

# Clitics in the wild

Empirical studies on the microvariation of the pronominal, reflexive and verbal clitics in Bosnian, Croatian and Serbian

Zrinka Kolaković

Edyta Jurkiewicz-Rohrbacher

Björn Hansen

Dušica Filipović Đurđević

Nataša Fritz

Open Slavic Linguistics 7



## Open Slavic Linguistics

Editors: Berit Gehrke, Denisa Lenertová, Roland Meyer, Radek Šimík & Luka Szucsich

In this series:

1. Lenertová, Denisa, Roland Meyer, Radek Šimík & Luka Szucsich (Eds.). *Advances in formal Slavic linguistics 2016*.
2. Wiland, Bartosz. *The spell-out algorithm and lexicalization patterns: Slavic verbs and complementizers*.
3. Marušič, Franc, Petra Mišmaš & Rok Žaucer (eds.). *Advances in formal Slavic linguistics 2017*.
4. Blümel, Andreas, Jovana Gajić, Ljudmila Geist, Uwe Junghanns & Hagen Pitsch (eds.). *Advances in formal Slavic linguistics 2018*.
5. Dočekal, Mojmír & Marcin Wağiel (eds.). *Formal approaches to number in Slavic and beyond*.
6. Wağiel, Marcin. *Subatomic quantification*.
7. Kolaković, Zrinka, Edyta Jurkiewicz-Rohrbacher, Björn Hansen, Dušica Filipović Đurđević & Nataša Fritz. *Clitics in the wild: Empirical studies on the microvariation of the pronominal, reflexive and verbal clitics in Bosnian, Croatian and Serbian*.

ISSN (print): 2627-8324  
ISSN (electronic): 2627-8332

# Clitics in the wild

Empirical studies on the microvariation of the pronominal, reflexive and verbal clitics in Bosnian, Croatian and Serbian

Zrinka Kolaković

Edyta Jurkiewicz-Rohrbacher

Björn Hansen

Dušica Filipović Đurđević

Nataša Fritz

Zrinka Kolaković, Edyta Jurkiewicz-Rohrbacher, Björn Hansen, Dušica Filipović Đurđević & Nataša Fritz. 2022. *Clitics in the wild: Empirical studies on the microvariation of the pronominal, reflexive and verbal clitics in Bosnian, Croatian and Serbian* (Open Slavic Linguistics 7). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/339>

© 2022, Zrinka Kolaković, Edyta Jurkiewicz-Rohrbacher, Björn Hansen, Dušica Filipović Đurđević & Nataša Fritz

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/> 

ISBN: 978-3-96110-336-2 (Digital)

978-3-98554-032-7 (Hardcover)

ISSN (print): 2627-8324

ISSN (electronic): 2627-8332

DOI: 10.5281/zenodo.5792972

Source code available from [www.github.com/langsci/339](http://www.github.com/langsci/339)

Collaborative reading: [paperhive.org/documents/remote?type=langsci&id=339](http://paperhive.org/documents/remote?type=langsci&id=339)

Cover and concept of design: Ulrike Harbort

Typesetting: Eberhard Gade, Edyta Jurkiewicz-Rohrbacher

Proofreading: Alexandr Rosen, Amir Ghorbanpour, Amy Amoakuh, Brett Reynolds, Cesar Perez Guarda, Christopher Straughn, Jeroen van de Weijer, Krystyna Kupiszewska, Marten Stelling, Jean Nitzke

Fonts: Libertinus, Arimo, DejaVu Sans Mono

Typesetting software: Xe<sub>La</sub>T<sub>E</sub>X

Language Science Press

xHain

Grünberger Str. 16

10243 Berlin, Germany

[langsci-press.org](http://langsci-press.org)

Storage and cataloguing done by FU Berlin

Freie Universität  Berlin

# Contents

<b>Acknowledgements</b>	<b>xi</b>
<b>List of abbreviations</b>	<b>xv</b>
<b>I Preliminaries</b>	<b>1</b>
<b>1 Introduction and overview</b>	<b>3</b>
1.1 Topic of the book . . . . .	3
1.2 Clitics and microvariation . . . . .	4
1.3 Empirical orientation . . . . .	8
1.4 Structure of the volume . . . . .	9
<b>2 Terms and concepts in the light of theoretical approaches to the study of clitics in BCS</b>	<b>11</b>
2.1 Introduction . . . . .	11
2.2 Clitics . . . . .	12
2.3 Systemic vs functional microvariation . . . . .	13
2.4 Parameters of microvariation . . . . .	16
2.4.1 Inventory . . . . .	17
2.4.2 Internal organisation of the clitic cluster . . . . .	18
2.4.3 Position of the clitic or the clitic cluster . . . . .	22
2.4.4 Clitic climbing . . . . .	32
2.4.5 Diaclisis and pseudodiaclisis . . . . .	34
2.5 Syntactic categories relevant for the description of microvariation	34
2.5.1 Complement-taking predicates . . . . .	34
2.5.2 The control vs raising distinction . . . . .	35
2.5.3 Types of complements . . . . .	40
2.5.4 Different types of reflexives . . . . .	41
<b>3 Empirical approach to clitics in BCS</b>	<b>51</b>
3.1 Introduction . . . . .	51

## Contents

3.2	From researchers' intuition to triangulation of methods . . . . .	53
3.2.1	Triangulation of methods . . . . .	53
3.2.2	Research validity . . . . .	54
3.3	Empirical approach in the current study . . . . .	55
3.3.1	Chosen strategy . . . . .	55
3.3.2	Corpus studies . . . . .	56
3.3.3	Psycholinguistic experiments . . . . .	59
3.3.4	Experiment chosen for our study . . . . .	66
<b>4</b>	<b>Corpora for Bosnian, Croatian, and Serbian</b>	<b>69</b>
4.1	Introduction . . . . .	69
4.2	Some remarks on corpus types . . . . .	69
4.2.1	The meanings of the term <i>corpus</i> in modern linguistics . . . . .	69
4.2.2	Types of corpora as text collections . . . . .	70
4.3	Overview of traditionally compiled corpora for BCS . . . . .	71
4.3.1	Bosnian corpora . . . . .	71
4.3.2	Croatian corpora . . . . .	72
4.3.3	Serbian corpora . . . . .	75
4.4	{bs,hr,sr}WaC . . . . .	77
4.4.1	The concept of the Web as a Corpus . . . . .	77
4.4.2	{bs,hr,sr}WaC in a nutshell . . . . .	78
4.4.3	WaC content . . . . .	79
4.4.4	Sources of noise . . . . .	80
4.5	Corpora for spoken BCS . . . . .	82
4.5.1	Bosnian . . . . .	82
4.5.2	Croatian . . . . .	82
4.5.3	Serbian . . . . .	83
4.6	Discussion . . . . .	83
4.6.1	The scope of available data . . . . .	83
4.6.2	Variation in spoken BCS . . . . .	85
4.6.3	Clitic climbing in BCS . . . . .	86
<b>II</b>	<b>Parameters of variation</b>	<b>87</b>
<b>5</b>	<b>Parameters of variation: Inventory, internal organisation of cluster, and position</b>	<b>89</b>

<b>6</b>	<b>Clitics and variation in grammaticography and related work (Bosnian, Croatian, Serbian)</b>	<b>91</b>
6.1	Introduction . . . . .	91
6.2	Detecting variation . . . . .	92
6.3	Inventory . . . . .	93
6.3.1	Inventory of pronominal clitics in BCS standard varieties	93
6.3.2	Inventory of verbal clitics in BCS standard varieties . . .	95
6.3.3	Reflexive markers <i>se</i> and <i>si</i> in BCS standard varieties . .	98
6.3.4	Polar question marker <i>li</i> . . . . .	99
6.4	Internal organisation of the clitic cluster . . . . .	100
6.4.1	Clitic ordering within the cluster in BCS standard varieties	100
6.4.2	Morphonological processes within the cluster . . . . .	103
6.5	Position of the clitic or the clitic cluster . . . . .	106
6.5.1	General remarks on clitic placement in BCS standard varieties . . . . .	106
6.5.2	Placement with respect to breaks in BCS standard varieties	106
6.5.3	Placement with regard to different types of hosts in BCS standard varieties . . . . .	109
6.5.4	Second position, second word and delayed placement . .	110
6.5.5	The limits of phrase splitting in BCS standard varieties .	114
6.6	Summary . . . . .	124
6.6.1	Clitic inventory . . . . .	124
6.6.2	Clitic cluster and morphonological processes within it .	124
6.6.3	Position of clitics or clitic cluster: second position . . . .	125
6.6.4	Second position, delayed placement and phrase splitting	125
<b>7</b>	<b>Clitics in dialects (Bosnian, Croatian, Serbian)</b>	<b>127</b>
7.1	Introduction . . . . .	127
7.2	An overview of BCS Štokavian dialects . . . . .	128
7.3	Available data . . . . .	132
7.3.1	Types of available data . . . . .	132
7.3.2	Data quality . . . . .	133
7.3.3	Examples in this chapter . . . . .	135
7.4	Inventory . . . . .	136
7.4.1	Pronominal clitics . . . . .	136
7.4.2	Verbal clitics: Aoristal/conditional clitics of the verb <i>biti</i> 'be' . . . . .	145
7.4.3	Reflexive clitic <i>si</i> . . . . .	145
7.4.4	Stress on clitics in BCS dialects . . . . .	147

## Contents

7.5	Internal organisation of the clitic cluster . . . . .	147
7.5.1	Clitic ordering within the cluster . . . . .	147
7.5.2	Morphonological processes within the cluster . . . . .	152
7.6	Position of the clitic or the clitic cluster . . . . .	153
7.6.1	Second position . . . . .	153
7.6.2	Delayed placement of clitics . . . . .	154
7.6.3	Phrase splitting . . . . .	156
7.6.4	Clitic first (1P) . . . . .	160
7.6.5	Endoclitics . . . . .	162
7.7	Clitic climbing . . . . .	164
7.8	Diaclisis . . . . .	166
7.9	Clitic doubling . . . . .	167
7.10	Summary . . . . .	168
7.10.1	Inventory . . . . .	168
7.10.2	Internal organisation of the clitic cluster . . . . .	169
7.10.3	Position of the clitic or the clitic cluster . . . . .	170
<b>8</b>	<b>Clitics in a corpus of a spoken variety (Bosnian)</b>	<b>173</b>
8.1	Introduction . . . . .	173
8.2	State of the art: Clitics in spoken BCS . . . . .	174
8.3	Research questions . . . . .	174
8.4	Corpus of Bosnian interviews . . . . .	175
8.4.1	Data quality . . . . .	175
8.4.2	Sociolinguistic features of the corpus . . . . .	177
8.5	Principles of analysis of spoken language . . . . .	177
8.6	Data preparation . . . . .	183
8.6.1	Annotation scheme . . . . .	183
8.6.2	Inventory-related categories . . . . .	183
8.6.3	Position-related categories . . . . .	185
8.7	Inventory . . . . .	188
8.7.1	Distribution of clitics in the corpus of spoken Bosnian . . . . .	188
8.7.2	Pronominal clitics . . . . .	189
8.7.3	Verbal clitics . . . . .	190
8.7.4	Reflexive clitics . . . . .	191
8.7.5	Clusters . . . . .	191
8.8	Internal organisation of the clitic cluster . . . . .	193
8.8.1	Clitic ordering within the cluster . . . . .	193
8.8.2	Morphonological processes within the cluster . . . . .	194



8.9	Position of the clitic or the clitic cluster . . . . .	196
8.9.1	General distribution . . . . .	196
8.9.2	Placement of single CLs . . . . .	196
8.9.3	Placement of clusters . . . . .	197
8.9.4	Relationship between the length of preceding phrases and clitic placement . . . . .	198
8.9.5	Phrase splitting . . . . .	205
8.10	Diaclisis . . . . .	208
8.11	Impact of certain syntactic structures on clitic placement . . . .	210
8.11.1	Impact of structures occurring before clitics . . . . .	210
8.11.2	Positioning of clitics within fluencemes . . . . .	212
8.12	Summary . . . . .	212
<b>9</b>	<b>Parameters of variation: Conclusions</b>	<b>215</b>
9.1	Introduction . . . . .	215
9.2	Parameters of microvariation: Global patterns . . . . .	217
9.2.1	Clitic inventory . . . . .	217
9.2.2	Clitic cluster and morphonological processes . . . . .	218
9.2.3	Absolute first position and clitics after the conjunctions <i>a</i> and <i>i</i> . . . . .	219
9.2.4	Second position, delayed placement and phrase splitting	220
9.2.5	Heaviness of the initial constituent . . . . .	220
<b>III</b>	<b>Clitic climbing</b>	<b>223</b>
<b>10</b>	<b>Approaches to clitic climbing</b>	<b>225</b>
10.1	Introduction . . . . .	225
10.2	Theoretical approaches to clitic climbing . . . . .	228
10.2.1	Definitions of clitic climbing . . . . .	228
10.2.2	Clitic climbing and optionality . . . . .	230
10.2.3	Clitic climbing, restructuring (or clause union) and movement . . . . .	231
10.2.4	Outlook . . . . .	235
<b>11</b>	<b>Constraints on clitic climbing in Czech compared to BCS: Theory and observations</b>	<b>237</b>
11.1	Introduction . . . . .	237
11.2	Island constraints . . . . .	238
11.2.1	Infinitives in comparative sentences with <i>než/nego</i> . . . .	238

## Contents

11.2.2	Clauses with inflected verbs . . . . .	240
11.2.3	Phrases with gerunds . . . . .	241
11.2.4	Adjective phrases . . . . .	242
11.2.5	Depth and kind of embeddedness of infinitive phrases .	244
11.2.6	Embedded wh-infinitives . . . . .	251
11.3	Constraints related to the raising–control distinction . . . . .	253
11.3.1	Object-controlled complements . . . . .	253
11.3.2	Object control constraint related to case . . . . .	255
11.3.3	Object control person-case constraint . . . . .	258
11.3.4	Object control and animacy of the referent of the clitic .	260
11.3.5	Object control reflexive constraint . . . . .	262
11.4	Constraints related to mixed clitic clusters . . . . .	263
11.4.1	Pseudo-twins . . . . .	263
11.4.2	Phonologically identical pronominal and reflexive clitics with different governors . . . . .	264
11.4.3	Morphologically different clitics with similar syntactic function and different governors . . . . .	267
11.5	How clitics climb . . . . .	270
11.5.1	Clitic cannot climb over clitic . . . . .	270
11.5.2	All-or-nothing constraint . . . . .	272
11.6	Sentential negation . . . . .	274
11.7	Constraints related to information structure . . . . .	275
11.7.1	Infinitive as a whole as the topic of a sentence . . . . .	275
11.7.2	Infinitive as a whole as the focus of a sentence . . . . .	276
11.8	Summary . . . . .	276
11.8.1	Overview . . . . .	276
11.8.2	Island constraints . . . . .	277
11.8.3	Constraints related to the raising–control distinction . .	279
11.8.4	Constraints related to mixed clitic clusters . . . . .	280
11.8.5	How clitics climb . . . . .	281
11.8.6	Sentential negation . . . . .	281
11.8.7	Constraints related to information structure . . . . .	281
11.9	Further Perspectives . . . . .	283
<b>12</b>	<b>Introductory remarks to corpus studies on clitic climbing</b>	<b>285</b>
12.1	Corpus-driven studies on clitic climbing . . . . .	285
12.2	Operationalising the constructions in question . . . . .	285
12.3	Choice of corpora . . . . .	290
12.4	Choice of matrix verbs . . . . .	291

12.5	Data collection . . . . .	295
<b>13</b>	<b>A corpus-based study on clitic climbing out of <math>da_2</math>-construction and the raising-control distinction (Serbian)</b>	<b>297</b>
13.1	Introduction . . . . .	297
13.2	The $da$ -complement and CC in Serbian . . . . .	298
13.3	Results . . . . .	300
13.4	Discussion: Constraints on CC from $da_2$ -complements . . . . .	304
13.5	Conclusions . . . . .	308
<b>14</b>	<b>A corpus-based study on clitic climbing out of infinitive complements in relation to the raising-control dichotomy and diaphasic variation (Croatian)</b>	<b>311</b>
14.1	Introduction . . . . .	311
14.2	Clitic climbing and diaphasic variation . . . . .	312
14.2.1	Clitic climbing and diaphasic variation in Romance languages . . . . .	312
14.2.2	Diaphasic variation in Croatian . . . . .	313
14.3	Research questions . . . . .	314
14.4	Methods . . . . .	314
14.5	Results and discussion . . . . .	316
14.5.1	Data distribution . . . . .	316
14.5.2	Testing correlations with a logistic regression model . . . . .	323
14.6	Conclusions . . . . .	325
<b>15</b>	<b>Experimental study on constraints on clitic climbing out of infinitive complements (Croatian)</b>	<b>327</b>
15.1	Introduction . . . . .	327
15.2	Research questions . . . . .	328
15.3	The test set-up . . . . .	332
15.3.1	Selection of matrix verbs . . . . .	333
15.3.2	Experiment design . . . . .	337
15.3.3	Stimuli . . . . .	341
15.3.4	Participants . . . . .	348
15.3.5	Recruiting participants . . . . .	350
15.3.6	Procedure . . . . .	350
15.4	Data analysis . . . . .	353
15.4.1	Data preprocessing . . . . .	353
15.4.2	Statistical analysis . . . . .	354

