

*Björn Hansen, Zrinka Kolaković, Edyta Jurkiewicz-Rohrbacher*

# Clitic Climbing and Stacked Infinitives in Bosnian, Croatian and Serbian – A Corpus-Driven Study<sup>1</sup>

**Abstract** Although clitics (CLs) have been very often analysed for Bosnian, Croatian and Serbian (BCS), only few studies approach clitic climbing (CC) in BCS. According to Čamdžić & Hudson (2002) and Aljović (2004), CC out of infinitive complements is obligatory. In the present paper, we focus on constructions with stacked infinitives and address the following research question: “Can pronominal CC appear in the context of stacked infinitives?” Based on material extracted from three web corpora {bs, hr, sr}WaC, we conclude that pronominal CC does not always occur in the case of stacked infinitives in all three languages examined. We identify the following constraints: 1. CLs in the same case but depending on two different verbs block CC. 2. Reflexivity of the infinitive embedding further infinitives seems to be involved in the blocking of CC.

**Keywords** Clitic climbing, stacked infinitives, web corpora, Bosnian, Croatian, Serbian

## 1 Introduction

The syntax of clitics in Bosnian, Croatian and Serbian, by some authors called Serbo-Croatian (BCS), has been the target of intense theoretical research. The placement of clitics (CL) is usually associated with the left edge of the sentence, the so-called ‘second position’. Most works on CL in Bosnian, Croatian and Serbian address the nature of this second position effect, mainly within formal

- 1 This study was carried out within the research project ‘Microvariation of the Pronominal and Auxiliary Clitics in Bosnian, Croatian and Serbian. Empirical Studies of Spoken Languages, Dialects and Heritage Languages’ funded by the Deutsche Forschungsgemeinschaft (HA 2659/6-1, 2015-2018). We are grateful to the anonymous reviewer for helpful comments on an earlier version of the paper.

theoretical frameworks (primacy of syntactic vs prosodic processes, for an overview see Bošković 2004, Browne 2003, 2004, 2014, Franks & King 2000, Franks 2010). Descriptively speaking, CLITIC CLIMBING (CC) refers to sentence structures in which “the clitic is associated with a verb complex in a subordinate clause but is actually pronounced in construction with a higher predicate (for instance, the matrix verb which selects that subordinate clause), even though it may have no obvious semantic or syntactic connection to that verb” (Spencer & Luís 2012: 162). An example of CC out of an infinitival complement is given in (1) where the clitical pronoun *ga* has to move from the infinitival into the matrix clause:

(1) Marija **ga**<sub>2</sub> mora<sub>1</sub> vidjeti<sub>2</sub>.  
 Marija him.ACC must.3PRS see.INF

(1') \*Marija mora<sub>1</sub> vidjeti<sub>2</sub> **ga**<sub>2</sub>.  
 ‘Marija must see him.’ Aljović (2004)

Čamdžić & Hudson (2002: 326) argue that in BCS CC “[...] is obligatory when the complement is an infinitival form and marginally possible when the complement is a *da* clause”. Two years later Aljović (2004) claims the same: in the case of restructuring verbs, CC out of infinitive complements “is not an option but a necessity”. However, they do not provide any empirical evidence. A further work dealing with CC is Stjepanović (2004) but her focus is on *da*-constructions where CC is claimed to be optional. There are, actually, no empirical studies specifically dealing with CC in BCS based on natural data. The syntactic conditions of CC are thoroughly described only for Czech by Junghanns (2002), Dotlačil (2004), Rezac (2005) and Hana (2007) who propose several constraints on CC in this West Slavonic language. As we assume that the word order behaviour of clitics is based on syntactic constraints, we shall refrain from conjecturing about restrictions imposed by allegedly prosodic features.

