Zrinka Kolaković

Institut für Slavistik, Fakultät für Sprach-, Literatur- und Kulturwissenschaften, Universität Regensburg

Universitätsstraße 31, DE-93053 Regensburg

Institut für Slawistik, Fakultät für Kulturwissenschaften, Alpen-Adria-Universität Klagenfurt

Universitätsstraße 65-67, 9020 Klagenfurt

Zrinka.Kolakovic@aau.at

Edyta Jurkiewicz-Rohrbacher

Institut für Slavistik, Fakultät für Sprach-, Literatur- und Kulturwissenschaften, Universität Regensburg; University of Helsinki

Universitätsstraße 31, DE-93053 Regensburg

Edyta.Jurkiewicz-Rohrbacher@sprachlit.uni-regensburg.de

Björn Hansen

Institut für Slavistik, Fakultät für Sprach-, Literatur- und Kulturwissenschaften, Universität Regensburg

Universitätsstraße 31, DE-93053 Regensburg

Bjoern.Hansen@sprachlit.uni-regensburg.de

CLITIC CLIMBING, THE RAISING-CONTROL DICHOTOMY AND DIAPHASIC VARIATION IN CROATIAN[1]

In the paper, we discuss the phenomenon of clitic climbing (CC) out of infinitive complements in contemporary Croatian. Based on the first theoretical work and some empirical findings on CC in Czech and Bosnian, Croatian and Serbian (BCS) and the observation that differences in CC linked to register have been reported for some languages, we elaborate on the claim that CC varies in respect of both register and the Raising-Control Dichotomy. The following research questions are addressed: Does clitic climbing out of the single infinitive in Croatian depend on the type of

complement-taking predicate (CTP) with respect to the Raising-Control Distinction? Does CC appear with equal frequency in standard and colloquial Croatian if the type of CTP verb (Raising vs Control) as a variable remains constant?

Our study is based on data for two types of complement-taking predicates: a) Raising (8 different verbs) and b) Subject Control (8 non-reflexive + 8 reflexive verbs). The data was extracted from the Forum subcorpus of hrWaC v2.2 and from the Croatian Language Repository and Croatian National Corpus. Our data suggest that not only the Raising-Control Dichotomy, but also diaphasic variation have an impact on CC from infinitive complements.

1. Introduction

Descriptively speaking, CLITIC CLIMBING (CC) "refers to 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" (Spencer and Luís 2012: 162). An example of CC out of an infinitive complement is given in (1) where the reflexive clitic *se* is realised in the second position of the matrix clause (Wackernagel position); in other cases, however, CC does not take place, as may be seen in (2) where the clitic *ih* stays in the complement.

(1) I   mi   **se**$_2$ planiramo$_1$ baciti$_2$ na   posao.[2]
     and   we   REFL   plan.1PRS   throw.INF   at   work.ACC
     'We also plan to throw ourselves into work.'   [hrWaC v2.2]

(2) Bojim$_1$ **se**$_1$ testirati$_2$ **ih**$_2$.
     be.afraid.1PRS   REFL   test.INF   them.ACC
     'I am afraid to test them.'   [hrWaC v2.2]

Although clitics (CLs) in Bosnian, Croatian, and Serbian (BCS)[3] have attracted considerable attention in the syntactic literature (cf. Franks and King 2000, Browne 2004, Bošković 2004, Progovac 2005, Diesing and Zec 2011, 2017), the

---

[2] We index CTPs and their respective CLs with 1 and infinitive complements and their respective CLs with 2.
[3] The label Bosnian/Croatian/Serbian (BCS) is used to refer to the Štokavian language use common to the varieties used in Croatia, Serbia, Bosnia-Herzegovina and Montenegro, and when referring to other works in which that term is used. Since our empirical study targets the language structures in codified language, we later use the single label Croatian. The question whether we are dealing with independent languages or with national variants of a so-called polycentric language is not relevant to our study.

syntactic conditions for and constraints on CC are seriously understudied in comparison to Czech, for example (e.g. Rezac 2005, Junghanns 2002, Dotlačil 2004, Hana 2007, Rosen 2014). There are very few studies on the constraints on CC in BCS: Stjepanović (2004), Aljović (2004, 2005), Jurkiewicz-Rohrbacher; Kolaković and Hansen (2017), Jurkiewicz-Rohrbacher; Hansen and Kolaković (2017), and Hansen; Kolaković and Jurkiewicz-Rohrbacher (2018).[4] The latter three papers are the only descriptions of CC in BCS based on empirical investigation. Jurkiewicz-Rohrbacher; Hansen and Kolaković (2017) show the Raising-Control Dichotomy of matrix predicates to be a relevant factor for CC out of da_2-complements[5] in Serbian. Hansen; Kolaković and Jurkiewicz-Rohrbacher (2018) show that CC out of stacked infinitives is possible (see ex. (3)), but not obligatory in BCS. Moreover, they detect two possible constraints on CC out of stacked infinitives. First, CC might be blocked if two CLs depending on two different CTPs are in the same case. Second, reflexivity of the infinitive embedding further infinitives decreases the probability of CC (Hansen; Kolaković and Jurkiewicz-Rohrbacher 2018: 265f).

