")]{0,5}[tag="Vmr.*"&word!="(šta)|(gde)|([Vv]am)|(vise)|(takođe)"][!(word="da"|tag="V.*"|tag="C.*"|tag="Z"|tag="P[iq].*"|word="(šta|sta|što|sto|kada?|gdj?e|kako|koliko|t?ko|onda|kada?|otkako|otkada?|odakle|dokle|pošto)"|tag="Rr"|word=".*\..*"|lemma="što")]{0,5}[(word="(me)|([mj]u)|(joj)|(i[hm])|(ga)|([nv]as)")|(word="je"&tag!="V.*")|(word="[mt]i"&tag!="Pp[12]-pn")|(word="te"&tag!="(Pd-[fm][sp][ga])|(Cc)")]{1,4}[tag!="V.r.*"]within<s/>

Fig. 1. Query example for Pattern 3

5.2 Choice of Tested Verbs

It goes without saying that due to imperfect annotation and search limitations, not all instances of the patterns in question could be retrieved. Apart from the well-acknowledged problems of recall, also the precision of searches caused problems.

While the size of InterCorp allowed for one general query per pattern, and filtering out wrong examples manually was possible, in srWaC the final formula brought us results exceeding our human capacities, but not yielding any viable results.

We narrowed the search in srWaC to a selection of non-reflexive CEPs from the list of 42 CEPs retrieved from InterCorp (five of each type plus two additional object-control predicates, as we expected smaller frequencies in this type), which had appeared more than once and which allowed only da_2-complements. Additionally, we had to take into account differences in frequencies of particular

syntactic types of CEPs with the da_2-complement. Some raising verbs are particularly frequent, more so than for example object-control verbs. Therefore, we did not necessarily choose raising verbs with the highest frequencies, and we selected those belonging to different semantic types (modal and phasal verbs).

Finally, we tried to handle polyfunctionality in relation to syntactic type by eliminating verbs that due to their semantics can belong to different classes or appear with different *da*-complements. As such distinctions had not been annotated, they could not be easily distinguished on the query surface. This is why, for example, the verbs *(na)učiti* 'to learn/to teach', *znati* 'to know', *htjeti* 'to want/will', *morati* 'to must' and *trebati* 'to have to', *dati* 'to give' were excluded from the list of potential candidates.[4]

verb	frequency of da_2 in InterCorp	estimated frequency of da_2 in srWaC	relative frequency of CC out of da_2 in srWac[4]	syntactic type
moći 'can'	69	37526	0.0043	raising
nastaviti 'to continue'	2	1028	0.0029	raising
početi 'to start'	26	6546	0.0100	raising
prestati 'to stop'	5	1203	0.0116	raising
sm(j)eti 'to be allowed'	7	2027	0.0039	raising
nam(j)eravati 'to intend'	3	465	0.0021	subject control
nastojati 'to strive'	7	721	0.0027	subject control
pokušati 'to try'	7	4794	0.0047	subject control
um(j)eti 'to be able to'	8	1209	0.0016	subject control
usp(j)eti 'to succeed'	9	4331	0.0009	subject control

[4] Proportion of the estimated frequency of CC out of da_2-complements to the estimated frequency of all da_2-complements for the given CEP.

dozvoliti 'to allow'	7	2528	0	object control
narediti 'to order'	5	1174	0	object control
nat(j)erati 'to force'	2	502	0	object control
zamoliti 'to ask'	3	1584	0	object control
pustiti 'to let'	3	534	0	object control
primorati 'to force'	0	248	0	object control
pomoći 'to help'	0	331	0	object control

Tab. 1. Selected CEPs

The size and precision of results still posed processing problems. As no gold standards have been broadly acknowledged we decided to follow some suggestions by S. Wallis [19], and accordingly the precision of queries was estimated through sampling. We took random samples of 100, which usually should give no more than a 10% margin of error at a confidence level of 95% regardless of the population size. We calculated the binominal probability confidence interval using the Clopper-Pearson exact method. We recalculated raw frequencies into estimated frequencies on the basis of the worst-case scenario of the obtained confidence intervals. These are used in the analyses in the next section.

6   RESULTS AND DISCUSSION

6.1   Constraints on CC from da$_2$-complements into the Matrix Clause in Serbian
Although part of our data is based on the worst-case scenario, our material provides empirical evidence that CC out of *da$_2$*-complements into matrix clauses is indeed possible, although it is most likely a marginal phenomenon. In addition to two examples of CC into raising predicate clauses obtained from the Intercorp subcorpus, our samples yielded 69 correct sentences with CC, from which we estimated a worst-case scenario of 286 CC cases in the whole examined population. The frequencies of CC normalized to the frequency of a *da$_2$*-complement for a particular verb are summarized as part of Table 1 as well as in Figure 2. Analysis of frequencies shows that CC out of *da$_2$*-complements occurs with verbs of varying frequencies. The Chi-square test of dependence between the syntactic type and clitic climbing yields a significant result (p = 7.948e-11).
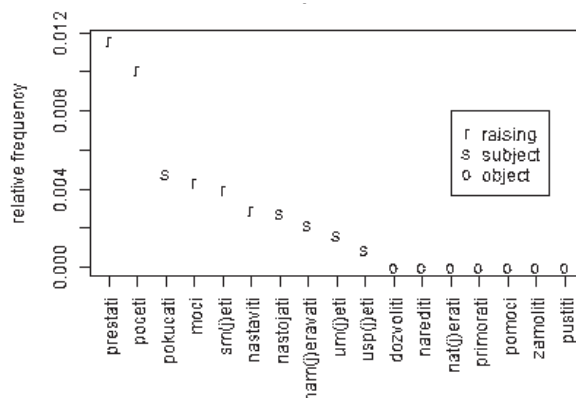
186