## 2 Research question

The few existing studies which mention CC in BCS focus on the structure ‘complement taking predicate + infinitive’ as in (1), none of them, however, deals with what we call STACKED INFINITIVE CONSTRUCTIONS, i.e. complement taking predicates (CTP) showing multiple embedding of two or more infinitives, as in example (2):

- (2) Pokušavao<sub>1</sub> je prestat<sub>2</sub> pušiti<sub>3</sub>, (...)  
 try.PTCP.SG.M be.3SG stop.INF smoke.INF  
 ‘He tried to quit smoking, (...)’ (hrWaC v2.2)

We believe that precisely stacked infinitives are an ideal test case for constraints on CC because they contain all types of combinations of CL and therefore allow to identify possible contexts of blocked CC (on Czech see Hana 2007: 122–132). A further reason to restrict the search to stacked infinitives is a methodological one. Since the structure CTP + one infinitive is rather frequent, we would have been forced to work with samples that would not have enabled us to detect possible constraints, since in the samples frequently occurring raising CTP-like e.g. modal or phrasal verbs would have predominated.

In the following, we are going to test Čamdžić & Hudson’s (2002: 326) and Aljović’s (2004) claim that CC is obligatory in infinitival complements. Our research question is:

“Is CC obligatory in the context of stacked infinitives (embedding of two or more infinitives)?; i.e. can stacking of infinitives block CC?”

Our study is corpus-driven; we will present the actually attested constructions and their frequencies.

### 3 Data extraction & methodology

We extract the data from three massive, morphosyntactically tagged web corpora: bsWaC v1.2, hrWaC v2.2 and srWaC v1.2 (Ljubešić & Klubička 2014). We look for CLs in three different positions in the context of infinitive stacking (we allowed 2 to 4 infinitives in a row). The following examples (3), (4) and (5) illustrate the possible positions of the clitics:

- |     |                                       |                     |                     |                         |                              |
|-----|---------------------------------------|---------------------|---------------------|-------------------------|------------------------------|
|     |                                       | CTP                 | Infinitive          | Infinitive              | CL                           |
| (3) | I vi                                  | možete <sub>1</sub> | pomoći <sub>2</sub> | zaustaviti <sub>3</sub> | <b>ga</b> <sub>3</sub> (...) |
|     | and you.NOM                           | can.2PRS            | help.INF            | stop.INF                | him.ACC                      |
|     | ‘You can also help to stop him (...)’ |                     |                     | (bsWaC v1.2)            |                              |

- |     |   |                        |      |                    |                           |
|-----|---|------------------------|------|--------------------|---------------------------|
|     | CTP   | CL                     |      | Infinitive         | Infinitive                |
| (4) | Morate <sub>1</sub>   | <b>ih</b> <sub>3</sub> | samo | znati <sub>2</sub> | prepoznati <sub>3</sub> . |
|     | must.2PRS   | them.ACC               | only | know.INF           | recognize.INF             |
|     | ‘You just have to know how to recognize them.’ (srWaC v1.2) |                        |      |                    |                           |

	CL	CTP	Infinitive	Infinitive
(5) Ona	<b>nas</b> <sub>3</sub>	mora <sub>1</sub>	naučiti <sub>2</sub>	kontrolirati <sub>3</sub> .
she.NOM	us.ACC	must.3PRS	learn.INF	control.INF
	'She has to learn how to control us.' (bsWaC v1.2)			