(3) […] i u svakome trenutku **ga**$_4$ možemo$_1$ odlučiti$_2$
and in any moment him.ACC can.1PRS decide.INF
prestati$_3$ držati$_4$ […].
stop.INF hold.INF

'[…] and in any moment, we can decide to stop holding him […].' [hrWaC v2.2]

However, the constraints on CC out of a single infinitive complement in Croatian remain unclear.

The present paper is empirically oriented. Section 2 introduces the Raising-
-Control Distinction, and summarizes the discussion on CC out of infinitive complements. Section 3 presents the importance of diaphasic variation for CC in Spanish and Portuguese. Spanish and Portuguese are of interest because their clitic system shows some common features with Croatian. Our research questions are presented in Section 4. The choice of data and the collection process

---

[4] Čamdžić and Hudson (2002) do mention CC in BCS, but they do not elaborate on the constraints.
[5] Complements headed by *da* 'that' do not behave in a uniform way. In most approaches the distinction between da_1- and da_2-complements depends on tense marking. The former are regularly marked for tense and the latter are not. In contrast to da_1-complements, da_2-complements only allow the verbal form coinciding with the present tense; other tenses are impossible (cf. Ivić 1970, Browne 2003).

are explained in Section 5 and 6, while Section 7 describes the results in detail, and is followed by the final Section 8, which draws conclusions and offers suggestions for future research.

2. Raising-Control Distinction and CC

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. It forms the basis for a typology of complement-taking predicates (CTP).

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 contrast, in a control construction the matrix verb and the embedded verb each assign a subject its thematic role. Davies and Dubinsky (2004: 4–8) list some relatively robust, cross-linguistically applicable tests[6] which have been proposed in the literature in order to distinguish raising constructions from control constructions:

1) In the case of Raising predicates the subject argument receives its semantic role of agent from the embedded complement, whereas in the case of Subject predicates the subject argument receives it from the matrix predicate.

2) Raising matrix predicates do not impose the selectional restriction +/- human.

3) In the case of Raising predicates passivization does not change the propositional meaning of the sentences.

A distinction should be made between Subject and Object Control constructions depending on the argument selected as controller (first or second argument). According to Stiebels (2015: 422), verbs denoting commissive speech acts (e.g. *obećati* 'promise') are typical Subject Control whereas predicates which refer to directive speech acts (e.g. *zamoliti* 'to request') or which have a causative component belong to the canonical class of Object Control predicates. In the present paper, we will deal exclusively with Subject Control.

---

[6] Due to lack of space, we will not provide examples here: for examples see Jurkiewicz-Rohrbacher et al. (2017b: 181f).

The importance of the Raising-Control Dichotomy for CC in BCS has been rec-
ognized in very few papers. Aljović (2005) observes that in BCS, CC is only
available out of complements whose subject is empty and coreferential with the
matrix subject,[7] although in a footnote she acknowledges that CC is also pos-
sible when the subject of the embedded complement is coreferential with a ma-
trix indirect object in the dative, i.e. out of object-controlled infinitives (Aljović
2005). Jurkiewicz-Rohrbacher et al. (2017) show, on solid empirical evidence,
that neither pronominal nor reflexive CLs can climb from object-controlled da_2-
-complements. For Czech, a range of various constraints to CC tightly connected
with Object Control has been identified in the theoretical literature (Rezac 2005,
Dotlačil 2004, Hana 2007). Apparently, in Czech there are no restrictions to CC
out of infinitive complements governed by Raising and Subject Control CTPs
(Rezac 2005, Hana 2007). In work on BCS, CC is sometimes seen in the context
of "restructuring", a process that in one way or the other transforms two clauses
into one. Aljović (2005) assumes that CC is obligatory with restructuring predi-
cates (cf. Aljović 2005). Although we are aware of the fact that restructuring and
raising predicates are not identical, we would like to point out that Stjepanović
(2004: 198–204) observes that restructuring verbs behave like raising verbs. In
contrast, Hansen; Kolaković and Jurkiewicz-Rohrbacher (2018) demonstrated
that in BCS, CC out of stacked infinitives, i.e. multiply embedded infinitives, is
not obligatory. However, there are still no empirical data for CC out of single in-
finitive complements. Furthermore, as far as we know, theoretical syntacticians
who asserted that there are no constraints on CC out of subject controlled infini-
tives did not take into account that there are different types of Subject Control
predicates. We believe, for instance, that simple Subject Control predicates such
as *planirati* 'to plan' (ex. (1)) may behave differently with respect to CC than re-
flexive Subject Control predicates such as *bojati se* 'to be afraid' (ex. (2)). Since
the latter type of Subject Control predicate has its own reflexive CL *se*, CC out
of their infinitives might be more restricted than in the case of simple Subject
Control predicates. In other words, similar constraints could appear on CC out
of infinitive complements of reflexive Subject Control predicates as on CC out
of Object Control predicates.