## Contents

15.5	Results: Regression model for acceptance rates . . . . .	355
15.6	Discussion . . . . .	362
15.6.1	Raising, subject and object control predicates . . . . .	362
15.6.2	Simple subject control predicates and subject control predicates with REFL <sub>LEX</sub> CL <i>se</i> . . . . .	366
15.6.3	Object control predicates with a pronominal CL controller in the dative and with a pronominal CL controller in the accusative . . . . .	368
15.6.4	Object control predicates with a REFL <sub>2ND</sub> CL <i>si</i> controller and with a REFL <sub>2ND</sub> CL <i>se</i> controller . . . . .	369
15.6.5	Object control predicates with a pronominal CL controller in the dative and with a REFL <sub>2ND</sub> CL <i>si</i> controller	371
15.6.6	Object control predicates with a pronominal CL controller in the accusative and with a REFL <sub>2ND</sub> CL <i>se</i> controller . . . . .	372
15.6.7	Reflexive subject control predicates and object control predicates with a REFL <sub>2ND</sub> CL <i>se</i> controller . . . . .	373
15.6.8	Pronominal and reflexive (REFL <sub>LEX</sub> CL <i>se</i> and REFL <sub>2ND</sub> CLs <i>se</i> and <i>si</i> ) infinitive CLs . . . . .	375
15.6.9	Dative and accusative infinitive CL complements . . . . .	378
15.7	Reaction time analysis . . . . .	382
15.8	Conclusions . . . . .	383
<b>16</b>	<b>On the heterogeneous nature of constraints on clitic climbing:</b>	
	<b>Complexity effects</b> . . . . .	<b>389</b>
16.1	Introduction . . . . .	389
16.2	Complexity . . . . .	390
16.2.1	The complexity of a system . . . . .	390
16.2.2	Ontological modes of complexity . . . . .	391
16.3	Systemic constraints related to the embedding type . . . . .	392
16.3.1	Islands . . . . .	392
16.3.2	Tied islands . . . . .	393
16.3.3	True islands . . . . .	393
16.3.4	Complexity effects in embeddings . . . . .	395
16.4	Systemic constraints related to the interaction of matrix and embedding . . . . .	396
16.4.1	Constraints in the light of empirical evidence . . . . .	396
16.4.2	Predicate type (CTP) . . . . .	398
16.4.3	CL type . . . . .	398

16.4.4	CL case . . . . .	400
16.4.5	Mixed cluster effects . . . . .	401
16.4.6	Complexity effects related to the interaction of matrix and embedding . . . . .	402
16.5	Non-systemic factors related to the diaphasic dimension . . . . .	405
16.6	Interaction, optionality and preferences . . . . .	406
<b>IV</b>	<b>Final remarks</b>	<b>409</b>
<b>17</b>	<b>Overall summary</b>	<b>411</b>
17.1	Scope of the book . . . . .	411
17.2	Empirical approach . . . . .	412
17.3	The role of prescriptivism in the BCS clitic systems . . . . .	413
17.4	Results: Parameters of variation . . . . .	415
17.4.1	Inventory . . . . .	415
17.4.2	Internal organisation of the clitic cluster . . . . .	415
17.4.3	Morphonological processes within the cluster . . . . .	416
17.4.4	Position of clitics and clitic clusters . . . . .	416
17.4.5	Constraints on clitic climbing . . . . .	418
	<b>Appendix A: Stimuli design</b>	<b>421</b>
	<b>Appendix B: Explanation of statistical measures</b>	<b>425</b>
	<b>References</b>	<b>427</b>
	<b>Index</b>	<b>455</b>
	Name index . . . . .	455



# Acknowledgements

This research was carried out within the project “Microvariation of the Pronominal and Auxiliary Clitics in Bosnian, Croatian and Serbian. Empirical Studies of Spoken Languages, Dialects and Heritage Languages”, in the years 2015 until 2019 generously supported by the German Research Foundation (HA 2659/6-1). Further, we received financial support from BAYHOST, DAAD and the women’s affairs officer at the Faculty for Linguistics, Literatures and Cultural Studies at the University of Regensburg as well as from the Faculty of Humanities of the University of Klagenfurt.

It took slightly longer to complete than we expected, and would not have been possible without the contribution of many people, whom we would like to thank here.

We are honoured to publish our results with a high quality open-access publisher. Thus, we wish to express our gratitude to the two anonymous reviewers who provided us with valuable comments enriching the contents of the book, and to editorial board members of *Open Slavic Linguistics*, especially Roland Meyer, Luka Szucsich, and Radek Šimik, and Sebastian Nordhoff from Language Science Press for words of encouragement and support while editing this book.

Our project benefited considerably from seminars on statistics, programming, natural language processing, and corpus linguistics organized by Nikola Ljubešić, Maja Miličević, and Tanja Samardžić on behalf of the Regional Linguistic Data Initiative (<https://reldi.spur.uzh.ch/>) financed by the Swiss Science Foundation.

Our workflow would not have been as smooth without our undergraduate student assistants Theodora Tiha Loos and Eberhard Gade, who performed the data extraction and annotation, and the manuscript conversion to  $\text{\LaTeX}$ . We are very grateful to have had the book proofread by Krystyna Kupiszewska, an amazing and patient translator. Branimir Brgles from the Department of Onomastics and Etymology at the Institute of Croatian Language and Linguistics in Zagreb plotted the dialectological maps for Chapters 7 and 9.

Naturally, we benefited from the feedback we received at conferences, round tables, seminars, and talks all over Europe. In particular, we owe thanks to Alexandr Rosen, Uwe Junghanns, Irenäus Kulik, Petr Karlík, and Anna Dušková for their invaluable comments and notes contributing to Chapters 10 and 11.

## *Acknowledgements*

We would like to recognise the impact of brainstorming with Václav Cvrček, Jana Pekarovičová, Petar Vuković, Alexandr Rosen, Jiří Hana, Pavel Kosek, Anna Dušková, Lenka Nerlich, Marek Nekula, participants of the round table *Language, culture and variation* held in Regensburg on 4<sup>th</sup> November 2016 and of the round table *Mechanisms and constraints on Clitic Climbing* held in Regensburg on 19<sup>th</sup> October 2017. It influenced our approach to clitics and variation outlined in Chapter 2 and the data gathering and annotation strategy, which were crucial for Chapters 8 and 15. We applied Pavel Kosek's advice on measuring heaviness of initial constituents in the data annotation for Chapter 8.

Moreover, we had the honour of presenting our project at the seminar *Jezi-koslovne rasprave* at the Institute of Croatian Language and Linguistics on 14<sup>th</sup> December 2017 and at the Department of Slavonic Studies of Humboldt Universität zu Berlin on 13<sup>th</sup> February 2019. Invaluable comments, suggestions and ideas for improvement made among others by Anita Peti-Stantić, Roland Meyer, Luka Szucsich, and Radek Šimík, who as an enthusiastic audience definitely had an impact on our work in Chapters 15 and 17.

Finally, we were always able to count on constructive discussion with the participants of the research seminar at the Department of Slavonic Studies of the University of Regensburg who followed our work on the project through all these years.

Jasmina Moskovljević Popović answered our questions regarding raising and control and gave us her monograph *Ogledi o glagolskoj potkategorizaciji* which was more than useful during work on Part III of our book dedicated to clitic climbing. Ivana Kurtović Budja, Željka Brlobaš, Ljiljana Kolenić, Anita Celinčić, Perina Vukša Nahod, and Milica Dinić Marinković provided us with dialectal data needed for Chapter 7. Thanks to professor Tilman Berger from the University of Tübingen we were able to work with the full corpus Bosnian Interviews (Stevanović 1999).

And last but not least, in the preparatory and final stages of the project we received useful comments from our consultants Petar Vuković and Petar Kehayov.

The corpus studies would not have been possible without two very supportive figures from the NLP world: Tomaž Erjavec and Nikola Ljubešić. They helped us solve various problems with CQL queries and other technical issues.

The speeded yes-no psycholinguistic acceptability judgment experiment with 336 native speakers as participants presented in Chapter 15 was a serious undertaking that would have been impossible if not for many colleagues who helped us find valuable contacts at different universities and faculties. These were, first and foremost, Ivana Brač, Tomislava Bošnjak Botica, Ana Ostroški Anić, and Siniša Runjić from the Institute of Croatian Language and Linguistics in Zagreb.



The experiment would not have been possible without great effort and enthusiasm of staff members at various institutions of higher education in Croatia. For organising the necessary permissions, providing quiet rooms, additional computers, and IT support for conducting the experiment, as well as for encouraging students to participate in the study, we would like to express our gratitude to:

- Boris Lazarević, Milan Poljak, and Klaudija Carović-Stanko from the Faculty of Agriculture, University of Zagreb,
- Ivan Botica, Zrinka Jelaska, Barbara Kušević, and Iva Nazalević Čučević from the Faculty of Humanities and Social Sciences, University of Zagreb,
- Tea Žakula from the Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb,
- Đuro Njavro, Robert Manestar, and Ivanka Rajh from the Zagreb School of Economics and Management,
- Marina Novak, Sanja Kiš Žuvela, Monika Jurić Janjik, Ivan Ćurković, and Petra Mitrović from the Academy of Music, University of Zagreb,
- Ivan Grgurević, Borna Abramović, and Jasmina Pašagić Škrinjar from the Faculty of Transport and Traffic Sciences, University of Zagreb,
- Tereza Rogić Lugarić from the Faculty of Law, University of Zagreb
- Mirko Bošnjak, Boženko Ćosić, and Ružica Rajšić, from the Student Centre, University of Zagreb,
- Krunoslav Zmaić, Siniša Ozimec, and Tihomir Florijančić from the Faculty of Agriculture, University of Osijek,
- Senka Blažetić, Irena Labak, Ljiljana Krstin, Kristina Mandić, Tanja Žuna Pfeiffer, and Mario Dunić from the Department of Biology, University of Osijek,
- Časlav Livada, Drago Žagar, Marijana Širić, Tomislav Matić, Mario Milož, Dalibor Buljić, Krno Miličević, Dario Došen, Ana Šokčević, Luka Kruljac, and Tomislav Lovrić from the Faculty of Electrical Engineering, University of Osijek,
- Mirjana Milić and Ana Penjak from the Faculty of Kinesiology, University of Split,

## *Acknowledgements*

- Josip Lasić, Marko Rimac, Renata Relja, Snježana Dobrota, Marita Brčić Kuljiš, Marija Lončar, Snježana Žana Dimzov, and Nataša Torlak from the Faculty of Humanities and Social Sciences, University of Split,
- Anđelko Domazet, Domagoj Runje, Željko Matas, and Danko Kovačević from the Catholic Faculty of Theology, University of Split.

Part of the participants were recruited in Rokovci-Andrijaševci during the 2017 Christmas holidays. This was possible thanks to Damir Dekanić, Ante Rajković, Martin Majer, Martina Markota, Blanka Vincetic, Sanja Uremovic, and Ana Koprta from the Rokovci-Andrijaševci municipality .

Finally, we can say without any exaggeration that it was not easy to deliver this book. We would not have succeeded in this task without the people who stood by us during the past few years. Being there for us took a lot of patience, care, love and tolerance. We owe our children, spouses, parents, parents-in-law, relatives, friends, and colleagues a debt of gratitude for their help and understanding.

# List of abbreviations

## Abbreviations used in the running text

1P	first position
2P	second position
2W	second word position
BCS	Bosnian, Croatian, Serbian
Bs	Bosnian
CC	clitic climbing
CL	clitic
CNC	Croatian National Corpus
CP	complementiser phrase
CQL	Corpus Query Language
CQP	Corpus Query Processor
CTP	complement-taking predicate
Cr	Croatian
Cz	Czech
DP	delayed position
DPh	host constituent in DP
DPi	initial constituent in DP
DSE	discourse structuring element
ECM	exceptional case-marking
EDU	elementary discourse unit
HrAL	Croatian Adult Spoken Language Corpus
IP	inflectional phrase
I-verbs	indicative selecting verbs
KSJ	Korpus srpskog jezika
MSD	morphosyntactic descriptions
NETK	Neetiketirani korpus savremenog srpskog jezika
NP	noun phrase
OCBT	Oslo Corpus of Bosnian Texts
PCC	person-case constraint
PP	prepositional phrase

## List of abbreviations

RT	reaction time
Sr	Serbian
SrpKor2003	Korpus savremenog srpskog jezika
S-verbs	subjunctive selecting verbs
TP	tense phrase
VP	verbal phrase
WaC	Web as Corpus
vP	verb shell

## Abbreviations used in glosses

ACC	accusative	PST	past
AOR	aorist	PTCP	participle
COND	conditional auxiliary	PASS	passive
DAT	dative	PL	plural
F	feminine	PRS	present
FUT	future auxiliary	Q	question particle <i>li</i>
GEN	genitive	REFL	reflexive clitic <i>se</i> or <i>si</i>
IMP	imperative	SG	singular
IMPERF	imperfect	PTCP.ADV.PRS	present adverbial participle
INF	infinitive		
M	masculine	PTCP.ADV.PST	past adverbial participle
NEG	negation		
N	neutral		

**Part I**

**Preliminaries**



# 1 Introduction and overview

## 1.1 Topic of the book

The present monograph is a data-oriented, empirical in-depth study of the system of clitics in Bosnian, Croatian, and Serbian.<sup>1</sup> Clitics are elements which, like affixes, cannot occur freely in a clause but need a host to lean on. The book deals with expressions such as those highlighted in the following examples (the hosts are *spremna* ‘ready’ in example (1) and *jučer* ‘yesterday’ in example (2)):

- (1) *Spremna sam mu se vratiti i oprostiti [...].*  
ready be.1SG him.DAT REFL return.INF and forgive.INF  
‘I am ready to return to him and forgive him [...].’ [hrWaC v2.2]
- (2) *Jučer ga je Marin snimao na središnjem terenu.*  
yesterday him.ACC be.3SG Marin record.PTCP.SG.M on central site  
‘Yesterday Marin was recording him on the central site.’ [hrWaC v2.2]

Since the seminal work *Über ein Gesetz der indogermanischen Wortstellung* by the Swiss linguist Jakob Wackernagel (1892), clitics have received continuous attention from linguists.<sup>2</sup> They are particularly well described in the Romance languages, (Ancient) Greek, and Czech, for instance.

Cross-linguistically, CLITICS (CLs) can be defined as “elements with some of the properties characteristic of independent words and some characteristic of affixes, in particular, inflectional affixes within words. Such elements act like single-word syntactic constituents in that they function as heads, arguments, or modifiers within phrases, but like affixes in that they are “dependent”, in some way or another, on adjacent words” Zwicky (1994: xii).

CLs are interesting for several reasons. First, they have a special phonological structure and they combine features of both syntactic words and affixes, thus blurring the boundary between the morphological and the syntactic system of the language. Second, in languages like Bosnian, Croatian, and Serbian (BCS),

---

<sup>1</sup>For detailed information on clitics see Section 2.2.

<sup>2</sup>Walkden et al. (2020) is a recent English translation of Wackernagel’s (1892) work.

## 1 Introduction and overview

which are usually claimed to have so-called free word order, allowing for positional permutations of phrases depending on information structure, CLs differ from other elements with a similar syntactic function in that their position is fixed. The placement of these CLs is usually associated with the left edge of the sentence, the so-called “second position”.<sup>3</sup> Third, some CLs have non-clitic counterparts which have the same meaning and syntactic function but differ as to word order. They are what Zwicky (1977: 3) calls SPECIAL CLITICS; i.e. unaccented bound forms which act as “variant[s] of a stressed free form[s] with the same cognitive meaning and a similar phonological makeup”. This is a hard nut to crack for functional frameworks which tend to explain structures by functional or cognitive mechanisms. The second position effect, however, seems to have a purely formal syntactic and/or prosodic basis. Nevertheless, formal approaches also have to struggle with the idiosyncratic word order behaviour setting CLs apart from other syntactic elements. A major problem is the ordering of the CLs in clusters, where verbal CLs show up in two different positions (CL *je* ‘is’ vs all other verbal CLs in BCS). As Franks et al. (2004: 12) argue, “the study of clitics can shed light on the interfaces between syntactic, morphological, and phonological linguistic representations.” We should add that they are also an ideal test case for usage-based approaches and/or explanations connected to the notions of repeated morphs, syntactic complexity and long-distance dependencies.