In example (3), the pronominal CL *ga* remains *in situ*, following its infinitival governor *zaustaviti*. In (4), however, the pronominal clitic *ih*, which is a complement of the infinitive *prepoznati*, climbed into the matrix clause and follows the higher CTP *morati*. A structurally similar situation is found in (5), where the pronominal clitic *nas*, which is a complement of the infinitive *kontrolirati* moved to the matrix clause and precedes the higher CTP *morati*. Both (4) and (5) are perfect examples of CC. Nevertheless, they differ in respect to the word order. Therefore, our queries accounted for both above described word order patterns<sup>2</sup>. Here is example of the query CTP + INF (2,4) + CL<sup>3</sup>:

```
[!(word="(me)|([mj]u)|(joj)|(i[hm])|(ga)|([nv]as)")|
(word="je"&tag!="V.*")|(word="[mt]i"&tag!="(Pp[12]-[sp]n|Pd-mpn)")|(word="te"&tag!="(Pd-[fm][sp][nga])|(Cc)")]{1,4}[tag="Vm.*"&lemma!="ht-j?eti"&word!="nemoj[mt][eo]"][!(tag="C.*"|lemma="\Z"|tag="P[iq].*"|(tag="(V.*y)|(Va[ae]fpmn).*)|(Vc[efp]m).*)|(Vm.*)"&word!="ćeš")|tag="Rr"|word=".\..*"|lemma="što")]{0,4}[tag="V.n"]{2,4}[!(tag="C.*"|lemma="\Z"|tag="P[iq].*"|(tag="(V.*y)|(Va[ae]fpmn).*)|(Vc[efp]m).*)|(Vm.*)"&word!="ćeš")|tag="Rr"|word=".\..*"|lemma="što")]{0,4}[(word="(me)|([mj]u)|(joj)|(i[hm])|(ga)|([nv]as)")|(word="je"&tag!="V.*")|(word="[mt]i"&tag!="(Pp[12]-[sp]n|Pd-mpn)")|(word="te"&tag!="(Pd-[fm][sp][nga])|(Cc)")]{1,4}[!(tag="Rr"|tag="(V.*y)|(Va[ae]fpmn).*)|(Vc[efp]m).*)|(Vm.*)"&word!="ćeš")|word=".\..*"|lemma="što")]{0,4}within<s/>
```

- 2 We are aware of the fact that the reverse order infinitive complements-CTP is possible in BCS, but we did not take it into account, because it represents information about structurally marked word order. Additionally, infinitive + infinitive + CTP poses difficulties for corpora, where the sentence clause border is not annotated. Hence, the precision of the queries in question would be very low.
- 3 Index of morphosyntactic descriptions MSD at <http://nl.ijs.si/ME/V5/msd/html/msd-hr.html#msd.msds-hr>.

In the query, we excluded all forms of the lemma *htjeti* (‘will’, ‘want’) since the corpus annotation does not offer disambiguation of its function as an auxiliary verb, which in combination with the infinitive forms the future tense, or as modal verb. Furthermore, we excluded the forms *nemoj*, *nemojmo* and *nemojte*, which in combination with the infinitive express prohibitive in BCS.

In order to obtain most occurrences of the constructions, we could not restrict the query only to the core elements of the construction (CTP, Infinitive stack, CL), but we allowed empty positions, so elements such as clitics governed by CTP could appear. Nevertheless, we excluded from empty positions most elements marking the sentence clause, such as conjunctions, other main verbs, and punctuation signs.

The resulting recall required manual processing, also due to errors in tagging. Since hrWaC v2.2 is two and a half times bigger than srWaC v1.2 and five times bigger than bsWaC v1.2 the query returned proportionally higher results, which are almost impossible to process manually. Therefore, for hrWaC v2.2 we generated three samples via NoSketch Engine (function “Sample”) which comprise a quarter of the originally retrieved hits.

Apart from empty positions which decreased the recall, some duplicates and hits which were linked only to CC out of the first infinitive, as in the example given in (6), had to be excluded manually.

- (6) (...) možemo<sub>1</sub> **im**<sub>2</sub> pomoći<sub>2</sub> popraviti<sub>3</sub> ponašanje (...)
   
can.1PRS them.DAT help.INF correct.INF behaviour.ACC
   
‘(...) we can help them to correct their behaviour (...)’ (bsWaC 1.2)

The reason for that is the fact that in accordance with our research question we focus on CL depending on the second infinitive, as is it only in that case that stacked infinitives may or may not block CC. The sentences in which two clitics appeared, one as a complement of the first infinitive and the other as a complement of the second (or in rare cases of the third) infinitive were taken into consideration, see the example in (7):