---

[7] Hence, out of complements of Raising and Subject Control CTPs.

3. CC and Diaphasic variation

For a perspective on the relationship between CC and diaphasic variation it is worth looking at Romance languages. As a matter of fact, Spanish and Portuguese are languages with CLs that can climb. In the literature about variation in Spanish CC, several authors have pointed out the relevance of register:[8] generally it can be said that CC is less frequent in Spanish written texts than in Spanish spoken texts (e.g. Davies 1995, Cacoullos 1999). Davies (1995) reported a systematic difference between registers: his data show that the difference between registers with respect to CC can be as high as 30% (Davies 1995: 373f). Cacoullos (1999) reports higher rates of CC in sociolinguistic interviews than in essays (89% compared to 68%). De Andrade (2010) replicated the results of studies on Spanish CC for European Portuguese data, and confirmed the relevance of register for CC also in this language. Using basic statistical correlation testing, he showed that CC rates in the formal (newspaper interviews and novels) and informal register (sociolinguistic interviews) were significantly different (de Andrade 2010: 99).

Although we are aware of the differences in the stratification of the languages mentioned above, we take these results as a point of departure for addressing diaphasic variation in Croatian. Due to lack of space, we cannot give a full account of the stratification of Croatian. In the corpus study below, we decided to analyse CC in standard and in colloquial Croatian. Although they share many similarities, the latter has language elements which are not part of the codified norm (cf. Langston and Peti-Stantić 2014: 30).

4. Research questions

Based on the considerations presented in Sections 2 and 3 we investigate the claim that CC is dependent both on the Raising-Control Dichotomy and on register, and address the following research questions:

---

[8] It is obvious that the authors who worked on the impact of register on CC in Spanish and Portuguese have a different understanding of register. In its broadest sense register is a language variety defined by the context of usage (Čolak 2015: 31).

Q1: Does clitic climbing out of single infinitives in Croatian depend on CTP with respect to the Raising-Control Distinction?

Q2: Does CC appear equally often in standard and colloquial Croatian if the CTP verb type (Raising vs Control) as a variable remains constant?

5. Source of data

To answer our research questions we had to find appropriate sources of data, which would reflect the Croatian language, which is under direct influence of prescriptive norms on the one hand, and the Croatian language, which lies beyond the influence of prescriptive norms on the other. The Croatian Web Corpus (Ljubešić and Klubička 2014, henceforth hrWaC) is the only Croatian corpus which contains texts not only with standard but also with colloquial language features,[9] primarily in user-generated content such as comments, blogs or fora. This data variety is, like availability of meta-information (allowing to track where the texts come from), size, annotation and accessibility, a great advantage of hrWaC (cf. Jurkiewicz-Rohrbacher; Kolaković and Hansen 2017). However, since we are interested in an in-depth analysis of diaphasic variation and its influence on CC, we decided not to use the whole of hrWaC, but only its subcorpus Forum. This was motivated by two important factors. First of all, we could be sure that this subcorpus had not been externally standardized by any proofreader and adapted to the official norm. Second, the Forum subcorpus is still relatively big – it contains 237 485 906 tokens, which is crucial if we want to obtain an amount of data which is necessary for thorough statistical analysis. As the source of texts written in standard Croatian we decided to use the Croatian Language Repository (Ćavar and Brozović Rončević 2012, henceforth Repository) which contains 101 782 863 tokens. Since the amount of data collected for Subject Control reflexive CTPs in the Repository is drastically lower than the amount of data collected from the Forum subcorpus,[10] we decided to use the

---

[9]   Polančec and Mihaljević (2016: 444) report that while the language of the hrWaC corpus reflects standard and everyday colloquial Croatian, dialectal varieties are represented only to a limited extent. Their claims are based on the reduced presence of the dialectal invariant relativizers *ča* and *kaj*.

[10]   This should not come as a surprise since the Forum subcorpus is more than twice as big as the Croatian Language Repository.

Croatian National Corpus (Tadić 1998, henceforth CNC), which contains 216 847 110 tokens, as a supplementary source of data for standard Croatian in our study.[11]

6. Data collection

Once we established the list of matrix predicates,[12] we examined the behaviour of CLs in relation to the type of complement-taking predicate and standardness of Croatian language. For Raising and simple Subject Control predicates we extracted the data from the Forum subcorpus and the Repository, while the infrequent instances of reflexive Subject Control predicates for standard Croatian from the Repository had to be supplemented with additional data from the CNC.

The corpus queries took into account three word-order patterns (CL – clitic, CTP – complement-taking predicate, INF – embedded infinitive complement, to which we will refer as *the core elements* of the query):

1. CTP INF CL

2. CTP CL INF

3. CL CTP INF

If we look at the structures listed above, we see that 2 & 3 are cases of CC,[13] while in 1 CC does not take place.