### 1.2 Clitics and microvariation

Our starting point is the observation that there is a high degree of variation in the CL system of Bosnian, Croatian, and Serbian. First, as argued by von Waldenfels & Eder (2016), these Neo-Štokavian varieties share a largely convergent grammatical, lexical, stylistic, and orthographic basis but show multiple types of variation of a multifactorial nature. Second, there seems to be relevant variation within varieties. We acknowledge that there is a considerable body of research dedicated specifically to CLs in BCS. However, the research on the syntax of Bosnian, Croatian, and Serbian is divided into works with a formal theoretical orientation on the one hand, and descriptive studies on the other. Considering this split, it comes as no surprise that in the literature we find largely contradictory statements concerning the acceptability of certain structures. Moreover, most authors rely exclusively on linguistic intuition and work with constructed examples. In footnotes, authors sometimes admit that the data they are discussing in order to develop

---

<sup>3</sup>For detailed information on the second position see Section 2.4.3.



certain theoretical claims are either marginal or are rejected right away by other native speakers.

A good illustration of disagreements as to the grammaticality of certain examples is provided by the question whether CLs can climb out of the so-called *da*-construction.<sup>4,5</sup> For instance, the formal linguist Stjepanović (2004: 174, 197) argues that semi-finite *da*<sub>2</sub>-complements and infinitive complements allow climbing in a similar way, as in the following constructed example (3).<sup>6,7,8</sup>

- (3) Marija *ga* mora / želi da posjeti.  
 Mary him.ACC must.3PRS want.3PRS that visit.3PRS  
 ‘Mary has to/wants to visit him.’ (BCS; Stjepanović 2004: 174)

In contrast, Ćavar & Wilder (1994: 448) argue that clitic climbing out of finite complements is “blocked in all dialects” of BCS. Others, like Progovac (2005: 146), refer to individual variation in the sense that some speakers of Serbian do and others do not accept such sentences. All the above-mentioned authors rely exclusively on constructed examples. This indicates that we not only encounter conscious pre-selection of data best fitting the theoretical claims, but also have to deal with the question of data quality. We agree with Diesing et al. (2009: 60) who emphasise that current research on CLs “has [...] relied heavily on native speaker judgments that have been culled primarily from previously published work, or from interrogating native speaker linguists. While these are not uncommon methods in theoretical linguistics, it is well worth augmenting the database with other sources [...]”

We focus both on language-internally motivated variation (systemic microvariation) and on selected cases of sociolinguistic microvariation in the diatopic and the diaphasic dimensions. This distinction is meant to capture the fact that there are two basic types of conditioning factors: whereas systemic microvariation is triggered by purely language-internal factors, sociolinguistic microvariation in the narrow sense depends on features relating to space (diatopic dimension:

<sup>4</sup>For basic information on clitic climbing see Section 2.4.4 and for thorough theoretical and empirical data on clitic climbing out of *da*<sub>2</sub> and infinitive complements see Part III.

<sup>5</sup>For basic information on *da*-complements see Section 2.5.3, and for detailed information on CC out of *da*-complements based on empirical evidence see Chapter 13.

<sup>6</sup>We use the abbreviations Cr, Sr, and Bs for the varieties Croatian, Serbian, and Bosnian, respectively.

<sup>7</sup>Stjepanović uses the label Serbo-Croatian.

<sup>8</sup>In example (3), the pronominal CL *ga* ‘him’, which is generated by the *da*<sub>2</sub>-complement *posjeti* ‘visits’, climbs out and appears in the matrix clause, to the left of the matrix predicate *mora* ‘must’ and/or *želi* ‘wants’.

## 1 Introduction and overview

three standard languages, dialects), or to the modes of language use (diaphasic: e.g. standard vs non-standard, written vs spoken language). We do not deal with variation in the language use of social groups (diastratic dimension). Throughout the book we use the terms variation and microvariation interchangeably.

One example of sociolinguistic microvariation concerns the CL of the third person feminine accusative pronoun: whereas the Croatian handbook *Hrvatski jezični savjetnik* by Barić et al. (1999: 173) recommends the form *je*, Ham et al. (2014: 74) favour *ju*; compare the examples in (4).

- (4) a. Vidjela *sam je*.  
see.PTCP.SG.F be.1SG her.ACC  
'I saw her.' (Cr; Barić et al. 1999: 173)
- b. Vidim *ju*.  
see.1PRS her.ACC  
'I see her.' (Cr; Ham et al. 2014: 74)

We are interested both in the prescriptive norms of the three standard languages, Bosnian, Croatian, and Serbian, and in real language usage as found in web corpora and in dialects, that is, observable language data.<sup>9</sup> The main focus is on the dominant varieties based on Neo-Štokavian, but we also allow for short side-glances at other dialects like Old-/Middle (Torlak) Štokavian, Kajkavian and Čakavian. Throughout this book, we use the label Bosnian/Croatian/Serbian (BCS) to refer to the Štokavian language usage common to the varieties used in Croatia, Serbia, and Bosnia-Herzegovina. We do not examine standard language use in Montenegro because first, efforts to create a standard for the variety spoken there are still in their infancy and, second, there are considerably fewer resources and specific studies. When we refer to language structures as codified in national handbooks we use the single label: Croatian, Serbian or Bosnian. The same holds for language usage patterns found in web corpora; i.e. in texts from the top-level domains .hr, .sr, and .ba. 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.

As the focus of the present study is microvariation within the CL system of Serbian, Croatian, and Bosnian and not a cross-linguistic typology of CL systems, we are quite cautious with regard to data and findings from other languages. We agree with Rosen & Hana (2017: 4) that in many respects, CLs in

---

<sup>9</sup>The type, quality, and quantity of data available for different variants of BCS vary considerably. We discuss these topics in Chapter 4 as well as in Sections 7.3 and 8.4.1.

different languages are similar with regard to inventory, positioning, and internal order (CL clusters), but the mix of properties in each language may be unique. Therefore, we will focus on Serbian, Croatian, and Bosnian data and refrain from conjectures concerning larger groups of languages. Thus, we will not comment on South Slavic or Slavic in general. Readers with a particular interest in contrastive studies can consult the existing, rich literature: the general handbook of Franks & King (2000) on systems of Slavic CLs, Božović (2021) on clustering phenomena in South Slavic, and Migdalski (2016) on second position cliticisation in Slavic, which emphasises the diachronic perspective. Although we acknowledge findings and theoretical insights on CLs in languages including Russian, Polish, Bulgarian, Slovene, and Portuguese, we will use these data only in exceptional cases to generate research hypotheses. In the case of Portuguese, we have come across studies showing that register may have a significant effect on clitic climbing. This observation was used to generate research hypotheses for our corpus-based study on clitic climbing in infinitive complements in relation to the diaphasic variation presented in Chapter 14. No comparison with clitic climbing in Portuguese is offered. Similarly, we consider the linguistic material from Russian and Bulgarian, e.g. in Landau (2000, 2004, 2013), to be irrelevant for our study because Russian CLs differ fundamentally from Serbian, Croatian, and Bosnian CLs with respect to inventory (no CL pronouns, no CL reflexive), position (second position not obligatory), and cluster formation (not present). Polish does have CL pronouns and a reflexive, but these allow both second position and verb-adjacent position. Further differences include the presence of the conditional CL *by* and past tense endings, but no present tense forms of the copula/auxiliary are available. Moreover, there are no CL clusters in Polish. Bulgarian is utterly different as it shows CL doubling, while Slovene has proclitics. We make an exception for Czech, taking it into consideration in the case of clitic climbing, which is exceptionally well described for this language. The Czech CL system is highly comparable indeed as it shares with BCS its CL inventory (verbal, reflexive, and pronominal CLs) and some other crucial features such as cluster formation, second position effects, and morphological processes within the cluster.<sup>10</sup>

---

<sup>10</sup>Czech clitics can phonologically encliticise or procliticize (Lenertová 2001: 295 and citations therein), in contrast to BCS CLs. We think that this factor is irrelevant for the phenomenon of CC, which is observed in languages with phonologically diverse types of CLs. See Chapters 10 and 11 for further discussion on this matter.

### 1.3 Empirical orientation

The monograph offers an account of the range of language-internal and sociolinguistic microvariation in this component of the language system by integrating large amounts of data and findings from descriptive and prescriptive works on the one hand, and from theoretically oriented studies on the other. A selection of structures is tested in an array of corpus and experimental studies. Our aim is to bridge the gulf between fine-grained description and syntactic generalisation by putting traditional work by Croatian, Serbian, and Bosnian scholars on an equal footing with general linguistic studies with a purely theoretical orientation.

As to the empirical approach chosen, the current work is usage-based oriented and we use triangulation of methods: intuition/theory – observation – experiment.<sup>11</sup> The first step always involves a thorough analysis of the whole body of existing research literature, independently of the respective theoretical framework, which is quite unique in syntax research. We also systematically document the approaches advocated by the leading normativists of Croatian, Serbian, and Bosnian who frequently discuss or evaluate variants.

The state-of-the-art literature review shows that with the exception of a few studies the previously analysed data lack precise characteristics and descriptions of amount and origin, which calls into question their replicability.<sup>12</sup> We have not come across many studies on CLs in BCS combining the theoretical literature with either corpus or experimental evidence. Thus, we conclude that in contrast to our study, most of the previously undertaken efforts did not include the kinds of standard types of empirical evidence currently acknowledged in linguistics. Therefore, we believe it is necessary to verify the often contradictory theoretical claims against empirical data collected primarily from corpora – our first source of observations. Since corpora allow the application of statistical methods, some hypotheses can be verified already at this stage. A selection of hypotheses concerning factors determining variation in the usage of CLs, formulated on the basis of corpus material, are further tested in acceptability judgment experiment where the level of control can be adjusted for individual factors. We are convinced that corpora as recordings of natural language production can be supplemented with experimental data such as acceptability judgment data because they both provide evidence about syntax. However, they offer different kinds of evidence: while corpora reflect language production, acceptability data primarily reflect language comprehension. The corroborating results from studies using different

---

<sup>11</sup>For details on the empirical approach chosen see Chapter 3.

<sup>12</sup>Under “with exceptions” we refer to the studies of Diesing et al. (2009), Zec & Filipović-Đurđević (2017), and Diesing & Zec (2017); see Section 2.4.3.3.

kind of data and methods provide more insightful and reliable linguistic evidence in comparison to studies using only one type of data.

Our aim is to give an account of the range of variation encountered in the real usage of the CL systems of Bosnian, Croatian, and Serbian. We restrict ourselves to the three main types of CLs, namely pronominal, reflexive and verbal CLs, thus excluding the polar interrogative marker *li*. Finally, in our empirical studies we pay only marginal attention to the question of phrase splitting as this phenomenon has already been studied extensively elsewhere.<sup>13</sup>

## 1.4 Structure of the volume

The parts I, II, and III are the three main parts which form the core structure of this monograph.

Part I covers Chapters 1–4. Chapter 2 introduces the most important concepts and terms used in the monograph, presents the parameters of variation, and discusses the most influential works that examine BCS CLs within formal theoretical frameworks. Departing from theoretical approaches to CLs which are usually based on limited numbers of constructed examples, we decided to investigate the phenomena of interest empirically. Our approach is explained in detail in Chapter 3. In the subsequent Chapter 4 we first present electronically stored corpora that are easily accessible to the research community and then discuss which of them are the most suitable for our empirical studies.

As we explain in Chapter 5, Part II focuses on the parameters of microvariation identified in Chapter 2. The structure of Chapter 6 and Chapter 7 follows the parameters of variation identified in Chapter 2. In Chapter 6 the parameters of variation are explored in detail at the level of standard languages. The goal of that chapter is to identify possible diatopic variation between BCS standard varieties, i.e. between different standard languages. Furthermore, where information is available in the literature, we comment on diaphasic variation within one BCS (standard) variety. However, the process of identifying variation is based solely on descriptions in the literature. Nevertheless, Chapter 7 offers deeper analysis of the identified parameters of variation with respect to diatopic variation. Moreover, we elaborate on some factors of variation such as CL inventory, CL placement, and morphological processes within the CL cluster, based on the empirical data in Chapter 8.

The only parameter of variation which we do not examine in Chapter 6 is clitic climbing. The large number of mainly theoretical studies on CLs notwith-

---

<sup>13</sup>For basic information on phrase splitting see Section 2.4.3.5.

## *1 Introduction and overview*

standing, clitic climbing has not received much attention. Moreover, as a phenomenon it has also been overlooked in grammar books and related works written by native authors. This is why we decided to dedicate one whole part of the book to this topic. Therefore, drawing on empirical studies on clitic climbing in Czech, in Chapters 10–15 we study clitic climbing mechanisms in some detail and propose a series of constraints it is subject to. Finally, we offer an explanation for constraints on clitic climbing in terms of complexity in Chapter 16. Our data-driven study gives new insights into the understanding of clitic climbing achieved through probabilistic modelling. We hope that in the future this can also feed into existing formal theories of clitic climbing. Chapter 17 recapitulates the main findings and gives an outlook for further studies.

## 2 Terms and concepts in the light of theoretical approaches to the study of clitics in BCS

### 2.1 Introduction

The goal of this chapter is to present the most important terms and concepts used throughout the monograph in the light of existing approaches. As pointed out by Spencer & Luís (2012: 233), there are PHONOLOGICAL, MORPHOLOGICAL and SYNTACTIC APPROACHES to the study of CLs. In phonological approaches CL positioning is defined in terms of phonological phrasing, which often interacts with information structure. Morphological approaches treat CLs as morphological units, usually as a specific type of affix, whereas syntactic approaches define CLs as function words which are associated with specific syntactic positions. As we will see below, some authors propose MIXED APPROACHES combining, for example, phonological and syntactic rules.

In BCS, CLs have been studied by scholars of two major lines of research, which tend to ignore each other. On the one hand, CLs have been the subject of a large number of theory-driven studies by US-based linguists (for an overview, see Bošković 2000, 2004). On the other hand, some aspects of CLs have been discussed with respect to stylistic and prescriptive factors (e.g. Reinkowski 2001, Peti-Stantić 2007). Most theoretically oriented studies are attempts to explain the principles of second position (see Section 2.4.3) and CL ordering within the framework of a formal grammar theory (e.g. Radanović-Kocić 1988, 1996, Schütze 1994, Progovac 1996, Bošković 2000, 2004; see Section 2.4.2.1).

However, as mentioned in Chapter 1, our aim is to prepare a data-oriented, empirical in-depth study of variation in the system of CLs in Bosnian, Croatian, and Serbian. We mainly focus on systemic microvariation and on selected cases of sociolinguistic variation in the diatopic and the diaphasic dimensions.<sup>1</sup> These research aims mean that we are not a priori bound to a specific syntactic theory; we thus strive for descriptive category labels and terms that are maximally

---

<sup>1</sup>For more information on (micro)variation see Section 2.3.

compatible with different theoretical approaches. Therefore, our objective is to explore the range of variation of the CL system, which might or should inform future theoretical accounts.

The rest of this chapter is structured as follows: Section 2.2 offers a cross-linguistic definition of CLs, whose properties are exemplified on BCS language material. In Section 2.3 we present our approach to the terms systemic and functional (micro)variation and describe which features they refer to. Section 2.4 focuses on parameters of CL microvariation: inventory, internal organisation of the CL cluster, position of the CL or CL cluster, CC, diacsis, and pseudodiacsis. Syntactic categories relevant to the description of microvariation such as complement-taking predicates, complement types, and reflexive types are presented in Section 2.5.

## 2.2 Clitics

As already mentioned in Chapter 1, CLs can be defined as “elements with some of the properties characteristic of independent words and some characteristic of affixes, in particular, inflectional affixes within words. Such elements act like single-word syntactic constituents in that they function as heads, arguments, or modifiers within phrases, but like affixes in that they are “dependent”, in some way or another, on adjacent words” (Zwicky 1994: xii). CLs cannot bear an accent of their own and therefore need an accented word form, the so-called host, to form an independent syntactic word. In sharp contrast to affixes, CLs exhibit low selectivity towards their host, attaching to very different kinds of hosts (PROMISCUOUS ATTACHMENT). In the following examples the CLs attach to the personal pronoun *on* ‘he’ (1), the adverb *jučer* ‘yesterday’ (2) and the adjective *svakog* ‘every’ within the noun phrase (3):

- (1) [On] *ga je potvrdio.*  
 he him.ACC be.3SG confirm.PTCP.SG.M  
 ‘He confirmed it.’ [hrWaC v2.2]
- (2) [Jučer] *ga je Marin snimao na središnjem terenu.*  
 yesterday him.ACC be.3SG Marin record.PTCP.SG.M on central site  
 ‘Yesterday Marin was recording him at the central site.’ [hrWaC v2.2]
- (3) [Svakog] *ga je dana do studija vozio odani šofer*  
 every him.ACC be.3SG day to studio drive.PTCP.SG.M loyal chauffeur  
 Clifton [...].  
 Clifton  
 ‘Every day his loyal chauffeur Clifton drove him to the studio [...].’  
 [hrWaC v2.2]



In general, CLs can attach either to the left of a host or to its right; in the first case they are called enclitics, in the latter proclitics. For the sake of brevity, throughout this book we use the term clitic exclusively to denote enclitics.