- (7) (...) možete<sub>1</sub> **si**<sub>2</sub> dozvoliti<sub>2</sub> uskratiti<sub>3</sub> **mi**<sub>3</sub> sve (...)
   
can.2PRS REFL.DAT allow.INF curtail.INF me.DAT everything
   
‘(...) you can allow yourself to curtail everything from me (...)’ (hrWaC v2.2)

In those cases, our focus was on the clitic which is a complement of the second infinitive (here *uskratiti*) and, of course, on the relationship between two clitics, by which we mean the formation of a clitic cluster or clitic split as in the case of *si* and *mi* in example (7).

Although our queries allowed a maximum of four embedded infinitives, we found only three examples with three infinitives (see one of the examples in (8)) and no example of a bigger stack.

- (8) (...) samo **se**<sub>3</sub> ne smijem<sub>1</sub> zaboraviti<sub>2</sub> sjetiti<sub>3</sub>  
 only REFL NEG must.1PRS forget.INF remember.INF  
 reći<sub>4</sub> **im**<sub>4</sub> (...)  
 tell.INF them.DAT (hrWaC v2.2)  
 ‘(...) I only must not forget to remember to tell them (...)’

A corpus-driven study may help to determine factors which are responsible for CC or the lack of CC respectively, but it requires an additional manual annotation of samples. In the present study, our annotation scheme contains the language variety, the word order behaviour of CL, grammatical features of the CL and basic syntactic properties of the predicates the CL depends on (raising vs control).

#### 4 Results & discussion: clitic climbing and stacked infinitives

Our results give a clear answer to the research question. As can be seen in Figure 1, which presents the final distribution of the target constructions across each corpus, stacked infinitives as such do not prevent CLs from climbing into the matrix clause. We find both examples with CC (83,44–86,12 %) and without CC (13,88–16,56 %).

We have not found significant, language-specific differences in the distributions of the constructions with CC and without CC ( $\chi^2$  test, p-value 0.51). The low overall recall in srWaC v1.2 can be explained by the fact that especially in Serbian the infinitive competes with the semifinite *da*-construction, as in (9).

- (9) (...) stvarno moram<sub>1</sub> da počnem<sub>2</sub> da učim<sub>3</sub> (...)  
 really must.1PRS COMP start.1PRS COMP learn.1PRS  
 ‘(...) I really have to start to study (...)’ (srWaC v1.2)

Regarding our results, it is interesting to point out that even in those rare cases with three infinitives. The CL of the last infinitival complement could climb over three CTPs into the matrix clause, as shown in (10): *držati ga* (‘to hold him’)

- (10) (...) i u svakome trenutku **ga**<sub>4</sub> možemo<sub>1</sub>  
 and in any moment him.ACC can.1PRS  
 odlučiti<sub>2</sub> prestati<sub>3</sub> držati<sub>4</sub> (...)  
 decide.INF stop.INF hold.INF (hrWaC v2.2)  
 ‘(...) and in any moment, we can decide to stop holding him (...)’

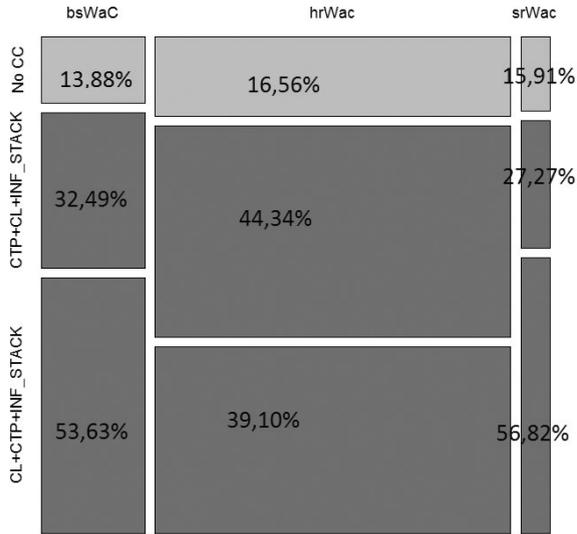


Figure 1: CC and stacked infinitives in {bs, hr, sr}WaC. ( $\Sigma$  of all examples 1492 = 317 (bsWaC v1.2) + 1087 (hrWaC v2.2) + 88 (srWaC v1.2))