---

[11] One may argue that the Forum can still be heterogeneous with respect to standardness since besides colloquial texts it may also contain texts, which are part of standard language. We will not dispute this; while some people prefer writing in a more relaxed style, others choose to stick to the norm. However, it must be pointed out that the Croatian Language Repository and the Croatian National Corpus are to some extent also heterogeneous with respect to standardness of their texts since they contain literary texts, and in those texts dialects, jargon and/or colloquial language are often used to portray characters.

[12] A list of all the main verbs, which have an infinitive as a complement, was extracted from hrWaC. We then classified CTPs with respect to the Raising and Control Distinction. To allow balanced data to be obtained it was necessary for all groups of predicate types to have the same number of representative verbs. The number of eight CTPs per syntactic type was determined by the number of Raising predicates, which formed a rather small group since we applied the criterion of obligatory raising of the subject (i.e. modal and phasal verbs). Due to lack of space, all selected CTPs are listed in Figure 2 and not in a separate section.

[13] We are aware of the fact that the structure CTP CL INF in which the CL directly precedes the infinitive is not a clear case of CC. Namely, Junghanns (2002: 67) warns that if we have the surface word order 'matrix predicate + CL + infinitive' where the CL is directly before the infinitive it is possible that the CL actually remains in the complement although on the surface it is placed left of the infinitive. Therefore, our query for the structure listed under 2. was so designed that at least one additional element had to appear between the CL and the infinitive.

While formulating queries on the basis of the pattern (see example query for Pattern 2 in Figure 1) we combined attributes of the morphological tag, the word form, and lemma. We allowed only occurrences of single pronominal and reflexive CLs: CC of a clitic cluster was excluded. Moreover, pronominal CLs were narrowed to third person only, and CTPs to present tense only, to avoid auxiliary CLs.

In order to improve our recall, we allowed other elements to appear between the core elements of the query, but the extra elements could be no longer than between two and four tokens. Nevertheless, additional core elements, expressions that would most probably mark the sentence and clause crossing were not allowed as extra elements.

```
[!(word="(me)|([mj]u)|(joj)|(i[hm])|(ga)|([nv]a[sm])|(se)"|(word="(je)|(si)"&tag!="V.*")|
(word="[mt]i"&tag!="(Pp[12]-[sp]n)|(Pd-mpn)")|(word="te"&tag!="(Pd-[fm][sp][nga])|(Cc)"))]{1,2}
[lemma="CTP" & tag="V.r.*"][!(tag="Rgp"|tag="C.*"|lemma="\Z"|tag="P[iq].*"|tag="V.*"|tag="Rr"|
word=".*\..*"|lemma="što"|tag="S.*"|(word="(me)|([mj]u)|(joj)|(i[hm])|(ga)|([nv]a[sm])|(se)")|
(word="(je)|(si)"&tag!="V.*")|(word="[mt]i"&tag!="(Pp[12]-[sp]n)|(Pd-mpn)")|
(word="te"&tag!="(Pd-[fm][sp][nga])|(Cc)"))]{0,4}
[word="([mj]u)|(joj)|(i[hm])|(ga)|(se)"|(word="(je)|(si)"&tag!="V.*")]
[!(tag="C.*"|lemma="\Z"|tag="P[iq].*"|tag="V.*"|tag="Rr"|word=".*\..*"|lemma="što"|
word="(me)|([mj]u)|(joj)|(i[hm])|(ga)|([nv]a[sm])|(se)"|(word="(je)|(si)"&tag!="V.*")|
(word="[mt]i"&tag!="(Pp[12]-[sp]n)|(Pd-mpn)")|(word="te"&tag!="(Pd-[fm][sp][nga])|(Cc)"))]{1,4}
[tag="V.n" & lemma!="biti"][!(tag="C.*"|lemma="\Z"|tag="P[iq].*"|tag="V.*"|
tag="Rr"|word=".*\..*"|lemma="što"|tag="S.*"|(word="([ms]e)|([mj]u)|(joj)|(i[hm])|(ga)|([nv]as)")|
(word="(je)|(si)"&tag!="V.*")|(word="[mt]i"&tag!="(Pp[12]-[sp]n)|(Pd-mpn)")|
(word="te"&tag!="(Pd-[fm][sp][nga])|(Cc)"))]{1,2}within<s/>
```

Figure 1. Query example for Pattern 2

The morphosyntactic tag sets for Croatian were developed on the basis of tag sets from the Multext-East project, but the Forum (hrWaC v2.2) and Repository use the latest, updated version 5,[14] while the CNC uses the older version 4.[15] For this reason, queries for the Forum and Repository could be designed together, while querying the CNC had to be prepared separately and full equivalence of queries could not be ensured. Therefore, in the latter case we used a simplified procedure, which involved multiple filtering. We first retrieved all instances of a given CTP in the present tense form with the query [lemma="CTP" & msd="Vmip.*"]. Within the result, we filtered out all instances of reflexive CLs in the region of seven words before or after the CTP. After that, we identified the instances of embedded complements, which could be no further than ten tokens after CTP, and then excluded the occurrences of the complementizer *da* up to 10 tokens after CTP to avoid the so-called *da-construction*. This simplified procedure of data collection with

---

[14] http://nl.ijs.si/ME/V5/msd/html/msd-hr.html
[15] http://nl.ijs.si/ME/V4/msd/html/msd-hr.html

multiple filtering was also the reason why the CNC was used only as a sup-plementary source of standard Croatian.