As the focus of the present monograph is on parameters of variation, we will not discuss the plethora of existing approaches and definitions of the CL category. Instead, we refer to the handbooks by Spencer & Luís (2012) and Franks & King (2000), which offer thorough overviews of the state of the art in the field.

In the following, we will outline the most important terms and concepts used throughout the book.

## 2.3 Systemic vs functional microvariation

First and foremost, we need to clarify what we mean by the term microvariation in relation to the syntax of CLs. Variationist linguistics has been developing as an independent research paradigm in the wake of William Labov's pioneering work on the social stratification of English and has brought to the fore a large number of variationist studies on English, but has not yet gained firm ground in South Slavistics, where prescriptive attitudes prevail among scholars and where the field of variation is still dominated by descriptive dialectology. In work on South Slavic syntax, variation does not therefore play an important role.

According to Walker (2013: 440) linguistic variation can be, informally speaking, understood as “different ways of saying the same thing”. As a matter of fact, it is more challenging to demonstrate an instance of syntactic than of phonological variation. This is because the former involves the non-trivial question of whether the given grammatical variants really present different ways of saying the same things or whether there are fine semantic or functional differences between them (Walker 2013: 441). As Walker (2013: 442) points out, the crucial methodological step in variation analysis is, therefore, defining precisely what is understood under “the same thing”, that is, circumscribing the variable context where the speaker has a true choice between forms. This can be achieved by following a form-based or a function-based approach, depending on the type of variable and the purposes of the study. As we do not intend to discuss in any detail the assumption of form-meaning isomorphism (where exactly one form corresponds to one meaning and vice versa), we avoid the function-based approach. Instead, in the present work, we follow the form-based approach. Hence, in order to examine syntactic variants, we extract CL forms that alternate with each other in a single (i.e. non-complementary) variable context or that are used for a single,

## 2 Terms and concepts in the light of theoretical approaches

identical meaning (Walker 2013: 443).<sup>2</sup> Identical communicative functions may or may not be present. This can be exemplified by the following two sentences which show variation as to the position of the CL *ga* ‘him’. In these two sentences the variable context is the same: in the matrix clause, it is the same subject control complement-taking predicate *žel(j)eti* ‘wish/want’ complemented with the same complement type *da*.<sup>3,4</sup> We index complement-taking predicates and their respective CLs with 1 and complements and their respective CLs with 2.

- (4) a. Mila *želi*<sub>1</sub>        da *ga*<sub>2</sub>        *vidi*<sub>2</sub>.  
      Mila want.3PRS that him.ACC see.3PRS
- b. Mila *ga*<sub>2</sub>        *želi*<sub>1</sub>        da *vidi*<sub>2</sub>.  
      Mila him.ACC want.3PRS that see.3PRS  
      ‘Mila wants to see him.’ (BCS; Aljović 2005: 11)

For the purposes of the present study, we would like to distinguish between systemic and functional factors. This distinction is meant to capture the fact that there are two basic types of conditioning factors:<sup>5</sup> First, SYSTEMIC MICROVARIATION, which is defined as purely language-internal, i.e. as variation between a dependent variable (e.g. CL position) and an independent variable encoded in the linguistic context. Second, variation in the traditional sociolinguistic sense, which depends on features relating to space (DIATOPIC), to social groups (DIATRATIC) or to the modes of language use in different situations (e.g. oral vs written, DIAPHASIC).<sup>6</sup> The focus of the present monograph is on the range of systemic factors and only secondarily on the conditional sociolinguistic factors. In some places we discuss the link between the two, but we refrain from a systematic sociolinguistic variationist account. We are mainly interested in the range and the limits of microvariation determined by linguistic contexts. The choice between

---

<sup>2</sup>Notice that we do not understand the term *context* in the informal way as items (words or passages) which precede and follow the studied item. Instead, we treat it as a construct of variables held constant in the study.

<sup>3</sup>For more information on complement-taking predicates see Section 2.5.1.

<sup>4</sup>For more information on complement types in BCS see Section 2.5.3.

<sup>5</sup>See the discussion on syntactic variables in Romaine (1981). She emphasises that purely syntactic variables differ from phonological variables because the latter always imply a social or stylistic factor (cf. Romaine 1981: 15).

<sup>6</sup>This distinction, which most variationist linguists can probably agree on, goes back to Coseriu (1980: 111): „Es gibt nämlich in einer historischen Sprache zumindest drei Arten der inneren Verschiedenheit, und zwar: *diatopische* Unterschiede (d.h. Unterschiede im Raume), *diatratische* Unterschiede (Unterschiede zwischen den sozial-kulturellen Schichten) und *diaphasische* Unterschiede, d.h. Unterschiede zwischen den Modalitäten des Sprechens je nach der Situation desselben (einschließlich der Teilnehmer am Gespräch).“

variants is governed by what one might call variable rules of grammar. We follow Walker (2010: 141) who argues that for a full understanding of variation we have to take formal or structural considerations into account. It goes without saying that we are unable to cover the whole area of variation delineated by the three dimensions in question.

We present an in-depth empirical study on the syntactic microvariation in the area of CC, for which we identify structural factors (i.e. constraints).<sup>7</sup> Among the diatopic conditioning factors, we mainly deal with variation between the standard norms of Croatian, Serbian, and to a lesser degree of Bosnian, as described in publications relevant for language corpus planning like authoritative reference handbooks used in schools, universities, and in the media. Furthermore, we give a literature-based account of variation in the BCS dialects spoken in Serbia, Croatia, Bosnia and Herzegovina, Montenegro, and Kosovo.

It is well known that the three standards show major differences in their lexicons, which includes cases where one and the same lexical unit belongs to different diastatic or diaphasic layers of the language.<sup>8</sup> As to core grammar, the differences are much more subtle. Piper (2009: 547), who discusses the differences between the Croatian and Serbian standards, points out that both varieties or languages have the same parts of speech, grammatical categories, grammemes, and morphonological processes. Following Piper (2009: 542f) we can distinguish:

1. differences in the inventory of grammatical constructions (e.g. the reflexive impersonal construction with the second argument in the accusative),<sup>9</sup>
2. differences concerning variants of forms (e.g. dative forms of *ko/tko* 'who': *kome* in Serbian and *kome, komu* or *kom* in Croatian),
3. differences in the frequency of specific forms (short forms of the adjective in oblique cases),
4. differences in the stylistic value of specific forms (e.g. the preposition *u* plus genitive is perceived as archaic in Serbian but neutral in Croatian).

It is noteworthy that in his overview Piper (2009) mentions CLs as a feature dividing Serbian and Croatian. He observes that the frequency of specific forms and the stylistic values of specific forms vary. The Croatian usage of the impersonal reflexive and the CL form *si* are recognised as a difference in the inventory

---

<sup>7</sup>For basic information on CC see Section 2.4.4 below and for in-depth information on CC see Part III.

<sup>8</sup>A good reference source for this kind of difference is Samardžija (2015).

<sup>9</sup>For more information on the reflexive impersonal construction see Section 2.5.4 in this chapter.

of grammatical constructions. In addition, he also notes the difference with respect to phrase splitting, which “in modern standard Serbian [is] less common or felt as regionalism” (Piper 2009: 546).

As far as possible, we mark all examples in our study with abbreviations indicating the national varieties: Croatian (Cr), Serbian (Sr) and Bosnian (Bs). When an example has been culled from the web corpora we restrict ourselves to the corpus names hrWaC, srWaC, and bsWaC. It goes without saying that some authors stick to the glossonym Serbo-Croatian, for which we use the label BCS.

We also record variants discussed in the normative literature which do not gain approval as “good” or “correct” language use. These data can be interpreted as variants determined by diatopic, diastratic or diaphasic factors. Furthermore, we dedicate one chapter to the use of CLs in a spoken variety, specifically in Bosnian, taking into account the diaphasic dimension of variation. The diaphasic dimension is additionally addressed in a corpus study based on a web subcorpus containing texts from a Croatian forum. As our empirical approach is based on data from the literature, from web and oral corpora, and finally from psycholinguistic experiment, we have nothing to say about variation related to social factors.<sup>10</sup> This is a separate research question which would require a completely different research design.

We acknowledge that due to the lack of space and available human and language resources we are not able to study all three national variants of BCS with the same analytical depth.<sup>11</sup> Furthermore, not all investigated phenomena are equally common in all varieties. We therefore concentrate on varieties in which the most data for certain structures were available or easily accessible. The monograph thus has a certain bias towards Croatian.

## 2.4 Parameters of microvariation

In this section, we present the dimensions or parameters of variation relevant for the CL systems of BCS, and discuss previous approaches. Note that we understand the term parameters not in the sense of Universal Grammar but as a set of contexts and variables pertaining to CLs.

---

<sup>10</sup>Our empirical approach is presented in more detail in Chapter 3.

<sup>11</sup>We refer to resources such as available electronically stored and morphosyntactically annotated corpora of a sufficient size.

### 2.4.1 Inventory

An important parameter of (micro)variation is the inventory of CLs in the Bosnian, Croatian, and Serbian standard languages and their non-standard varieties. The inventory of CLs encompasses the following four types:

1. personal pronouns,
2. verbal CLs,
  - a) copula/past tense auxiliary *biti*,
  - b) conditional auxiliary,
  - c) future auxiliary,
3. the reflexive marker(s) *se* and *si*,
4. the polar question marker *li*.

As already mentioned in Chapter 1, in our project we cover mainly verbal, pronominal and reflexive CLs. We thus exclude proclitic elements like prepositions and only touch upon the polar question marker *li*. The question marker differs from pronominal, reflexive and verbal CLs in its syntactic function and lack of a nonclitic equivalent. Whereas we discuss the variation within each CL type in Chapters 6, 7, and in 8, in this chapter we discuss the reflexive marker in more detail because due to its multifunctionality it turns out to be an important factor in microvariation (see Section 2.5.4 below).

The BCS CLs, except the polar question marker *li*, belong to what since the seminal work by Zwicky (1977) has been called special clitics. This term has been contested and its usefulness has been called into question (see the discussion in Spencer & Luís 2012: 41–45). For our purposes, it suffices to point out that CLs in BCS have “a significantly different distribution from their non-clitic counterpart” (Franks & King 2000: 6). Whereas the full forms of verbs and personal pronouns can change their position in the sentence depending on information structure, the CLs in question have a much more fixed position in the sentence. An important feature setting CLs apart from their stressed counterparts is coordination, which is possible with the full forms but completely ruled out with CLs. There does not seem to be variation in this respect; see the sentence presented in (5a) and its permuted counterpart (5b):<sup>12</sup>

---

<sup>12</sup>We did not find a single instance of the string *i te i ga* in hrWaC, srWaC or bsWaC.

## 2 Terms and concepts in the light of theoretical approaches

- (5) a. Vara i [tebe] i [njega].  
cheat.3PRS and you.ACC and him.ACC
- b. \*Vara i te i ga.  
cheat.3PRS and you.ACC and him.ACC
- ‘(S/he) is cheating on both you and him.’ [hrWaC v2.2]

The question of placement concerns the internal organisation of CL clusters (see the next subsection) on the one hand and the position of the cluster in the sentence on the other.

### 2.4.2 Internal organisation of the clitic cluster

#### 2.4.2.1 Clitic ordering within the cluster

If several CLs occur in one clause, they usually occur in a cluster, i.e. “a string of clitics that neither allows insertion of non-clitic elements nor permutation of clitics, when they are contiguous” (Zimmerling & Kosta 2013: 181).<sup>13</sup> The CLs occupy a fixed slot within the cluster, available only to this particular CL or type of CL (Zimmerling & Kosta 2013: 182). All theoretical models face a significant problem in this relative order of CLs, since it does not correlate with any other ordering rule in BCS. Authors arguing in favour of a morphologically oriented approach to the ordering of CLs assume a morphological template similar to affix order within a word, whereas in syntactic approaches the linear order within the cluster is explained in terms of syntactic positions.

Bošković (2001: 63) claims that the generative syntactic account of CL order is more principled, since under this account CL order within the cluster matches the structural height or position of the CLs in the clause structure. He further argues that conversely, morphological template analysis merely provides a formal way of stating the idiosyncrasies of BCS CL ordering (Bošković 2001: 64). The proponents of the syntactic approach tend to seek general explanations relating the positions to universal syntactic heads. A major problem arises because ordering within CL clusters differs cross-linguistically. Even a closely related language like Czech shows different ordering (e.g. reflexive before pronominal).<sup>14</sup> A second typologically interesting feature of the ordering sequence is that unlike e.g. clusters in Romance languages it contains not only pronominal and reflexive

<sup>13</sup>For more information on diaclysis, i.e. situations where one CL can be in clausal second position (or in delayed placement), while an additional clusterising CL is placed to its right, see Sections 2.4.5, 7.8, and 8.10.

<sup>14</sup>See also the putative generalisations concerning CL ordering in Romance languages discussed in Heggie & Ordóñez (2005).

elements, but also verbal elements. In contrast, proponents of a morphological approach usually refrain from such generalisations and may explain a given pattern as “a caprice of history as any property of the language faculty” (Spencer & Luís 2012: 319).

Slightly revising the proposal in Franks & King (2000: 29), we argue in favour of the following ordering within the cluster for BCS:

$li > \text{VERBAL}^* > \text{PRON}_{\text{DAT}} > \text{PRON}_{\text{ACC}} > \text{PRON}_{\text{GEN}} > \text{REFL} > je$   
\* except  $je = \text{PRS.3SG}$  of *biti* ‘be’

In contrast to Franks & King (2000), we do not use the label AUX because the ordering sequence does not seem to distinguish between the copula and the auxiliary uses of the forms of *biti* ‘be’.

The most puzzling feature of this ordering sequence and without doubt a major challenge for any theory of BCS CL ordering is presented by the position of the present tense third person singular verbal CL *je* ‘is’. Unlike other verbal CLs it follows the pronominal (and reflexive) CLs, and appears in cluster-final position as a sort of outlier.<sup>15</sup> The verbal CLs are thus split between the left and the right periphery of the cluster.

In order to remain consistent with the above-mentioned idea that CL order should match the structural height or position of the CLs in the clause structure, Mišeska Tomić (1996), Progovac (2005), and Franks (2017) account for differences in the slot verbal CLs occupy in the cluster with the existence of distinct heads: one for *je*, lower than for pronominal CLs, and one for other verbal CLs, higher than for pronominal CLs.

Although this solution sounds very attractive, Bošković (2001: 126) presents examples against it. He bases his argumentation on the observation that it is possible to insert a constituent between the pronominal CLs and *je* and that in the case of verbal phrase ellipsis, verbal phrase fronting, and parenthetical placement *je* behaves like other verbal CLs and precedes pronominal (and reflexive) CLs.<sup>16</sup> This serves as evidence that *je* is higher in the syntax than pronominal CLs: that is, it does not diverge from other verbal CLs in terms of generation place, only its phonological form is the last to occur on the surface, after pronominal CLs.

<sup>15</sup>In standard BCS varieties the verbal CL *je* is omitted after the reflexive CL *se*, but this is not always the rule in non standard varieties, for more information see Sections 6.4.2.2, 7.5.2.2, and 8.8.2. Moreover, in non-standard varieties the reversed order where the CL *je* precedes the reflexive CL *se* is attested, see Sections 7.5.1 and 8.8.

<sup>16</sup>Interestingly, Franks (2017: 224) starts his explanation from exactly the opposite statement: nothing can be inserted between pronominal CLs and be.3sg CL *je* – but his example comes from Bulgarian.

## 2 Terms and concepts in the light of theoretical approaches

As to the reasons why *je* must be pronounced in the tail of the cluster, Bošković (2001: 130f) suggests that *je* is in the process of losing its cliticood.

Finally, Migdalski (2020) proposes a syntactic approach in which CL variants of *biti* are pure phi-feature bundles. In this approach *je* specifies only the number feature (which is also present in the participle structure), whereas other CLs also carry the person feature, which results in different projections (Aux<sup>0</sup> for *je* and T<sup>0</sup> for the others) and leads to different ordering of the CLs. This, however, does not explain why only *je* is affected and not *biti*.3PL *su*.

It must be pointed out that *je* behaves peculiarly also in other respects, and not only regarding the slot it occupies. It is morphologically different from other CLs since the cliticised form originates from the root and not from the ending as is the case for all other verbal clitics (as pointed out e.g. by Mišeska Tomić 1996). Secondly, it participates in the morphological processes of suppletion, omission, and haplology of unlikes (see the next section) to a far greater extent than some other verbal CLs.<sup>17,18</sup> In the current work we do not strive to explain the slot taken by *je* in the cluster, but we do take a closer look at morphological processes.