## 5 Conclusion & further perspectives

To conclude, our corpus-driven study based on data from three web corpora has shown the range and frequency of word order patterns of CL in constructions with stacked infinitives in BCS. We have found that:

- i. Clitics can climb within stacked infinitives.
- ii. In stacked infinitive constructions, CC is found in around 83,44–86,12 % and the lack of CC in 13,88–16,56 % of all cases.
- iii. There are no significant, language-specific differences in the distributions of the researched constructions.

Coming back to our research question from Section 2, we can draw the conclusion that CC in BCS is not always obligatory (contra Čamdžić & Hudson 2002: 326). This might be explained in two ways: first, CC *per se* is facultative or, second, CC is obligatory but subject to constraints.

Following the latter assumption, our corpus-driven study allows formulation of a few hypotheses concerning possible constraints on CC:

- i. We found some evidence for ‘Same case different governors constraint’: CC might be blocked if two CL depending on two different CTPs have the same case as in ex. (7) where two clitics in Dative are split (*si, mi*). It is worth pointing out that this constraint may be a subtype of ‘object control case constraint’ (see Dotlačil 2004 and Rezac 2005 for more details).

- ii. Reflexivity of the infinitive embedding further infinitives seems to play a crucial role in blocking clitic climbing (Odds Ratio test with 95% confidence level yields 502.8000,  $p < 0.0001$ ) as in ex. (7) and (8)<sup>4</sup>.
- iii. We have also found a significant relation between the syntactic type of the infinitive governing further infinitives (Chi-square test 95.78,  $p < 0.0001$ ), but with medium size effect (Cramer's  $V = 0.2535$ ). CC from infinitive stacks governed by object-control infinitive (as the predicate *pomoći* 'to help' in ex. (3)) or by subject-control infinitive is more restricted than from raising. Our findings from (ii) help explain this fact: raising verbs are never reflexive, while every sixth subject-control and every eighth object-control verb in our data set is reflexive.

More findings could be obtained by extending the annotation schema. In the future, we intend to explore whether grammatical or lexical properties of the CL themselves influence CC, and how CL interacts with CL governed by other infinitives and CTP. This will allow a clearer picture of the nature of CC.

We have to be aware however, of the fact that the patterns of actual language usage described in this paper do not directly reflect constraints in a proper sense of the word. A corpus study can only provide first clues for possible constraints on CC. As not all combinations of CTPs and CL could be found in the corpora we envisage the triangulation of methods; i.e. we plan to carry out systematic experiments comprising acceptability judgements with a larger number of native speakers. As argued by Diesing, Filipović Đurđević & Zec (2009), the study of the syntax of clitics demands the combination of corpus and experimental data.

## 6 References

- Aljović, Nadira. 2004. Cliticization Domains: Clitic Climbing in Romance and in Serbo-Croatian. In Olivier Crouzet, Hamida Demirdache and Sophie Wauquier-Gravelines (eds.), *Proceedings of JEL'2004 Domain(e)s*, 169–175. Université de Nantes.
- Aljović, Nadira. 2005. On clitic climbing in Bosnian/Croatian/Serbian. In Nedžad Leko (ed.), *Lingvistički vidici* 34:(05), 58–84. Sarajevo: Forum Bosnae.

4 One of the anonymous reviewers proposed the stricter formulation “reflexivity blocks CC”, we, however, would like to keep it in this way since in this first phase of our research we did not distinguish between different types of reflexive CLs. Lešnerová & Malink's (2008) study conducted on Czech suggest that different reflexives indeed behave differently in respect to CC. This may be the case in BCS as well and we plan to investigate it in more depth.