Having designed the CTP list and the queries we proceeded to data collection. Due to processing problems arising from recall and precision, and since manual revision of all the retrieved examples would have exceed our human capaci-ties, we decided to work with random samples of maximally[16] 100 examples per structure per CTP. Maximally one hit from one web page or text was taken for the sample. The samples were downloaded and revised manually, and then used in the analysis described in the next section.

7. Results and discussion

7.1. Data distribution

We did not manage to retrieve from the list of CTPs any occurrences of *sti-djeti se* 'to be ashamed' in the three queried patterns, while the verbs *sramiti se* 'to be ashamed' and *kretati* 'to go, to start' were identified only in the Forum. All in all, the analysed data set comprised 1566 observations from the Forum and 761 from the Repository and CNC, which also corresponds with the size of the corpora used. The samples obtained for different CTPs differed drastically.

We identified 1027 observations of Raising CTPs, 1118 of simple Subject Con-trol and only 182 of reflexive Subject Control predicates. The frequencies of individual lexemes in the data are shown below in Figure 2. The distribution is proportional to the absolute frequency of the lexemes in the whole hrWaC (not presented in the article due to lack of space). Simple Subject Control predicates are generally less frequent than Raising predicates, and reflexive Subject Con-trol predicates are even less frequent than simple Subject Control predicates. Therefore, it is strictly impossible to build frequency triplets between the three types of predicates.

---

[16] In the case of the less frequent simple and reflexive Subject Control predicates, when a query resulted in less than 100 examples, all the retrieved examples were checked manually.
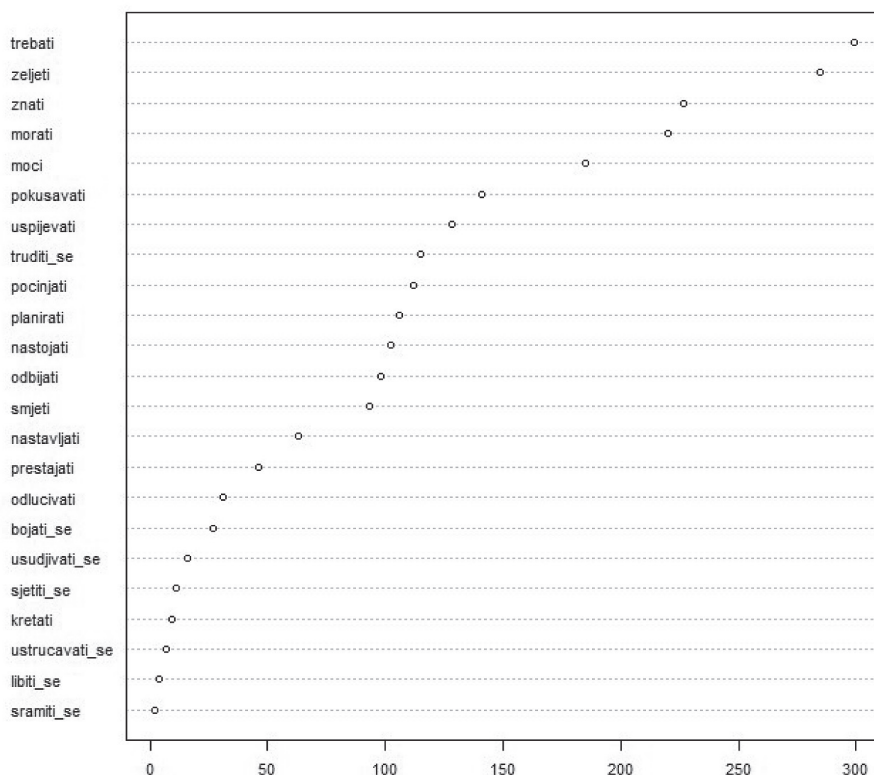
Figure 2. Distribution of different CTP lexemes in the data set: abscissa – number of observations per CTP lexeme, ordinate – CTP lexemes chosen for study

Since infinitive complements were not restricted in the query, we did not use them as independent variables. However, we examined their distribution in order to exclude the possibility of their having a significant impact on our results (e.g. if a particularly frequent complement dominated the data with a clear pattern). In the data we identified 837 distinct infinitive complements, the five most frequent being: *baviti se* 'be occupied with' (3%), *vratiti se* 'to return' (3%), *držati* 'to hold' (1.7%), *nositi* 'to carry' (1.4%), *dati* 'to give' (1.3%).

We will now discuss the distributions of our independent variables in the context of the studied dependent variable, which was the presence (presented as CC in our figures) or absence of CC (presented as noCC in our figures).

Our first independent variable was CTP Type with three levels: Raising, simple Subject Control, reflexive Subject Control. Our second independent variable was corpus with two levels: Standard (Repository + CNC) and Forum. CC occurred in 1850 cases, while in 477 cases, that is, in 20% of the sample, CC did not occur.