We distinguish SIMPLE CLUSTERS and MIXED CLUSTERS. In the former, CLs originate in one clause like in example (6):<sup>19</sup>

- (6) I stalno smo<sub>1</sub> mu<sub>1</sub> se<sub>1</sub> vraćali<sub>1</sub>.  
and constantly be.1PL him.DAT REFL return.PTCP.PL.M  
'And we kept returning to him.' [hrWaC v2.2]

In the latter, CLs originate in the matrix clause and in its infinitive or *da*<sub>2</sub>-complement, as in the case of clitic climbing. In the following example (7), in the cluster *si ga* the accusative pronoun *ga* 'him' depends on the embedded infinitive *ubiti* 'to kill' and the auxiliary *si* on *mogla* 'could'.

---

<sup>17</sup>For more information on omission see Meermann & Sonnenhauser (2016).

<sup>18</sup>This range of variation is hard to explain using purely formal approaches and would require a separate extensive study including a diachronic perspective. In fact, the explanation for the variation in the ordering and idiosyncrasy of the position of the verbal CL *je* within the cluster could be connected to the relative age of CLs, as suggested by Grickat (1972: 95; cf. also Zimmerling & Kosta 2013: 189). Pavlović (2013: 60) claims that in 12<sup>th</sup>–13<sup>th</sup> century Old Serbian vernacular texts the hierarchy of the CLs within the cluster was as follows: 1. the interrogative particle *li*, 2. the conditional forms of the verb 'be', 3. the dative pronominal CLs, 4. the accusative pronominal CLs, and 5. the present tense forms of the verb 'be'.

<sup>19</sup>In non-standard varieties various reversed orders of CLs within a cluster are attested: see Chapters 6, 7, and 8.



- (7) Kako  $si_1$   $ga_3$   $mogla_1$   $dati_2$   $ubiti_3$ ?  
 how be.2SG him.ACC can.PTCP.SG.F give.INF kill.INF  
 ‘How could you have him get killed?’ [hrWaC v2.2]

In some cases, CLs do not show up in a cluster but occupy separate positions (see Section 2.4.5 below).

#### 2.4.2.2 Morphological processes within the cluster

The ordering of CLs is not restricted only to the positioning of each CL: certain combinations of CLs within the cluster are subject to MORPHOLOGICAL PROCESSES. As Neeleman & van de Koot (2006: 685) note, many languages exhibit a resistance against accidental repetition of morphemes (REPEATED MORPH CONSTRAINT). One solution is the avoidance of such repetitions. In BCS, three types of such morphological processes can be found. The first is called SUPPLETION: either morpheme is associated “with a different realization, typically based on a subset or a superset of its features” (Neeleman & van de Koot 2006: 686). This is observed when the homophonous CLs, pronoun her.ACC *je* and be.3SG *je*, co-occur as in example (8a); the string *je je* is altered to *ju je*, see example (8b).<sup>20</sup>

- (8) a. \* On *je je* čekao u njihovoj ulici.  
 he her.ACC be.3SG wait.PTCP.SG.M in their street  
 b. On *ju je* čekao u njihovoj ulici.  
 he her.ACC be.3SG wait.PTCP.SG.M in their street  
 ‘He was waiting for her in their street.’ [srWaC v1.2]

The second type is identified for the co-occurrence of the reflexive *se* (PSEUDO-TWINS, Junghanns 2002: 79). In the example presented in (9) we have two lexical reflexive verbs (*bojati se* ‘be afraid’, *vratiti se* ‘return’), which would result in the repetition of *se*. Since only one reflexive CL *se* is present in the sentence, it is an instance of HAPLOLOGY, i.e. the deletion of one *se*:

- (9) Boji<sub>1</sub>  $se_{1+2}$  vratiti<sub>2</sub> u svoje rodno Cetinje [...].  
 fear.3PRS REFL return.INF in own birth Cetinje  
 ‘He is afraid to return to his hometown Cetinje [...].’ [hrWaC 2.2]

<sup>20</sup>The sequence *je je* can be labelled as incorrect only in standard BCS varieties, since it is attested not only in Štokavian, but also in Čakavian dialects; for more information and examples see Section 7.5.2.2.

## 2 Terms and concepts in the light of theoretical approaches

The third type involves the combination of the verbal CL *je* and the reflexive CL *se*. In this case *je* is deleted: see the example presented in (10b).<sup>21</sup> It is interesting to note that here the deletion affects phonologically similar but not identical morphs. This means that haplogy can occur when CLs are not phonologically identical (HAPLOGY OF UNLIKES, Rosen & Hana 2017).

- (10) a. ? *Idol se je nalazio u Brandenburg.*  
idol REFL be.3SG find.PTCP.SG.M in Brandenburg
- b. *Idol se nalazio u Brandenburg.*  
idol REFL find.PTCP.SG.M in Brandenburg  
'The idol was located in Brandenburg.' [srWaC v1.2]

Haplogy does not seem to affect the homophonous pronoun *her.GEN je*, which occurs in the reversed CL order *je se*: see the example presented in (11).

- (11) *Bojim je se.*  
fear.1PRS her.GEN REFL  
'I am afraid of her.' [bsWaC v1.2]

When it comes to such morphonological processes, CL clusters behave more like affixes than like words (Spencer & Luís 2012: 121f).

### 2.4.3 Position of the clitic or the clitic cluster

#### 2.4.3.1 Second position

As mentioned above, the CLs in BCS are special clitics. This means that they are subject to word order restrictions characteristic of this and only of this category. The single CL or CL cluster occupies what is frequently called the Wackernagel or second position in the clause (2P). There is a long debate on what 2P actually is.

CLs are positioned within a clause with respect to a constituent which serves as a host. In this book, we use the term 2P in a narrow sense as referring to the

---

<sup>21</sup>This type of morphonological process within the cluster is actually a feature of standard language, for more information on deletion or non-deletion of the verbal CL *je* in standard varieties, dialects, and spoken Bosnian see Sections 6.4.2.2, 7.5.2.2, and 8.8.2. Moreover, this type of so-called morphonological process does not have a purely morphonological nature. For instance, Ridjanović (2012: 564) shows that the verbal CL *je* which is a copula, will not be omitted, see Section 6.4.2.2. The partial syntactic nature of this constraint is also observable in the fact that it does not affect the homophonous pronoun *her.GEN je*, which occurs in the reversed CL order *je se*. The genitive pronoun *je* cannot be deleted since it is an argument, while the auxiliary verb *je* can be and to a high degree in standard varieties is deleted.

position after the first constituent of the clause. This covers the position after a full phrase and after a complementiser. Delayed placement may be triggered by so-called heavy phrases.<sup>22</sup> Generally speaking, CLs can attach to any type of phrase to which they bear or do not bear a syntactic relationship (promiscuous attachment, mentioned above). The possessive dative is a special case because it has a fixed position in the sentence: it either has to follow the noun/phrase denoting the “possessed entity” (12a) or it comes after the first stressed word in the phrase which it modifies (12b) (cf. Ridjanović 2012: 559); see our transformation of Ridjanović’s example (12c):

- (12) a. Starija sestra *mu* pjeva u horu.  
 older sister he.DAT sing.3PRS in choir
- b. Starija *mu* sestra pjeva u horu.  
 older he.DAT sister sing.3PRS in choir
- c. \*Pjeva *mu* starija sestra u horu.  
 sing.3PRS he.DAT older sister in choir  
 ‘His older sister sings in a choir.’ (Bs; Ridjanović 2012: 559)

This is a case of syntactic microvariation which is discussed in Section 8.9.

#### 2.4.3.2 Approaches to 2P effects: syntax, phonology and information structure

Many studies on BCS are attempts to explain the principles of 2P and CL ordering within the framework of Minimalism. The discussion essentially concerns the division of labour between syntactic structure on the one hand and phonology or prosody on the other. According to Bošković (2000), three different schools can be distinguished among generative models:<sup>23,24</sup>

1. The strictly syntactic approach explains 2P effects exclusively by syntactic mechanisms (e.g. Progovac 1996, 1993a, Franks 1997). In these accounts

<sup>22</sup>For more information on heavy phrases see the next Section 2.4.3.3.

<sup>23</sup>Recent developments include parametric approaches. Here we can mention the somewhat controversial suggestion of Runić (2014) and Bošković (2016) that 2P effects relate to the lack of a determiner phrase layer in languages and the lack of articles. This restriction seems to be too general and is criticised and modified by Migdalski (2021), who also provides an alternative (parametric) approach related to the loss of verbal morphology (Jung & Migdalski 2015, Migdalski 2016, 2020). As explained in Section 1.2 we refrain from both overall typological generalisations, and diachronical perspective. Thus, we do not discuss these two ideas in the current work.

<sup>24</sup>Although 2P cliticisation is rarely a topic of non-generative works, for Czech see Fried (1994).

it is argued that “clitic placement is a syntactic phenomenon and should be assimilated to other more familiar types of syntactic movement rules, rather than involving a special kind of phonological clitic placement operation. Clitics are syntactic entities—in particular, functional heads—and they move as such” (Franks 1997: 111). In Minimalism a clause is assumed to be headed by several functional projections, which hierarchically dominate the lexical projection of the verb. Accordingly, the CL is moved to the left-hand periphery in the syntax, where it leans to the right of the element that is in the so-called complementiser position. Progovac (1996: 412) argues that CLs move in syntax – their distribution is constrained not by phonological, but by syntactic principles. She claims that the strongest argument that the placement of CLs is sensitive to syntax/semantics comes from subjunctive-like complements (Progovac 1996: 422f). While in indicative-like complements CLs are strictly clause-bound and must attach to the local complementiser, in subjunctive-like complements CLs attach either to the local complementiser or to the matrix complementiser position.

2. The strictly phonological approach postulates that 2P is governed exclusively by phonological rules. This position is mainly represented by Radanović-Kocić (1988, 1996). According to this approach the target of the movement is not a syntactically defined constituent or syntactic position, but the intonational phrase (Radanović-Kocić 1996: 441).<sup>25</sup> Nevertheless, bear in mind that even in her so-called “strictly phonological approach” the position of the CL or the CL cluster is hard to explain only within the domain of phonology. As we show in Section 6.5.5, Radanović-Kocić (1988) uses syntax to explain variation and constraints on phrase splitting. For instance, she argues that whether CLs are placed after the first word of a two-word initial subject or after the whole phrase depends on the structure of the subject phrase (cf. Radanović-Kocić 1988: 112). Further, she claims that there is an important difference between initial two-word subject phrases and non-subject phrases, and concludes that only subject phrases can be split, whereas others cannot (Radanović-Kocić 1988: 111). Hence, it is more than obvious that in her “strictly phonological approach” Radanović-Kocić (1988) uses syntax to explain the limits of phrase splitting in BCS.

---

<sup>25</sup>Čavar & Wilder (1994: 441) argue against a purely phonological approach to the 2P phenomenon. In their view it is not desirable to assume phonological rules which have the power to move material around in phonological representations in order to capture marginal cases like phrase splitting (cf. Čavar & Wilder 1994: 441).

3. In mixed approaches, both the syntactic and the phonological components of the language system are responsible for the positioning of the CL. For example, Schütze (1994) assumes that the CL is moved by syntax, but specific contexts also permit phonological movements (weak syntax approach). In contrast, Bošković (2000, 2001) assigns the dominant role to phonology. According to the so-called weak phonology approach, movement takes place in the syntax, but in addition a phonological filter is in operation. In other words, the 2P is actually a constraint on phonological form representations which filters out all constructions where CLs are found in any other position of their intonational phrase than the second (Bošković 2000).

Within the mixed approach Ćavar & Wilder (1994: 431) treat CL forms as both syntactic CLs and phonological enclitics. More specifically, they consider CL placement in Croatian to be a syntactic process (Ćavar & Wilder 1994: 431). According to them the 2P effect can be best accounted for syntactically. However, they attribute the ill-formedness of the 1P to a phonological (prosodic) property of CLs (Ćavar & Wilder 1994: 431). According to Wilder & Ćavar (1994a) and Wilder & Ćavar (1994b) the CL 2P effect results from the interaction between a syntactic CL placement rule and a phonological filter.

A similar view is presented in Franks (2000). Second position CLs as verbal features on their way up the verbal extended projection form a syntactic cluster which ends up in the highest functional position of the clause (Franks 2000). If syntax leaves CLs without a proper host, a lower copy of the CL cluster is pronounced. In other words, phonological form plays a filtering role (Franks 2000). Similar claims can be found in Bošković (1995: 264): syntax proposes structures to phonology, which discards some syntactically well-formed structures since they violate certain phonological form requirements. In other words, the role of phonology is to filter out the output of the syntactic component (Bošković 1995: 264). This is actually in contradiction to Bošković (2000), where it is argued that no special syntactic procedure is involved in CL placement.

It is worth noticing that there is one major problem with the phonological and mixed approaches. Namely, as Diesing et al. (2009: 70) point out, the advocates of the idea of an intonational phrase do not provide any experimental acoustic evidence for the postulated pauses or intonation units. Moreover, as Ćavar & Seiss (2011: 136) put it, “all these accounts have in common that they cannot motivate or explain the intra-linguistic variation, i.e. the alternations of the different constructions”. Finally, as Zec & Filipović-Đurđević (2017: 175) observe, regardless

## 2 *Terms and concepts in the light of theoretical approaches*

of the theoretical frame of reference, only main clauses with initial arguments have been investigated.<sup>26</sup> The question is whether this somewhat impoverished empirical landscape can indeed give valid formal accounts of the bifurcation into two 2P types: 2W and after the first phrase (Zec & Filipović-Đurđević 2017: 175).

A factor which might be worth studying in more detail in future is INFORMATION STRUCTURE which, however, has not received much attention in the existing placement analyses (cf. Ćavar & Seiss 2011: 134, Diesing et al. 2009: 71f). Several authors (e.g. Diesing et al. 2009, Diesing 2010, Ćavar & Seiss 2011, Zec & Filipović-Đurđević 2017, Diesing & Zec 2011, 2017) discuss the possibility that information structure may have an influence on the positioning of CLs in simple clauses (see below).

In Diesing et al.'s (2009) acceptability judgment experiment, object argument sentences were more likely to be accepted with the CL after the first word when the first word was a demonstrative. The difference in the acceptance rate between split object argument constituents with adjectives and demonstratives was statistically significant (Diesing et al. 2009: 69). Based on the reported findings, Diesing et al. (2009: 69f) believe that the preferred status of demonstratives over adjectives as first word CL hosts suggests potential differences in information structure. This conjecture is based on the status of demonstratives as deictic and/or specific determiners in languages that do not otherwise have determiners (Diesing et al. 2009: 70). More specifically, they argue that it is more likely for a demonstrative than an adjective to be a point of contrast in Serbian (cf. Diesing et al. 2009: 70). Diesing et al.'s (2009) hypothesis is "that clitic positioning is an interface phenomenon, in the broadest sense of the term, with at least prosody, syntax, and information structure contributing to the selection between the competing configurations" in both the argument- and predicate-initial main clauses. This is further elaborated on in Diesing & Zec (2017: 13) with the conclusion that in the predicate case, prosody alone is responsible for the selection of hosts for 2W placement, while in the argument case, prosody interfaces with syntax and the information structure in the selection of hosts for 2W placement.

Ćavar & Seiss (2011: 139) explicitly argue that different word order positions of CLs are related to differences in their specific information theoretic properties. More specifically, they claim that both 2P types, i.e. 2W and 2P after the first phrase, can be best explained in purely syntactic terms (Ćavar & Seiss 2011: 133). In their approach, the assumed cases of phonological CL placement in the 2W type of placement are analysed as instances of split constituent constructions (Ćavar & Seiss 2011: 133, 136, 141). According to Ćavar & Seiss (2011: 145) the CL or CL cluster always attaches after the first syntactic constituent, which in information structure terms can be a topic or a contrastive focus. If the first syntactic

---

<sup>26</sup>Browne's (1975) detailed description is the only exception to this.

constituent is a split part of a syntactic constituent, it triggers a contrastive focus reading and consequently requires a specific intonational contour (Āavar & Seiss 2011: 145). In other words, according to Āavar & Seiss (2011: 133) word order variation is related to information structure: it implies scope differences in a hierarchical (syntactic) representation and not the scope-neutral phonological processes. In their analysis the prosody-syntax interface remains quite simple, since they do not utilize complex word rearrangement mechanisms outside of syntax, or at the level of phonological representation (Āavar & Seiss 2011: 133).

Avgustinova & Oliva (1995) discuss CL positioning in Czech and propose an explanation for 2P which is based on the approach to the communicative structure of the sentence proposed among others by Sgall et al. (1986). According to this approach, the first position is defined as “preceding lexical material as a single substantial communicative segment” (Avgustinova & Oliva 1995: 25).