Fig. 2. Relative frequencies of CC for the retrieved CEPs

Figure 2 shows that the two phasal verbs *prestati* and *početi* have the highest relative frequency of CC out of *da₂*-complements, followed by *pokušati, moći, sm(j) eti* and *nastaviti*. An interesting finding is that object control CEPs do not seem to allow CC. We did not find a single example for the predicates we selected.

Further, cases in which the CL is placed to the right of the verb of the *da*-complement are extremely rare, albeit possible in all syntactic types. It is also very clear that regardless of the type of CEP, CLs tend to be placed directly after the *da* complementizer, the position which some scholars assumed to be the only possible and correct one (see [1, p. 41], [2, p. 41]).

Furthermore, in the case of CC, CLs tend to be left of the matrix verb, but can appear between the CEP and the *da* complementizer as well. If there are auxiliaries belonging to the CEP, climbed CLs can form clusters with them, as shown as in example (6). These examples disprove Todorović's claim that "if the matrix verb is in the past or future tense, whose auxiliary clitics carry the tense feature, no clitic climbing is allowed out of the subjunctive *da*-complement[5]" [18, p. 166].[6]

(6) (...) *počeo*₁   ***im*₂**   *je*   *da*   *govori*₂   *o*
     start.PTCP.SG.M them.DAT be.3SG   COMP   speak.3PRS   about

     dolasku   *ove*   *grupe.*
     arrival.LOC   this.GEN   group.GEN   (srWaC v1.2)

     '(...) he began to speak to them about the arrival of this group.'

A reflexive CL can either climb with the pronominal, as in (7), or it can stay in the *da₂*-complement, as in (8).

(7) *U*   *poslednje*   *vreme*   ***mi*₂**   ***se*₂**   *pocelo²₁*
     in   past.ACC   time.ACC   me.DAT   REFL   start.PTCP.SG.N

     da   *desava*₂   *da*   *cujem*₃   (...)
     COMP   happen.3PRS   COMP   hear.1PRS   (srWaC v1.2)

     'Recently, it has started happening to me that I hear (...)'

---

[5] In her terminology, the subjunctive complement refers to *da₂*.

[6] As is known, many BCS Internet users do not use diacritics.

(8)	(...) i	počelo₁	**mi₂**	je	da	**se₂**
	and	start.PTCP.SG.N	me.DAT	be.3SG	COMP	REFL
	vrti₂	*u*	*glavi.*			
	spin.3PRS	in	head.LOC			(srWaC v1.2)

'(...) and I started to feel dizzy.'

The fact that two CLs that were generated by the same verb do not have to climb together over *da₂* was already observed by S. Stjepanović [17, p. 182]. Her examples, however, concern only two pronominal CLs and not the reflexive *se*. S. Stjepanović [17, p. 182] concludes that in the case of a split only a dative CL can climb, while an accusative CL stays in the *da₂*-complement. Additionally, we argue that if two CLs are generated in the *da₂*-complement and split, it is the pronominal that climbs, while the reflexive tends to stay in the *da₂*-complement. Moreover, it is worth mentioning that reflexive *se* did not climb with a pronominal CL if the matrix clause contained an auxiliary clitic. Since we did not find examples with three CLs (auxiliary, pronominal and reflexive) in a cluster, it seems that whenever there are three CLs in a sentence, the reflexive tends to stay in the *da₂*-clause.

Finally, it is worth mentioning that CC was not attested for the form *je* (acc.3sg.f). This needs further investigation, but could be due to error in tagging.

We also investigated embedded finite complements, but we did not observe any link between their semantic or syntactic properties and the inclination of their clitic pronominal complements' to climb.

6.2 Diaphasic and Diatopic Variation

Regarding diaphasic variation, S. Marković [11] suggests that the phenomenon of CC out of *da₂*-complements most typically occurs in the journalistic register, but can also be found in literary texts. Our data confirm both statements. First, the subcorpus of InterCorp, consisting only of Serbian literary texts, provides two examples obviously belonging to the literary register. Second, 36 examples were published on Internet sites with predominantly journalistic texts. As regards diatopic variation, S. Marković [11] claimed that CC out of *da₂*-complements is typical of language use in Bosnia. In our sample, ijekavian spelling, which is typical of language use in Bosnia, Croatia and Montenegro, occurred in only 8 examples while ekavian spelling was used in the remaining 63 examples (including InterCorp).