Figure 3 below shows the position of the CL for all studied CTPs. The plots on the left visualise the Forum, the plots on the right, the Standard (Repository + CNC). While Raising and simple Subject Control predicates show a strong tendency to appear in CC constructions, reflexive Subject Control shows the opposite trend. In addition, the preference for CC seems at first to be more frequent in the observations from standard language corpora for all three types of predicates. However, it is worth pointing out that the noCC rate in the Forum subcorpus clearly increases in the case of Raising CTPs *trebati* 'have to' and *moći* 'can', and simple Subject Control CTP *željeti* 'wish/want'.
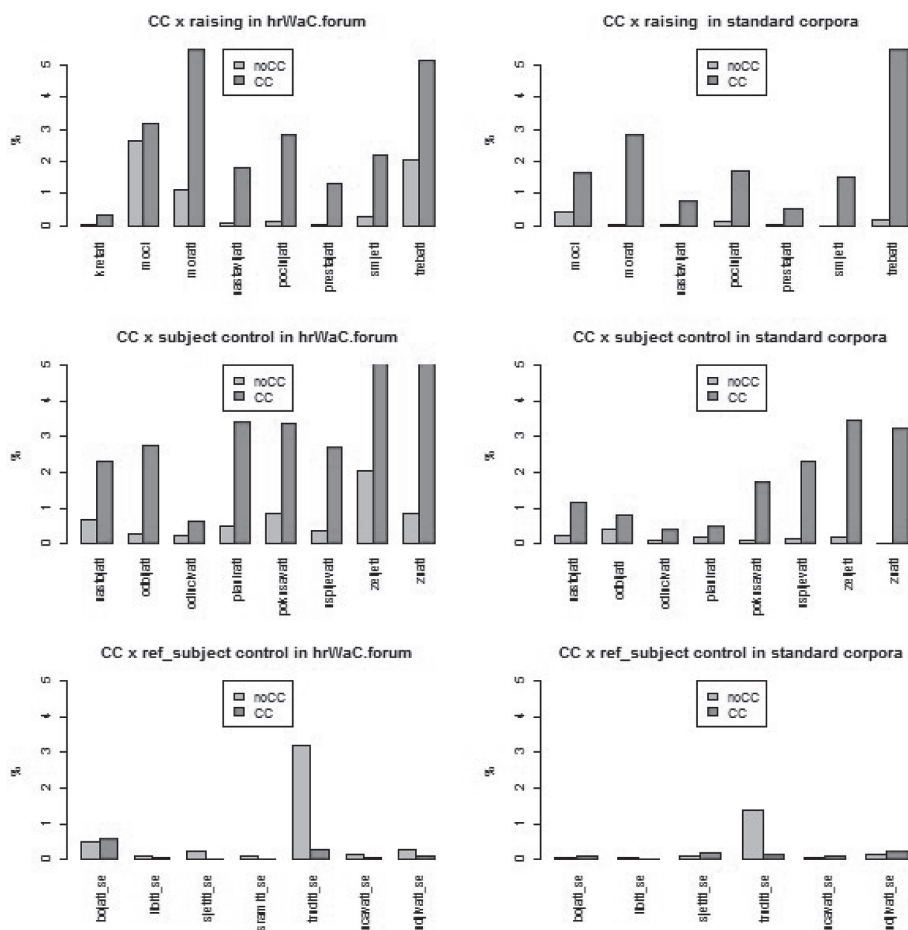
Figure 3. Verb-specific clitic positioning across different CTP types and corpora

In the study, we did not examine the relationship between the form of CL and CC. Nonetheless, in order to eliminate the possibility that the results are due to the type of CL and the case of the pronominal CL, we separately examined their distributions in the sample. We did not identify any significant difference between CL behaviour in sentences containing Raising or simple Subject Control CTPs. However, reflexive CLs of the infinitive complement in combination with reflexive Subject Control CTPs did not climb in any corpus.

7.2. Testing correlations

In order to statistically test the relationship between CC, CTP types and the corpus, we used a logistic regression model with CTP lexemes as random-effect. For computations, we used the generalized linear mixed model fit by maximal likelihood from the lme4 R-package (Bates 2010).

Random effects:

Groups Name	Variance	Std.Dev.
CtpVerb (Intercept)	0.4854	0.6967
Number of obs: 2327	groups:	CtpVerb, 23

Fixed effects:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept CC-Yes, Raising, Forum)	1.7498	0.2918	5.997	<0.001	***
simple Subject	-0.1784	0.3969	-0.449	0.6531	
reflexive Subject	-2.9451	0.5012	-5.877	<0.001	***
Standard	1.5790	0.2489	6.343	<0.001	***
simple Subject and Standard	-0.7625	0.3306	-2.307	0.0211	*
reflexive Subject and Standard	-0.7246	0.4978	-1.456	0.1455	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The results of the model confirm our preliminary observations – both the corpus and the difference between reflexive Subject Control and other predicate types influence the probability of CC occurring in a sentence. We will elaborate shortly on them.