A typological approach to CLs which combines syntactic and morphological features with information structure has been elaborated mainly by Anton Zimmerling on the basis of data from various Slavic languages. It is based on three principles:

1. the Template Principle,
2. Constituency Conditions predicting the choice of the CL host,
3. Barrier Rules generating derived word orders with clusterising CLs (Zimmerling & Kosta 2013: 203).

Zimmerling & Kosta (2013: 194) argue that the description of word order systems of clausal CLs “should base on syntactic constraints and be maximally independent from conjectures about restrictions imposed by allegedly purely phonetic or lexical properties of clitics”. In languages like BCS a class of clause-level CLs form ordered clusters which, following Franks & King (2000), are defined as “contiguous strings of clitics arranged in a rigid order according to language-specific rules called ‘Clitic Templates’”. A cluster is understood as “a string of clitics that neither allows insertion of non-clitic elements nor permutation of clitics, when they are contiguous” (Zimmerling & Kosta 2013: 181). Clusters are formed according to rules that are independent from other rules of ordering; in this sense, they arrange elements in an idiosyncratic order. The authors propose what we could call a BARRIER-TEMPLATE THEORY. They claim that this theory, introduced by Zaliznjak (1993: 287), is the only approach which explains delayed placement of clusters and diaclisis by one and the same underlying mechanism. Basically, CLs in BCS have a fixed position in the clause, i.e. they attach to the clause-initial element (2P). The authors note that this, however, holds only

for communicatively unmarked sentences, and thus they integrate information structure into their model. There are two main deviations from this basic 2P order. First, the whole CL cluster can end up to the right of clausal 2P (this corresponds to what we have labelled delayed placement). Second, some clusterising CLs remain in clausal 2P, while other clusterising CLs end up to the right of it (we use the term *diaclysis*) (Zimmerling & Kosta 2013: 196). The main hypothesis is that the sentence-initial phrase hosting the CLs may have properties of a barrier and move all or some clusterising CLs to the right of clausal 2P. The first option is referred to as a “blind” or “indiscriminating” barrier, the second option is referred to as a “selective” barrier (Zimmerling & Kosta 2013: 196): “[...] in 2P languages sentence-initial Barriers are either blind and move all clusterizing clitics *n* steps to the right of clausal 2P or selective and split the clusters by moving some clusterizing CLs *n* steps to the right of clausal 2P” (Zimmerling & Kosta 2013: 197). Both blind and selective barriers can be optional or obligatory. Furthermore, the authors distinguish communicative and grammaticalised barriers. Communicative barriers are phrases that affect the position of CLs due to the communicative status they acquire in a given sentence (Zimmerling & Kosta 2013: 198).

#### 2.4.3.3 Barriers and delayed position of clitics

A second type of placement is when for some reason the initial phrase(s) is not selected as the host and the CL cluster attaches to a phrase further to the right in the sentence. This phenomenon is sometimes referred to as “delayed clitic placement” (Zec & Inkelas 1990), “clitic third” (Čavar & Wilder 1994, Schütze 1994), “late placement of clusters” (Zimmerling & Kosta 2013), “Endstellung” (Reinkowski 2001) or “resumptive RSC” (Rhythmic Structure Constituent) (Alexander 2008, 2009). As there are cases like in example (13) where the CL attaches not to the second but even to the third phrase, we prefer the broader term *DELAYED POSITION* (DP).

- (13) [Pod uvjetima iz stavka 1. ovoga članka]<sub>phrase1</sub> [pravna  
under conditions from paragraph 1 this article legal  
osoba]<sub>phrase2</sub> [kaznit]<sub>phrase3</sub> će se za kaznena djela  
person punish.INF FUT.3SG REFL for criminal acts  
propisana Kaznenim zakonom [...].  
regulated.PASS.PTCP criminal law  
'Under the conditions from paragraph 1 of this article a legal entity will  
be punished for criminal acts prescribed by Criminal Law [...].'  
[hrWaC v2.2]



According to Ćavar & Wilder (1994: 439) delayed placement appears only in embedded infinitives and in main, i.e. root clauses, and is not found in subordinate clauses.<sup>27</sup>

Without taking an a priori stance as to the structural or functional nature of the DP of CLs or as to whether we are dealing with exceptions to 2P, we use the term BARRIERS as a descriptive label for the preceding phrases. There are two main divergences from the basic order:

1. delayed placement of clusters,
2. diaclisis.

According to Zimmerling & Kosta (2013: 196) in DP “the whole clitic cluster ends up to the right of clausal 2P”. We leave open the question whether in BCS it is the barriers that move a CL *n* steps to the right of the CL host or whether some other mechanism is in play.

As we show in Chapter 6, normative grammar handbooks tend to argue that CLs cannot or should better not be placed after phrases separated by a comma (cf. Reinkowski 2008: 132). For such instances Radanović-Kocić (1996: 435), who proposes a purely prosodic account of 2P, suggests the term “heavy constituent”. It is worth pointing out that Radanović-Kocić (1996) does not provide a precise definition of the “heavy constituent” concept. A similar observation on DP can be found in Bošković (1995: 264), where it is claimed that when the constituents preceding a CL within a clause are heavy, the CL does not have to occur in the 2P of its clause. Following Schütze (1994), Bošković (1995: 264) adds that phonologically heavy constituents such as preposed PPs form separate intonational phrases and as such they are followed by an intonational phrase boundary. However, an inspection of the theoretical literature suggests that the situation is not so clear cut. Zec & Inkelas (1990: 373) argue that the “p-constituent is heavy iff it branches”. They elaborate on branching conditions and claim that branching at the syntactic constituent level is neither a sufficient nor a necessary condition for heaviness and delayed placement (Zec & Inkelas 1990: 374f). This is exemplified with the help of the following two sentences:

- (14) a. \*Sa Petrom razgovarala je samo Marija.  
 with Peter talk.PTCP.SG.F be.3SG only Mary  
 Intended: ‘To Peter, only Mary spoke.’

---

<sup>27</sup>Our examples in Chapters 7 and 8 do not corroborate this claim.

## 2 Terms and concepts in the light of theoretical approaches

- b. Sa tim čovekom razgovarala je samo Marija.  
 with that man talk.PTCP.SG.F be.3SG only Mary  
 ‘To that man, only Mary spoke.’ (Zec & Inkelas 1990: 374)

Although the first constituent *sa Petrom* branches at the syntactic constituent level, it does not branch at the prosodic constituent level. Therefore, according to Zec & Inkelas (1990: 374) the example in (14a) is ill-formed. In contrast, the first constituent in (14b), *sa tim čovekom*, branches not only at the syntactic constituent level, but also at the prosodic constituent level, and therefore DP of the CL *je* does not result in an ill-formed sentence (Zec & Inkelas 1990: 374).

We would like to point out two facts regarding DP and heavy constituents. First, the initial phrases (or constituents) involved in DP are not necessarily “heavy” in a phonological sense of containing a large number of phonemes: compare our example (15) in which the initial phrase *ovakva vrsta pretrage* ‘this kind of search’ containing 19 phonemes does not host the verbal CL *će*, with the example presented in (16) in which the initial prepositional phrase *do zime* ‘by winter’ containing only six phonemes does not host the reflexive CL *se*.

- (15) [Ovakva vrsta pretrage] [bit] će dostupna za čitav HNK [...].  
 this kind search be.INF FUT.3SG available for entire HNK  
 ‘This kind of search will be available for the entire HNK [...].’  
 [hrWaC v2.2]
- (16) [Do zime] [planira] se završiti asfaltiranje građevine [...].  
 by winter plan.3PRS REFL finish.INF paving building  
 ‘It is planned that paving the building will be finished by winter [...].’  
 [hrWaC v2.2]

Second, the example with DP provided in (16) would not be well-formed if the heavy constituent concept were understood like in Zec & Inkelas (1990: 374f): compare their example in (14a) and our example in (16). Since, as we demonstrated, neither Bošković’s (1995: 264) nor Zec & Inkelas’ (1990: 373ff) measure of heaviness seems to be applicable to the language data which can be observed in corpora, we turn to the measure of heaviness expressed by the number of graphemes, as proposed by Kosek et al. (2018).

The approach taken by Kosek et al. (2018) is in accordance with information found in Stefanowitsch (2020: 90–93, as well as in references therein) on measuring the weight/length of linguistic units (syllables, words or phrases) in corpus studies. Operationalising (word) length for measurement purposes poses difficulties in itself (Stefanowitsch 2020: 90). A number of solutions can be found

in the literature, e.g. number of letters (cf. Wulff 2003), number of phonemes (cf. Sobkowiak 1993) and number of syllables (cf. Sobkowiak 1993, Stefanowitsch 2003).<sup>28</sup>

For the application of the measure proposed by Kosek et al. (2018) see our empirical study based on a corpus of spoken language in Chapter 8.

#### 2.4.3.4 Clitic first

Franks & King (2000: 225–234) discuss another type that departs from strict 2P, in which the requirement that an element precede the CL seems to be violated. They adduce cases where a form which usually lacks stress and attaches to the preceding element shows up in sentence-initial position (CLITIC FIRST, 1P), like in example (17) presented below.

- (17) *Su bíli u célo sèlo.*  
 be.3PL be.PTCP.PL.M in entire village  
 ‘They were in the entire village.’ (Sr; Okuka 2008: 148)

We will deal with cases of 1P in the chapters on variation in the standard languages (Chapter 6), in the dialects (Chapter 7), and in spoken language (Chapter 8).

#### 2.4.3.5 Phrase splitting

BCS differs typologically for example from Modern Czech, as the CL can attach not only to the first phrase but also to the first word of a phrase, allowing for example a noun phrase to be split (as in modified example (18b)):

- (18) a. [Običnim ljudima] *je* dosta rata i žele živjeti u  
 ordinary people be.3SG enough war and want.3PRS live.INF in  
 miru.  
 peace
- b. [Običnim *je* ljudima] *dosta* rata i žele živjeti u  
 ordinary be.3SG people enough war and want.3PRS live.INF in  
 miru.  
 peace
- ‘Ordinary people have had enough of war and want to live in peace.’  
 [hrWaC v2.2]

<sup>28</sup>Applicability in the domain of corpus linguistics limits the options for length operationalisation named here. However, other definitions do exist, such as phonetic length or mean phonetic length (Stefanowitsch 2020: 90f).

## 2 Terms and concepts in the light of theoretical approaches

It is worth pointing out that splitting is independent of 2P or DP. In permuted example (18b) above it is the first phrase which is split, while in example (19) below it is the second phrase:

- (19) [Prema izračunu]<sub>phrase1</sub>, [prosječan je broj sunčanih  
according.to calculations average be.3SG number sunny  
sati]<sub>phrase2</sub> godišnje čak 2500.  
hours yearly even 2500  
'According to calculations, the average number of sunny hours yearly is  
as high as 2500.' [hrWaC v2.2]

Splitting occurs for adverb phrases, adjective phrases, noun phrases and prepositional phrases. We will discuss the possibility of phrase splitting based on the existing research literature in Chapters 6 and 7, and based on our empirical study, in Chapter 8.

### 2.4.4 Clitic climbing

The main focus of our empirical studies on microvariation is on the 2P ordering rule for which the term CLITIC CLIMBING (CC) was established. CC occurs in sentences consisting of a matrix clause and an embedded verbal complement. Descriptively speaking, CC refers to a phenomenon whereby a CL that depends on the embedded complement appears in the matrix clause (see discussion in Chapter 10). Note that throughout this monograph we stick to the established term *climbing* even though we do not necessarily assume any movement operations. In example (20b) the pronominal CL *ga* 'him', which fills an argument position of the infinitival verb *vidjeti* 'to see', is realised in the second position of the matrix clause. In some theories, it is assumed that the CL "climbs" from the verbal complement into the matrix clause.<sup>29</sup> Throughout this book we annotate the relationship between CL and the governing predicate with small subscript numbers: in example (20a) below both infinitive and pronominal CLs are annotated with subscript number 2, which means that the infinitive *vidjeti* generated the pronominal CL *ga*. A matrix predicate is always annotated with subscript number 1 and every further verbal complement (infinitive or *da*<sub>2</sub>-complement), with the next number.<sup>30</sup>

<sup>29</sup>More information on verbal complements in BCS can be found in Section 2.5.3 below.

<sup>30</sup>For more information on *da*-complements and the distinction between *da*<sub>1</sub> and *da*<sub>2</sub> see the next section.

- (20) a. \* Milan mora<sub>1</sub> / želi<sub>1</sub> vidjeti<sub>2</sub> ga<sub>2</sub>.  
 Milan must want.3SG see.INF him.ACC
- b. Milan ga<sub>2</sub> mora<sub>1</sub> / želi<sub>1</sub> vidjeti<sub>2</sub>.  
 Milan him.ACC must want.3SG see.INF  
 ‘Milan must/wants to see him.’ (BCS; Stjepanović 2004: 179f)

An important question is how to detect CC. A clear case of CC is when the CL stands to the left of the matrix predicate (like *ga* before *mora* in example (20b) above). Junghanns (2002: 67), however, warns that if we have the surface word order matrix predicate + CL + infinitive (like in example (21) below) where the CL *ga* ‘him’ occurs directly before the infinitive *upoznati* ‘to get to know’, it cannot be ruled out that the CL is still in the complement.

- (21) Moram<sub>1</sub> ga<sub>2</sub> upoznati<sub>2</sub>.  
 must.1PRS him.ACC get.to.know.INF  
 ‘I have to get to know him.’ [hrWaC v2.2]

CC is a central source of both sociolinguistic and systemic variation in CL usage. Among others, it involves cases where CLs show up in two clusters. This happens e.g. in constructions with stacked infinitives where the CL could climb but for some reason stays in situ, leading to a split of the CLs between two positions, like in example (22) presented below (see Section 2.4.5 below).

- (22) [...] mogao<sub>1</sub> je<sub>1</sub> pokušati<sub>2</sub> spasiti<sub>3</sub> nas<sub>3</sub> od navale  
 can.PTCP.SG.M be.3SG try.INF save.INF us.ACC from attack  
 hohštaplera.  
 conman  
 ‘[...] he could have tried to save us from the attack of the conmen.’  
 [hrWaC v2.2]

Although selected examples of CC have generated controversy in the theoretical literature, hitherto the rules and especially the constraints on CC have not been adequately described. Most tellingly, there is not even an established linguistic term for CC in Serbian or Croatian. Hansen et al. (2013) proposed the ad hoc translation *uspon zanaglasnica* which, however, has not (yet) gained ground in Croatian linguistics. It is no exaggeration to say that the range of microvariation and the possible constraints on CC are a seriously understudied field of BCS syntax. Therefore, we will give a detailed and empirically valid account of this phenomenon in Part III.

### 2.4.5 Diacclisis and pseudodiacclisis

As mentioned above, in certain contexts one CL can be in clausal 2P (or in DP), while an additional clusterising CL is placed to its right (cf. Zimmerling & Kosta 2013: 196). As this phenomenon has not been discussed extensively in the literature on BCS, we use the cover term *DIACLISIS* which we borrowed from Greek linguistics.<sup>31</sup> We use the term for two different types: one for true inner clause diacclisis:

- (23) [...] *po gradovima su<sub>1</sub> predsednici opština se<sub>1</sub> odjednom*  
in cities be.3PL presidents counties REFL suddenly  
*opredeljivali<sub>1</sub> [...].*  
decide

‘In the cities, the county presidents were suddenly deciding [...].’

[Bosnian Interviews, BH]

and one for diacclisis happening in the context of the matrix predicate and its verbal complement(s), as in example (22) above. The latter case is labelled *PSEUDODIACLISIS*. If the difference between the two types is not relevant, we use *diacclisis* as a cover term for the sake of brevity. We discuss this phenomenon in more detail in Chapter 8 and in Chapters 13–15.

## 2.5 Syntactic categories relevant for the description of microvariation

### 2.5.1 Complement-taking predicates

As mentioned above, CC occurs in constructions involving a matrix clause that embeds a second verbal element. As there is no agreement as to the status of the embedded element (clause or non-clause – see discussion in Chapter 10), we would like to avoid the term “clause-embedding predicate” proposed in Stiebels’ (2015) work on control predicates. Instead, we prefer the well-established and more general term *COMPLEMENT TAKING PREDICATE* (CTP) used in the prominent typological work on complementation by Noonan (1985).<sup>32</sup> CTP is a more suitable term than “clause-embedding predicate” because it covers both control and

---

<sup>31</sup>The term is used e.g. by Janse (1998: 270) in work on CLs in Cappadocian Greek. In order to avoid confusion with phrase splitting we do not use the term “splitting” as proposed by Zimmerling & Kosta (2013: 196).

<sup>32</sup>“By complementation we mean the syntactic situation that arises when a notional sentence or predication is an argument of a predicate” (Noonan 1985: 42).