- Bošković, Željko. 2004. Clitic placement in South Slavic. *Journal of Slavic Linguistics* 12: (1), 37–90.
- Browne, Wayles. 2003. Razlike u redu riječi u zavisnoj rečenici. *Wiener Slawistischer Almanach*. Sonderband 57, 45–52.
- Browne, Wayles. 2004. Serbo-Croatian Enclitics for English-Speaking Learners. *Journal of Slavic Linguistics* 12: (1), 249–283 (Reprint from 1975).
- Browne, Wayles. 2014. Groups of Clitics in West and South Slavic Languages. In Elżbieta Kaczmarska and Motoki Nomachi (eds.), *Slavic and German in Contact: Studies from Areal and Contrastive Linguistics*. Slavic Eurasian Studies 26, 81–96. Hokkaido: Slavic Research Center.
- Čamdžić, Amela; Hudson, Richard. 2002. *Clitics in Serbo-Croat-Bosnian*. In John Harris et al. (eds.), *UCL Working Papers in Linguistics 14*, 321–353. London: Department of Phonetics and Linguistics University College London. [http://www.phon.ucl.ac.uk/home/PUB/WPL/02papers/camdzc\\_hudson.pdf](http://www.phon.ucl.ac.uk/home/PUB/WPL/02papers/camdzc_hudson.pdf). (23.06.2017).
- Diesing, Molly; Filipović Đurđević, Dušica and Zec, Draga. 2009. Clitic placement in Serbian: Corpus and experimental evidence. In Susanne Winkler and Sam Featherston (eds.), *The Fruits of Empirical Linguistics II: Product*, 59–73. Berlin/New York: Mouton de Gruyter.
- Dotlačil, Jakub. 2004. *The syntax of infinitives in Czech*. Master's Thesis. University in Tromsø. <http://jakubdotlacil.com/thesis.pdf> (03.05.2017).
- Franks, Steven. 2010. Clitics in Slavic. *Contemporary Issues in Slavic Linguistics* 10, 1–157.
- Franks, Steven; King, Tracy Holloway. 2000. *A Handbook of Slavic clitics*. New York/Oxford: Oxford University Press.
- Hana, Jirka. 2007. *Czech Clitics in Higher Order Grammar*. PhD Thesis. The Ohio State University.
- Junghanns, Uwe. 2002. Clitic climbing im Tschechischen. *Linguistische Arbeitsberichte* 80, 57–90.
- Lešnerová, Šárka; Malink, Marko. 2008. Clitic Climbing and Theta-Roles in Upper Sorbian and Czech. In Gerhild Zybatow et al. (eds.), *Formal Description of Slavic Languages: FDSL 5*, 396–407. Frankfurt am Main: Peter Lang. <http://philosophy.uchicago.edu/faculty/files/malink/14%20Clitic%20Climbing%202008.pdf>. (16.05.2017).
- Ljubešić, Nikola; Klubička, Filip. 2014. {bs,hr,sr} WaC – Web corpora of Bosnian, Croatian and Serbian. In Felix Bildhauer and Roland Schäfer (eds.), *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*, 29–35. Gothenburg, Sweden.
- Rezac, Milan. 2005. The syntax of clitic climbing in Czech. In Lorie Heggie and Francisco Ordóñez (eds.), *Clitics and affix combinations. Theoretical perspectives*, 103–140. Amsterdam: Benjamins.

<http://minimalism.linguistics.arizona.edu/AMSA/PDF/AMSA-202-0602.pdf>.  
(02.05.2017).

Spencer, Andrew; Luís, Ana R. 2012. *Clitics. An Introduction*. Cambridge: Cambridge University Press.

Stjepanović, Sandra. 2004. Clitic Climbing and Restructuring with “Finite Clause” and Infinitive Complements. *Journal of Slavic Linguistics* 12: (1), 173–212.