7   CONCLUSIONS

In this paper, we addressed the syntactic mechanism of clitic climbing in the context of *da₂*-complements, which are characterized by the presence of a verb inflecting for person and number. This is an interesting topic because e.g. for Czech it is claimed that finite complements block CC. The point of departure of our study was the observation that there is a large disagreement as to the acceptability of CC out of *da*-complements. Whereas S. Stjepanović [17] allows the grammaticality of CC out of *da*-complements mainly within a unified formal theory of CC in BCS, other authors reject the grammaticality of this structure outright. We presented the results of

a corpus-based study which had to overcome the various shortcomings of the available corpora of Serbian. We proposed solutions to enhance precision and recall by developing sophisticated CQL queries. Our data allow the following answers to our research questions given in Sections 2 and 3:

Q1: Serbian da_2-complements do marginally allow CC. In these cases, the climbed CL can form a cluster with the auxiliary CL of the matrix verb. We thus in principle agree with [17], but have to point out that we are dealing with a highly marginal construction. Examples did not support the occurrence of CC for all CL forms.

Q2: CC is possible in raising and in subject control contexts. It is, however, most probably blocked in the case of object control. This is in line with what has been claimed for Czech.

We found some further evidence for the following constraints: first, if two CLs are generated in a da_2-complement and split, it is the pronominal that climbs and the reflexive that stays in the complement; second, reflexive *se* does not climb if there is an auxiliary clitic in the matrix clause. This suggests that the pronominal CL and reflexive *se* behave differently, which leads to the conclusion that CC is not a unified syntactic mechanism. Finally, we were able to reject Todorović's hypothesis [18] that perfect or future auxiliaries block CC.

Since we are dealing with a rare phenomenon, which seems to be restricted not only syntactically, but also stylistically and regionally, the next study should involve native speakers who would judge the acceptability of such examples.

References

[1] Browne, W. (2003). Razlike u redu riječi u zavisnoj rečenici. *Wiener Slawistischer Almanach*, 57: 45–52.

[2] Ćavar, D. and Wilder, C. (1994). "Clitic third" in Croatian. *Linguistics in Potsdam*, 1:25–63.

[3] Čermák, F. and Rosen, A. (2012). The case of InterCorp, a multilingual parallel corpus. *International Journal of Corpus Linguistics*, 17(3):411–427.

[4] Davies, W. and Dubinsky, S. (2004). *The grammar of raising and control: A course in syntactic argumentation*. Blackwell, Malden.

[5] Franks, S. and King, T. H. (2000). *A Handbook of Slavic clitics*. Oxford University Press, Oxford.

[6] Gołąb, Z. (1964). The problem of verbal moods in Slavic languages. *International Journal of Slavic Linguistics and Poetics*, 8:1–36.

[7] Ivić, M. (1970). O upotrebi glagolskih vremena u zavisnoj rečenici: prezent u rečenici s veznikom *da*. *Zbornik za filologiju i lingvistiku*, 13(1):43–54.

[8] Junghanns, U. (2002). Clitic climbing im Tschechischen. *Linguistische Arbeitsberichte*, 80:57–90.

[9] Ljubešić, N. and Klubička, F. (2014). {bs,hr,sr}WaC — Web corpora of Bosnian, Croatian and Serbian. In *Proceedings of the 9th Web as Corpus Workshop* (WaC-9), pages 29–35, Gothenburg, Sweden.

[10] Ljubešić, N., Klubička, F., Agić Ž., and Jazbec I. (2016). New Inflectional Lexicons and Training Corpora for Improved Morphosyntactic Annotation of Croatian and Serbian. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 4264–4270, ELRA, Paris, France.

[11] Marković, S. (1955). Položaj zamjeničke enklitike u vezi sa naporednom upotrebom infinitiva i prezenta sa svezicom *da*. *Naš jezik*, 1–2:33–40.

[12] Mišeska-Tomić, O. (2003). The Syntax of the Balkan Slavic future tenses. *Lingua*, 114: 517–549.

[13] Progovac, Lj. (2005). *A Syntax of Serbian: Clausal Architecture*. Slavica Publishers, Bloomington.

[14] Rezac, M. (2005). The syntax of clitic climbing in Czech. In *Clitics and affix combinations. Theoretical perspectives*, pages 103–140, Benjamins, Amsterdam, Nederlands.

[15] Stiebels, B. (2015). Control. In *Syntax – theory and analysis. An international handbook*, pages 412–446, De Gruyter, Berlin, Germany.

[16] Spencer, A. and Luís, A. R. (2012). *Clitics. An Introduction.* Cambridge University Press, Cambridge.

[17] Stjepanović, S. (2004). Clitic Climbing and Restructuring with "Finite Clause" and Infinitive Complements. *Journal of Slavic Linguistics*, 12(1):173–212.

[18] Todorović, N. (2012). *The Indicative and Subjunctive da-complements in Serbian: A Syntactic--Semantic Approach.* PhD Thesis, University of Illinois, Chicago.

[19] Wallis, S. (2014). Coping with imperfect data. Accessible at: `https://corplingstats.wordpress.com/2014/04/10/imperfect-data/,` retrieved 2017-01-31.