## 2 Terms and concepts in the light of theoretical approaches

Accordingly, in raising constructions (with *seem*) the subject does not receive its semantic role directly from the matrix predicate but from the embedded predicate. In a control construction (with *try*), in contrast, the matrix verb and the embedded verb each assign a subject role (Fried & Östman 2004: 64f). In Principles and Parameters accounts, control constructions are characterised by the presence of two syntactic arguments: a surface subject and a non-overt infinitival subject called big PRO (Wurmbrand 1999: 600). Control always involves a relationship of obligatory (full or partial) co-reference between the non-overt first argument of the complement predicate (CONTROLLEE) and one of the arguments of the matrix predicate (CONTROLLER). In the following example, the first argument of the verb in the complement is interpreted as co-referential with the subject of the matrix clause (marked with  $\chi$ ):

- (25) [...] Dalibor $\chi$  im $\gamma$  je obećao $\chi$  pomoći.  
Dalibor them.DAT be.3SG promise.PTCP.SG.M help.INF  
‘[...] Dalibor promised to help them.’ [hrWaC v2.2]

Davies & Dubinsky (2004: 4–8) list relatively robust, cross-linguistically applicable tests proposed in the literature in order to distinguish raising from control constructions:

1. In raising constructions the subject argument of the matrix predicate has the same semantic role as the subject argument of the embedded predicate. In example (26a), the subject argument *poslodavac* ‘employer’ receives its semantic role of agent from *poništiti* ‘repeal’ (raising) whereas in example (27a) the subject *operater* ‘operator’ receives it from the matrix predicate *pokušao* ‘tried’ (control).

(26) Raising

- a. Poslodavac može poništiti rješenje [...].  
employer can.3PRS repeal.INF settlement  
‘The employer can repeal the settlement [...].’
- b. Rješenje može biti poništeno.  
settlement can.3PRS be.INF repeal.PASS.PTCP  
‘The settlement can be repealed (by the employer).’

[hrWaC v2.2]



## 2.5 Syntactic categories relevant for the description of microvariation

### (27) Control

- a. Operater *je* pokušao ručno obustaviti  
 operator be.3SG try.PTCP.SG.M manually stop.INF  
 reaktor [...].  
 reactor  
 ‘The operator manually tried to stop the reactor [...].’
- b. \* Reaktor *je* pokušao ručno biti obustavljen.  
 reactor be.3SG try.PTCP.SG.M manually be.INF stop.PASS.PTCP  
 Intended: ‘An attempt was made to manually stop the reactor  
 (by the operator).’ [hrWaC v2.2]

2. In subject control constructions the subject argument of the raising predicate does not show any selectional restrictions. For example, the raising verb *trebati* ‘have to’ does not determine the selectional restriction +/– human: compare examples presented in (28) and (29):

- (28) A oni<sub>human+</sub> trebaju platiti za ono što su  
 and they need.3PRS pay.INF for that what be.3PL  
 napravili.  
 do.PTCP.PL.M  
 ‘And they have to pay for what they did.’ [hrWaC v2.2]

- (29) [Idejna rješenja]<sub>human–</sub> trebaju biti poslana u JPEG i PDF  
 idea solutions need.3PRS be.INF sent in JPEG and PDF  
 obliku na sljedeću e-mail adresu [...].  
 format on following e-mail address  
 ‘Ideas for a solution should be sent in JPEG and PDF format to the  
 following e-mail address [...].’ [hrWaC v2.2]

3. In raising constructions passivisation does not change the propositional meaning of the sentences as shown in the permutations of sentences (26b) and (27b).

A distinction is made between subject and object control constructions depending on the argument selected as controller (first or second argument). Whereas predicates that have only one individual argument besides the predicative (verbal) argument are always subject control predicates, polyvalent predicates may show either a subject or an object control reading.<sup>34</sup> According

<sup>34</sup>We do not want to discuss the special cases of partial, split or switch control. For a more detailed account of control see Stiebels (2007, 2015), Landau (2000), Moskovljević (2008), and Słodowicz (2008).

## 2 Terms and concepts in the light of theoretical approaches

to Stiebels (2015: 422), verbs denoting commissive speech acts (e.g. *obećati* ‘promise’) are typical SUBJECT CONTROL predicates, whereas predicates which refer to directive speech acts (e.g. *zamoliti* ‘request’) or which have a causative component belong to the canonical class of OBJECT CONTROL predicates, exemplified here in (30) and (31) by sentences with the so-called *da*-construction:

- (30) On<sub>X</sub> mi<sub>Y</sub> je obećao da će<sub>X</sub> se vratiti u  
 he me.DAT be.3SG promise.PTCP.SG.M that FUT.3SG REFL return.INF in  
 Kragujevac [...].  
 Kragujevac  
 ‘He promised me to come back to Kragujevac, [...]’ [srWaC v1.2]
- (31) Dekan<sub>X</sub> je sve prisutne<sub>Y</sub> zamolio da kažu<sub>Y</sub> svoje  
 dean be.3SG all present ask.PTCP.SG.M that say.3PRS own  
 utiske [...].  
 impressions  
 ‘The Dean kindly asked the attending members to share their  
 impressions [...]’ [srWaC v1.2]

The raising–control distinction as outlined above is orthogonal to the distinction of matrix verbs proposed by Progovac (1993b) and applied by Todorović (2015). In the following we explain why we do not use this classification, although it has been developed and used by scholars dealing with BCS CLs.

Progovac (1993b: 116) distinguishes two basic groups of verbs: those which select opaque complements (I-verbs, or indicative-selecting verbs) and those which select transparent complements, allowing for domain extension (S-verbs, which select subjunctive-like complements).<sup>35</sup> I-verbs are mostly verbs of saying, believing and ordering, such as *kazati* (‘tell’), *v(j)erovati* (‘believe’) or *narediti* (‘order’). S-verbs are mainly verbs of wishing and requesting, such as *žel(j)eti* (‘want/wish’), *ht(j)eti* (‘want/will’), *moći* (‘be able to’) and *tražiti* (‘ask for’).

According to Progovac (1993b: 116), “the following local dependencies in Serbo-Croatian are clause bound with I-verbs, but can cross clause boundaries with S-verbs: licensing of negative polarity items (NPIs), clitic climbing, and topic preposing”.

We would like to point out that the distinction between S- and I-verbs might not be as clear as it appears at first sight, i.e. as presented by Progovac (1993b). First, these semantic verb classes are quite heterogeneous: verbs of ordering, like the mentioned *narediti* ‘order’, in fact select subjunctive-like complements which do not allow past or future tense.

<sup>35</sup>A similar distinction is applied by Landau (2004) to Balkan languages and Hebrew.

## 2.5 Syntactic categories relevant for the description of microvariation

Second, there are cases where the dependency relation seems to go in the opposite direction. That is, it seems that a complement can change the class of a verb. For instance, if verbs of saying co-occur with subjunctive-like complements, semantic coercion occurs. A verb of saying is interpreted as a verb of ordering, as in the following example:

- (32) Rekao *sam im* da budu oprezni, objasnio  
 say.PTCP.SG.M be.1SG them.DAT that be.3PL careful explain.PTCP.SG.M  
 tko *sam*, što *sam*.  
 who be.1SG what be.1SG  
 ‘I told them they should be careful and explained who I am, what I am.’  
 [hrWaC v2.2]

Even though the I- and S-classification of verbs seems too simplistic, a claim in Progovac (1993b: 119) and Progovac (2005: 146) has particular relevance to our study: that S-verbs allow CC, whereas I-verbs do not. Progovac discusses the following minimal pairs of sentences:

- (33) a. Milan kaže<sub>1</sub> da ga<sub>2</sub> vidi<sub>2</sub>. I-Verb  
 Milan say.3PRS that him.ACC see.3PRS  
 ‘Milan says that he sees him.’  
 b. \*Milan ga<sub>2</sub> kaže<sub>1</sub> da vidi<sub>2</sub>.  
 Milan him.ACC say.3PRS that see.3PRS  
 Intended: ‘Milan says that he sees him.’ (BCS; Progovac 1993b: 119)
- (34) a. Milan želi<sub>1</sub> da ga<sub>2</sub> vidi<sub>2</sub>. S-Verb  
 Milan want.3PRS that him.ACC see.3PRS  
 b. ?Milan ga<sub>2</sub> želi<sub>1</sub> da vidi<sub>2</sub>.  
 Milan him.ACC want.3PRS that see.3PRS  
 ‘Milan wants to see him.’ (BCS; Progovac 1993b: 119)

However, as we show in Chapter 13, the claim that “with S-verbs clitics originating in the embedded clause can optionally climb to the second position of the matrix clause” (Progovac 1993b: 119), which concerns structures such as those in (34b), is somewhat problematic. Progovac (1993b: 119) herself is unsure whether the sentence is grammatically correct, since she uses the question mark. Moreover, in a footnote in her later publication (Progovac 2005: 146), she admits that some speakers of Serbian, including the linguist Vesna Radanović-Kocić, do not accept CC. In Chapter 13 we give an empirical answer to the question of the extent to which the structure presented in example (34b) is possible, i.e. used by

native speakers of Serbian. Finally, we would like to emphasise that a more fine-grained subclassification of S-verbs, as offered by the raising–control distinction, is called for.

### 2.5.3 Types of complements

As mentioned above, CC involves structures with a matrix and an embedded complement. In BCS, the latter can be encoded either by a phrase with an infinitive (as in (35a)) or by a phrase introduced by the element *da* (as in (35b)) sometimes treated as a complementiser; see the examples from Stjepanović (2004: 174).<sup>36</sup>

- (35) a. Marija *ga*<sub>2</sub> mora<sub>1</sub> / želi<sub>1</sub> posjetiti<sub>2</sub>.  
 Marija him.ACC must want.3PRS visit.INF
- b. Marija *ga*<sub>2</sub> mora<sub>1</sub> / želi<sub>1</sub> da posjeti<sub>2</sub>.  
 Marija him.ACC must want.3PRS that visit.3PRS  
 ‘Marija must/wants to visit him.’ (BCS; Stjepanović 2004: 174)

It has long been known that *da*-complements do not behave in a uniform way. Ivić (1970) proposes to distinguish two complement types headed by *da* depending on tense marking: complements with “mobile present tense” and complements with “immobile present tense”, the former being regularly marked for tense and the latter not. This distinction goes back to Gołąb (1964) and was further elaborated on by Browne (2003: and earlier) who uses the labels *DA*<sub>1</sub>- and *DA*<sub>2</sub>-COMPLEMENT. Here is an example from Browne (2003: 39) with the CTP *saznati* ‘find out’, which allows present (36a), past (36b) or future tense marking (36c).

- (36) a. Saznao sam da crtaš zmiju.  
 find.out.PTCP.SG.M be.1SG that draw.2PRS snake  
 ‘I found out that you were drawing a snake.’
- b. Saznao sam da ste crtali zmiju.  
 find.out.PTCP.SG.M be.1SG that be.2PL draw.PTCP.PL.M snake  
 ‘I found out that you had been drawing a snake.’
- c. Saznao sam da ćeš crtati zmiju.  
 find.out.PTCP.SG.M be.1SG that fut.2SG draw.INF snake  
 ‘I found out that you would draw a snake.’ (BCS; Browne 2003: 39)

<sup>36</sup>Note that in the glossing of our examples we do not account for the polyfunctionality of the glossed morpheme *da* and that we simply gloss it lexically as ‘that’.

## 2.5 Syntactic categories relevant for the description of microvariation

In contrast, *da*<sub>2</sub>-complements only allow the verbal form coinciding with the present tense; other tenses are impossible. Ivić (1972) speaks about the “immobility” of the present tense (*nemobilnost prezenta*) whereas Đukanović (1994: 119) assumes that tense marking is blocked (*vremenska umrtvljenost, neovremenjenost*). It is claimed that *da*<sub>2</sub>-complements occur with CTPs with volitional meaning, e.g. with the verb *žel(j)eti* ‘wish/want’ in the following example (37a) and in its two permutations.

- (37) a. *Želim da<sub>2</sub> crtaš zmiju.*  
want.1PRS that draw.2PRS snake  
‘I want you to draw a snake.’
- b. \**Želim da<sub>2</sub> si crtala zmiju.*  
want.1PRS that be.2SG draw.PTCP.SG.F snake  
Intended: ‘I want you to have drawn a snake.’
- c. \**Želim da<sub>2</sub> ćeš crtati zmiju.*  
want.1PRS that FUT.2SG draw.INF snake  
Intended: ‘I want you to draw a snake in the future.’
- (BCS; Browne 2003: 39)

Todorović (2015) proposes to make a distinction between indicative and subjunctive complements. As we do not want to discuss the link between the two complement types and any semantic (i.e. modal) features, we stick to the terms *da*<sub>1</sub>- vs *da*<sub>2</sub>-complement. For the relationship between the semantics of the CTP and the selection of the *da*-complement we refer to the in-depth empirical study by Hansen, Wald & Kolaković (2018), who show that the semantics of the CTP does not directly determine the selection of the complement type (contra Todorović 2015).

### 2.5.4 Different types of reflexives

Veering away from problems directly relating to the inventory of CLs, we would like to discuss a wider issue concerning the REFLEXIVE CL *se*. Like other Slavic languages, BCS displays a wide range of usages of the reflexive marker and we assume that these may differ in their syntactic behaviour, e.g. with respect to CC. In example (38) *se* has a different status than in example (39) because in the first case its function is to hide the first argument (impersonal use), while in the second it is used to indicate reciprocity.

## 2 Terms and concepts in the light of theoretical approaches

- (38) Reklo *bi se da je* to srebrn upaljač.  
say.PTCP.SG.N COND.3SG REFL that be.3SG that silver lighter  
'It seemed to be a silver cigarette lighter.' (=One would say [...])  
(Cr; Moulton 2015: 106)
- (39) Hrvati *i mi ne moramo da se volimo*, ali moramo da se  
Croats and we NEG must.1PRS that REFL love.1PRS but must.1PRS that REFL  
poštujemo.  
respect.1PRS  
'Croats and we do not need to love each other, but to respect each other.'  
(Sr; Moulton 2015: 103)

As it is not our aim to either give an exhaustive overview of the existing research literature or to develop our own theoretical account of the different types of constructions, we restrict ourselves to the identification of some syntactic types of constructions with the element *se* which may differ as to CC. As a matter of fact, there is a considerable body of research dealing with reflexives in the Slavic languages in general and in BCS in particular. The topic has attracted the attention of different scholars working in both formal and cognitive-functional frameworks. For our typology we draw on the recent study of reflexives from a broader Slavistic perspective, Fehrmann et al. (2010). Among the studies specifically dealing with reflexives in BCS, we turn to the cognitivist work by Moulton (2015).

### 2.5.4.1 The approach of Fehrmann, Junghanns, and Lenertová (2010)

In this section, we will discuss the reflexive markers based on the first two steps of our triangulation of empirical methods outlined in Chapter 3 (intuition/theory – observation – experiment). We start with data from the literature – in this case data from Fehrmann et al. (2010) – and verify them by searching for qualitative empirical data in corpora in the sense of a corpus-illustrated approach. Keeping our research question in view, we will focus on evidence for microvariation in the use of reflexive markers.

As our study deals with variation in CL positioning and not with different semantic or structural types of the reflexive marker, we will restrict ourselves to testing a small number of types for differences especially in relation to CC. For the purposes of our study, it suffices to identify several types of reflexive constructions. Moulton (2015) distinguishes six semantic types which partially overlap with the list of surface configurations of reflexives in ten Slavic languages discussed by Fehrmann et al. (2010).<sup>37</sup> They present a unified account of two

<sup>37</sup>Moulton (2015) distinguishes reflexive verbs, possessive reflexive verbs, reciprocal verbs, passive constructions, impersonal constructions, and middle verbs.

## 2.5 Syntactic categories relevant for the description of microvariation

different lexical reflexive markers REFL1/REFL2 based on the framework of a two-level semantics.<sup>38</sup> The authors exclude “the relatively small group of reflexive verbs that synchronically have no non-reflexive counterparts”. These are stored in the lexicon as a unit (verbs like *smijati se* ‘laugh’).

Fehrmann et al. (2010) analyse seven descriptive surface types which differ among others as to the argument affected (first vs second), the argument blocking vs argument binding distinction and the presence of additional semantic features.<sup>39</sup> Based on the possibility or exclusion of so-called by-phrases, they argue that two reflexives “are necessary, but also sufficient, for the analysis of all [...] uses, regardless of whether an external or an internal argument is affected” (Fehrmann et al. 2010: 206). The term by-phrase is used as a label covering very different surface manifestations including prepositional phrases, dative phrases and others. The main idea is that a REFL affects one of the arguments of the verbal predicate preventing the canonical realisation of this argument as subject or object (Fehrmann et al. 2010: 208). Put simply, if a semantic specification of the affected argument is possible (via a so-called by-phrase as mentioned above), the authors propose REFL1 as an argument-blocking device. In contrast, if a semantic specification of the affected argument is not possible they refer to REFL2 as an argument-binding device. In the latter case the affected argument receives an arbitrary human interpretation. This distinction is claimed to be of a categorical nature which does not seem to allow for microvariation. In the following, we will show that this claim does not withstand closer scrutiny.