The intercept in the model is CC occuring for Raising CTPs in the Forum and is used as a reference level for effects. The estimate of the intercept, which is log odd, can be recalculated to probability.[17] That is, the probability of CC oc-

---

[17] Using the formula: $P=e^{\log O}/(1+e^{\log O})$, where P-probablity, O-odds.

curring when a Raising CTP is used in colloquial Croatian is 0.85. The other estimates refer to the change of log odd, when particular effects are compared with the intercept. Thus, there is no substantial difference between Raising and simple Subject CTPs in colloquial Croatian, but CC with reflexive Subject Control CTPs occurs in colloquial Croatian with a probability of only 0.23.

The change from colloquial to standard Croatian is significant, and has a positive effect on CC in the presence of Raising CTPs. This means that the probability increases to 0.96. This increase likely applies also to simple Subject Control verbs, but the increase is significantly lower than for Raising CTPs, the probability of CC being only 0.94. The change from colloquial to standard Croatian has little impact on reflexive Subject Control verbs.

8. Conclusions

Returning to our research questions, we can conclude that both diaphasic variation and the Raising-Control Dichotomy do have an impact on CC from infinitive complements.

A1: The difference between Raising and simple Subject Control CTPs is statistically significant only in standard Croatian, while CC with reflexive Subject Control CTPs is significantly less frequent in comparison to the other two types of CTPs regardless of the diaphasic variation.

A2: Diaphasic variation is a significant factor influencing the probability of CC. Unlike in Romance languages, in Croatian CC appears more frequently in standard language than in colloquial Croatian.

These findings allow some tentative observations to be made which should feed into future research. Although in standard Croatian CC out of single infinitive complements appears highly probable with Raising CTPs, it does not seem to be absolutely obligatory (pace Aljović 2005, in accordance with Hansen; Kolaković and Jurkiewicz-Rohrbacher 2018); colloquial language in particular allows the lack of CC to a certain degree. Furthermore, our assumption that the differentiation between simple and reflexive Subject Control CTPs hitherto neglected in theoretical syntactical research on CL could actually shed new light on mecha-

nisms of CC was justified. Moreover, the significant avoidance of CC with reflexive Subject Control predicates requires further empirical investigation. As in the case of reflexive Subject Control CC inevitably leads to mixed clitic clusters, we might conjecture that we could be dealing with a strategy to avoid such mixed clusters. However, this needs further investigation.

References:

Aljović, Nadira. 2004. Cliticization Domains: Clitic Climbing in Romance and in Serbo-Croatian. *Proceedings of JEL'2004 Domain(e)s*. Eds. Crouzet, O. et al. Université de Nantes. 169–175.

Aljović, Nadira. 2005. On clitic climbing in Bosnian/Croatian/Serbian. *Lingvistički vidici* 34/5. 58–84.

de Andrade, Aroldo Leal. 2010. The Application of Clitic Climbing in European Portuguese and the Role of Register. *Selected Proceedings of the 12th Hispanic Linguistics Symposium*. Eds. Borgonovo, Claudia et al. Cascadilla Proceedings Project. Somerville, MA. 97–108

Bates, Douglas. 2010. *lme4: Mixed-effects modeling with R*. Springer.

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* 57. 39–44.

Browne, Wayles. 2004. Serbo-Croatian Enclitics for English-Speaking Learners. *Journal of Slavic Linguistics* 12/1. 249–283. (Reprint from 1975).

Cacoullos, Rena T. 1999. Construction frequency and reductive change: diachronic and register variation in Spanish clitic climbing. *Language variation and change* 11. 143–170.

Čamdžić, Amela; Hudson, Richard. 2002. Clitics in Serbo-Croat-Bosnian. *UCL Working Papers in Linguistics* 14. Eds. Harris, J. et al. Department of Phonetics and Linguistics University College London. London. 321–353.

Čolak, Majda. 2015. Domene, registri i žanrovi u hrvatskome jeziku. *Tabula* 13/1. 29–44.

Ćavar, Damir; Brozović-Rončević, Dunja. 2012. Riznica: The Croatian Language Corpus. *Prace filologiczne* 63. 51–65.

Davies, Mark. 1995. Analyzing Syntactic Variation with Computer-Based Corpora: The Case of Modern Spanish Clitic Climbing. *Hispania* 78/2. 370–380.

DAVIES, WILLIAM; DUBINSKY, STANLEY. (2004) *The grammar of raising and control: A course in syntactic argumentation*. Blackwell. Malden.

DIESING, MOLLY; ZEC, DRAGA. 2011. Interface effects: Serbian clitics. *Experiments at the Interfaces, Syntax and Semantics*. Volume 37. Ed. Runner, J. T. Emerald. Bingely. 1–30.

DIESING, MOLLY; ZEC, DRAGA. 2017. Getting in the first word: Prosody and predicate initial sentences in Serbian. *Glossa: A Journal of General Linguistics* 2/1. 1–25.

FRANKS, STEVEN; KING, TRACY H. 2000. *A Handbook of Slavic clitics*. Oxford University Press. Oxford.

HANA, JIRKA. 2007 *Czech Clitics in Higher Order Grammar*. PhD Thesis. Charles University in Prague.

HANSEN, BJÖRN; KOLAKOVIĆ, ZRINKA; JURKIEWICZ-ROHRBACHER, EDYTA. 2018. Clitic climbing and stacked infinitives in Bosnian, Croatian and Serbian – A corpus-driven study. *Grammar and Corpora 2016*. Eds. Fuß, E. et al. Heidelberg University Publishing. Heidelberg. 259–268.

IVIĆ, MILKA. 1972. Problematika srpskohrvatskog infinitiva. *Zbornik za filologiju i lingvistiku* 15/2. 117–138.

JUNGHANNS, UWE. 2002. Clitic climbing im Tschechischen. *Linguistische Arbeitsberichte* 80. 57–90.

JURKIEWICZ-ROHRBACHER, EDYTA; KOLAKOVIĆ, ZRINKA; HANSEN, BJÖRN. 2017. Web Corpora – the best possible solution for tracking rare phenomena in underresourced languages: clitics in Bosnian, Croatian and Serbian. *CMLC-5+BigNLP 2017*. Eds. Bański, Piotr et al. Institut für Deutsche Sprache. Mannheim. 49–55.

JURKIEWICZ-ROHRBACHER, EDYTA; HANSEN, BJÖRN; KOLAKOVIĆ, ZRINKA. 2017. Clitic climbing, finiteness and the raising--control distinction. A corpus-based study. *Jazykovedný* časopis 68/2. 179–190.

LANGSTON, KEITH; PETI-STANTIĆ, ANITA. 2014. *Language planning and national identity in Croatia*. Palgrave Macmillan. New York.

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.

POLANČEC, JURICA; MIHALJEVIĆ, ANA. 2016. Croatian non-standard invariant relativizers in European perspective. *Jezikoslovlje* 17/1-2. 429–451.

PROGOVAC, LJILJANA. 2005. *A Syntax of Serbian: Clausal Architecture.* Slavica Publishers. Bloomington.

REZAC, MILAN. 2005. The syntax of clitic climbing in Czech. *Clitics and affix combinations. Theoretical perspectives*. Eds. Heggie, Lorie; Ordóñez, Francisco. Benjamins. Amsterdam. 103–140.

Rosen, Alexandr. 2014. Haplology of Reflexive Clitics in Czech. *Slavic and German in Contact: Studies from Areal and Contrastive Linguistics*. Eds. Kaczmarska, E.; Nomachi, M. Slavic Research Center. Hokkaido. 97–116.

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

Stiebels, Barbara. 2015. Control. *Syntax – theory and analysis. An international handbook*. Eds. Kiss, Tibor; Alexiadou, Artemis. De Gruyter. Berlin. 412–446.

Stjepanović, Sandra. 2004. Clitic Climbing and Restructuring with "Finite Clause" and Infinitive Complements. *Journal of Slavic Linguistics* 12/1. 173–212.

Tadić, Marko. 1998. Raspon, opseg i sastav korpusa suvremenoga hrvatskoga jezika. *Filologija* 30-31. 337–347.

Uspon zanaglasnica, dihotomija podizanje – kontrola i stilska varijacija

Sažetak

Rad je posvećen usponu zanaglasnica iz infinitivnih dopuna u suvremenome hrvatskom jeziku. Autori u obzir uzimaju postojeća teorijska i empirijska istraživanja o usponu zanaglasnica u češkome i bosanskome/hrvatskome/srpskome te zapažanja da u nekim jezicima postoji veza između uspona zanaglasnica i razlike u registrima. U radu se donose odgovori na sljedeća dva istraživačka pitanja: „Postoji li razlika u usponu zanaglasnica iz infinitivne dopune u hrvatskome jeziku između rečenica s glavnim glagolima podizanja i kontrole? Je li uspon zanaglasnica jednako čest u hrvatskome standardnom jeziku i u njegovu razgovornom varijetetu ako se tip glavnoga glagola (podizanje – kontrola) kao čimbenik drži konstantnim?"

Istraživanje se temelji na podatcima za dva tipa glavnih predikata: a) podizanje (8 različitih glagola) i b) subjektna kontrola (8 nerefleksivnih + 8 refleksivnih glagola). Analiza se temelji na podatcima iz potkorpusa *Forum Hrvatskoga mrežnog korpusa* (hrWaC v2.2), *Hrvatske jezične riznice* i *Hrvatskoga nacionalnog korpusa*. Na temelju analizirane građe zaključujemo kako na uspon zanaglasnica utječe i razlika među glavnim glagolima (podizanje – kontrola) i varijacija u registru (hrvatski standardni jezik ili razgovorni varijetet).

Ključne riječi: infinitivne dopune, podizanje i kontrola, varijacija s obzirom na registar, uspon zanaglasnica, sintaksa, hrvatski
Keywords: infinitive complements, Raising-Control Dichotomy, diaphasic variation, clitic climbing, syntax, Croatian