Leaving aside the formal machinery used in the two-level semantics approach, we condense the main ideas and apply the stipulated distinctions to the diverse usages of *se* in BCS. We critically discuss the question whether the distinction between REFL1 and REFL2 is sufficient for a typology of *se* usages in BCS. The point of departure is the following list of surface configurations with *se* proposed by Fehrmann et al. (2010):

1. “Reflexive passive” where the second argument is realised in the nominative. The passive is restricted to transitive verbs and, as the authors claim, does not allow the so-called by-phrase in the form *od strane* for the expression of the first argument. In the following example (40a) the agent (i.e. the builder) allegedly remains unspecified.

---

<sup>38</sup>This framework distinguishes Semantic Form and Conceptual Structure, where the former mediates between the latter and the syntax (originally going back to Bierwisch 1986).

<sup>39</sup>Fehrmann et al. (2010) do not use the terms “first” and “second argument”, they use the terms “external” and “internal argument” instead. The distinction first vs second argument is found among others in Role and Reference Grammar.

## 2 Terms and concepts in the light of theoretical approaches

- (40) a. Kuća se gradi.  
 house REFL build.3PRS  
 ‘The house is being built.’
- b. \* Kuća se gradi od strane radnika.  
 house REFL build.3PRS from side builders  
 Intended: ‘The house is being built by builders.’  
 (BCS; Fehrmann et al. 2010: 205)

However, if we verify this claim with selected data from the web corpora, we do find instances of the by-phrase *od strane* specifying the reference of the first argument – see our example presented in (41).<sup>40</sup>

- (41) Koristile su se razne metode od strane vladajuće  
 use.PTCP.PL.F be.3PL REFL various methods from side ruling  
 stranke i njihovih poslušnika [...].  
 party and their minions  
 ‘Various methods were used by the ruling party and their minions  
 [...].’ [hrWaC v2.2]

2. “Reflexive impersonal” where the first argument is affected; in standard Croatian and Serbian the second argument has to be realised in the nominative as in (42a), whereas – as Fehrmann et al. (2010) claim – in colloquial Croatian it can also appear in the accusative as in example (42b) below (cf. Katičić 1986: 146).<sup>41</sup> In contrast to the passive, the impersonal is not restricted to transitive verbs. Based on example (43a), Fehrmann et al. (2010: 214) assume that in BCS no by-phrase is possible: see their permutation of the example in (43b). This is in line with the claims of Silić & Pranjković (2007: 318).<sup>42</sup>

<sup>40</sup>This possibility seems to have been noted in Croatian grammaticography. While Barić et al. (1999: 257) allow for the insertion of the subject only in the case of participle passive with an animate subject (which they call *agens u užem smislu*), Silić & Pranjković (2007: 318) claim that it is generally possible to insert subjects in passive sentences.

<sup>41</sup>Both Katičić (1986: 146) and Barić et al. (1999: 260) agree that such constructions with an object in the accusative belong to Croatian substandard; see „Preoblika obezličanja ne primjenjuje se na prelazne glagole s izrečenim objektom u pomnije dotjeranom hrvatskom književnom jeziku i zato je to oznaka nešto manje brižna izražavanja” (Katičić 1986: 146).

<sup>42</sup>It is possible to find similar examples in hrWaC v2.2, for instance.

- (i) [...] čuje se vodu kako lupa o zidove suđerice [...].  
 hear.3PRS REFL water how hit.3PRS about walls dishwasher  
 ‘[...] one hears the water splashing against the dishwasher walls [...].’ [hrWaC v2.2]



## 2.5 Syntactic categories relevant for the description of microvariation

- (42) a. Čuje se kiša.  
hear.3PRS REFL rain  
b. Čuje se kišu.  
hear.3PRS REFL rain  
'One hears the rain.' (Cr; Fehrmann et al. 2010: 214)
- (43) a. Plesalo se sve do zore.  
dance.PTCP.SG.N REFL all to dawn  
'One danced until dawn.'  
b. \*Plesalo se sve do zore od strane žena.  
dance.PTCP.SG.N REFL all to dawn from side women  
Intended: 'One danced until dawn.'  
(BCS; Fehrmann et al. 2010: 223,  
adapted from Progovac 2005: 72)

However, as in the case of the passive, in our corpora we were able to find examples containing the by-phrase, like the one presented in (44). This would speak in favour of an interpretation as REFL1 and not REFL2 in the terms of Fehrmann et al. (2010).

- (44) Gdje ste Vi bili tada, kada se prozivalo  
where be.3PL you be.PTCP.PL.M then when REFL call.out.PTCP.SG.N  
narod od strane bivše premijerke Jadranke Kosor [...]?  
nation from side former prime.minister Jadranka Kosor  
'Where were you then, when the nation was called out by the  
former prime minister, Jadranka Kosor [...]?' [hrWaC v2.2]

3. Genuine reflexive and reciprocal where the second argument is coreferential with the first; the semantic specification of the affected second argument is obtained via identification or coindexation with the referent of the subject. In example (45) the person referred to by the subject washes him or herself; example (39) from Moulton (2015: 103) provided above at the beginning of the section is a case of a reciprocal interpretation. In these cases a by-phrase with *od strane* is ruled out.

- (45) Pažljivo se umivam i nakon toga nanosim  
carefully REFL wash.face.1PRS and after that apply.1PRS  
hidratantnu kremu.  
hydrating cream  
'I carefully wash my face and apply a hydrating cream afterwards.'  
[hrWaC v2.2]

## 2 Terms and concepts in the light of theoretical approaches

4. Antipassive where the blocked second argument is usually interpreted as arbitrary; this construction allows for the overt expression of the affected argument in the form of a prepositional phrase which leads to a non-arbitrary interpretation, see the example presented in (46). Antipassives are restricted to a small group of verbs. Moulton (2015: 110), who uses the term agent attributive, shows that *on se tuče* ‘he gets into fights’ allows a by-phrase: *on se tuče s drugima*, ‘he beats others up.’

(46) *Kada se dete mnogo tuče, nadgledajte ga češće.*  
when REFL child much hit.3PRS oversee.IMP.2PL him more often  
‘When a child hits a lot (= is a frequent hitter), watch him closely.’  
(Sr; Moulton 2015: 111)

5. Middles look like passives but involve a non-episodic reading. “The subject is interpreted as having properties that do or do not allow for the action expressed by the predicate to be potentially performed on the subject by an implicit, generic agent in a specific way expressed by adverbial means.” (Fehrmann et al. 2010: 227). The first argument – i.e. the washer in example (47) below – is bound. No by-phrase seems possible.

(47) *Ova se haljina lako pere.*  
this REFL dress easily wash.3PRS  
‘This dress is easily washed.’  
(Cr; Kučanda 1998: 214)

6. Decausatives affect the second argument and denote situations with an unagentive interpretation which is derived from the identification of the second and the first argument – see example (48) provided below. In some cases a by-phrase with *od* and the genitive is possible – see the example presented in (49).

(48) *Potopio se brod, poginulo 36 ljudi, među njima*  
sink.PTCP.SG.M REFL ship die.PTCP.SG.N 36 people among them  
*i trudnica.*  
and pregnant.woman  
‘A ship sank, 36 people, including a pregnant woman, perished.’  
(Cr; Moulton 2015: 109)

(49) *Čarobničine tamnozeleno oči zamute se [od suza].*  
sorceress dark.green eyes blur.3PRS REFL from tears  
‘The sorceress’ dark green eyes blurred from tears.’  
(Cr; Katičić 1986: 145)

## 2.5 Syntactic categories relevant for the description of microvariation

In addition to these six types Fehrmann et al. (2010) distinguish a further type which they call the “involuntary state construction”, where the first argument is encoded in the dative and the predicate receives a stative reading. This type is attested, for example, in Polish, but not in BCS where there is a related but distinct construction for which we propose the term *FEEL-LIKE CONSTRUCTION*. Among others, it involves “a dispositional interpretation (‘x feels like V-ing’) but no overt dispositional element”. Marušič & Žaucer (2014) suggest (for Slovene) that a null-psych-verb is present. This construction contains a dative phrase expressing the first argument interpreted as an experiencer. We consider structures with a nominative subject to be feel-like constructions as well – see example (50) provided below.

- (50) Marku     *se*    igrala         košarka.  
Marko.DAT REFL play.PTCP.SG.F basketball  
‘Marko felt like playing basketball.’ (Sr; Stanojčić & Popović 2002: 249)

This discussion can be summarised as follows: the (slightly revised) list of surface configurations is indeed useful for capturing the range of usages of *se* in BCS. The REFL1/REFL2 dichotomy, however, turns out to be built on shaky empirical ground.<sup>43</sup>

### 2.5.4.2 Conclusion: how many types of *se* do we need to distinguish?

To conclude, the preceding discussion of the approach to a typology of reflexive markers proposed by Fehrmann et al. (2010: 228) provides a mixed picture. On the one hand, the authors propose a list of surface configurations which seems to be applicable to BCS and claim that in Slavic a crucial role should be assigned to the

---

<sup>43</sup>In a brief digression, we would like to comment on middles and the feel-like construction. More in-depth analyses explaining the additional semantic elements mentioned above are certainly needed, but we do not agree with Fehrmann et al. (2010: 228) who assume the presence of a modal operator of possibility in the structure of middles. Generally, we would argue against a broad understanding of the term modality. Real modal constructions differ both in form and function. Neither the “feel-like” nor the “involuntary action” semantic component (e.g. for Polish) belongs to the semantic domain of possibility as a subdomain of modality *sensu stricto* (cf. van der Auwera & Plungian 1998). We think we are dealing with a meaning usually associated with so-called psych-verbs. As to the referential status of argument expression in the case of middles, the sentence interpretation is always generic, while in feel-like constructions it can be either specific or generic. Furthermore, we would like to point out that the status of the dative phrase (whether it is an external argument or not) needs a more elaborate discussion. Due to lack of space, we will refrain from deeper analysis and refer the reader to works on non-canonical subjects in Croatian by Kučanda (1998) and Hansen, Wald & Kolaković (2018).

## 2 Terms and concepts in the light of theoretical approaches

availability of a by-phrase. They convincingly argue for the distinction between argument blocking and argument binding. However, our first tentative empirical test of their claims reveals that this distinction becomes blurred because in natural language use we found evidence that many more reflexive constructions allow the use of the by-phrase than the authors who refer to the prescriptive norms of standard Croatian and Serbian claim. Our data indicate that in BCS only middles are a clear case for REFL2. All the remaining usages of *se* are either unclear or evidently belong to the REFL1 argument-blocking type. These usages, however, vary considerably. This leaves us with the sobering conclusion that for an empirically validated, full typology of syntactic types of reflexive markers much more work has to be done.

Therefore, we draw on the crucial observations by Fehrmann et al. (2010), but use an additional feature in our typology. The main idea is that a REFL affects one of the arguments of the verbal predicate, preventing the canonical realisation of this argument as subject or object (Fehrmann et al. 2010: 218). With regard to our empirical data, we, however, do not base our typology on the availability of the by-phrase. Instead, we propose a much simpler, robust typology referring to which argument (first vs second) is affected. Based on the discussion of the seven surface types above, we thus reach a threefold distinction:

1. FIRST ARGUMENT AFFECTED,
2. SECOND ARGUMENT AFFECTED,
3. FIRST AND SECOND ARGUMENT AFFECTED.

Additionally, we fully acknowledge the status of lexically determined usages of the reflexive marker. These reflexive verbs are of major interest to our study on CC as they appear both in CTPs (matrix verbs) and complements (both finite and semifinite). As to the grammatically determined usages, we more or less accept the list of surface types proposed by Fehrmann et al. (2010).

1. REFL<sub>LEX</sub> – *se* stored in the lexicon; in contrast to Fehrmann et al. (2010) we assume that this group comprises not only reflexiva tantum (reflexive verb as lemma, e.g. *smijati se* ‘laugh’), but also verbs which have a slightly different meaning than their non-reflexive counterpart (reflexive verb as lexeme, e.g. *šetati* ‘stroll/walk’ vs *šetati se* ‘take a walk’);
2. REFL<sub>1ST</sub> – constructions where only the first argument is affected: non-standard impersonal;

## 2.5 Syntactic categories relevant for the description of microvariation

3. REFL<sub>2ND</sub> – constructions where only the second argument is affected: reflexive/reciprocal, antipassive;
4. REFL<sub>1ST + 2ND</sub> – constructions where both the first and second argument are affected: passive, standard impersonal, middles, feel-like and decausative.

For the purposes of the present study, it is sufficient to distinguish these four types of usages of the reflexive marker *se*. In Parts II and III we comment on the reflexive CL *si* which can occur in the same contexts as the REFL<sub>2ND</sub> CL *se*. We use this tentative typology, including the corresponding abbreviations REFL<sub>LEX</sub>, REFL<sub>1ST</sub>, REFL<sub>2ND</sub> and REFL<sub>1ST + 2ND</sub>, throughout the book. The main focus of the psycholinguistic test is on lexical reflexives (REFL<sub>LEX</sub>) and genuine reflexives (sub-type of REFL<sub>2ND</sub>). We hypothesise that the lexical vs grammatical usage of *se* plays a role in CC.



## 3 Empirical approach to clitics in BCS

### 3.1 Introduction

The goal of this chapter is to present the ongoing discussion on data gathering practices in syntactic research on the one hand and to justify our choice of strategy on the other.

The methodological literature points out that the most widely used data source in syntactic research is speakers' intuitions (Schütze & Sprouse 2013: 27). In order to collect evidence which will enable them to describe syntactic structures, "syntacticians often rely on their own judgments, or those of a small number of their colleagues", about the acceptability of a structure/sentence in question (Dąbrowska 2010: 1). Some linguists (e.g. Newmeyer 1983: 48ff, Fanselow 2007: 353, Grewendorf 2007: 370, Phillips 2010) have argued that this kind of data is the most reliable and that it allowed the rapid development of linguistics. Others (e.g. Schütze 2016, Cowart 1997, Keller 2000, Featherston 2007) have replied that this can be problematic and that such an informal approach to collecting data leaves linguistics on shaky empirical ground.<sup>1</sup> As Clark & Bangerter (2004: 25) put it, using introspective methods "you imagine a wide range of utterances and situations and draw your conclusions. You are limited only by what you can imagine, but that turns out to be quite a limitation."

There are several crucial points in which typical informal linguistic judgments differ from the methodologically standardised practise of data gathering:

---

<sup>1</sup>Phillips & Lasnik (2003) try to defend generative grammar and to show that it is not built upon empirically weak foundations by presenting different kinds of experiments which have been used in the generative framework.

### 3 Empirical approach to clitics in BCS

1. judgments collected from very few speakers,<sup>2,3</sup>
2. linguists as participants,
3. judgments collected only for few tokens of a structure of interest,
4. relatively unsystematic data analysis (cf. Schütze & Sprouse 2013: 30).

In the following, we discuss some of these points.

We agree that stable measures of acceptability (or grammaticality, see below) can be obtained only if we average responses which were provided by a number of informants (cf. Dąbrowska 2010: 1f). As Dąbrowska (2010: 2) points out “[a]nother problem with linguists’ reliance on their own intuitions is observer bias: the possibility that judgments can be influenced by the observer’s beliefs and expectations”. A further reason why linguists and non-linguists tend to evaluate the same sentences differently can be disparities in experience. Hiramatsu (1999) and Snyder (2000) experimentally demonstrated the existence of SYNTACTIC SATIATION – a phenomenon where participants (linguists or students of linguistics) are prone to accept some types of ungrammatical or borderline structures due to repetitive exposition to such structures. Furthermore, some scholars (e.g. Schütze 2016: 47, Cowart 1997: 60, Snyder 2000: 575) warn about the additional danger of judgments made by linguists: it is possible that “correct answers” could have been learned from the linguistic literature or in the course of education.

Surely, if the intuitions of every native speaker are based on the same hard-wired language faculty which according to early generativist assumptions fully represents his or her linguistic competence, consulting a vast number of speakers about the same language phenomenon would only result in replication of one

---

<sup>2</sup>As Schütze & Sprouse (2013: 39) point out this can be by necessity. “In the case of languages spoken in remote locations and languages with few remaining speakers, collecting data from just one or two speakers may be all that a linguist can practically do [...]”. However, as we argue later, there is no reason to treat Bosnian, Croatian, and Serbian as languages for which more data could not or should not be collected.

<sup>3</sup>Schütze & Sprouse (2013) do not define how many exactly makes a very few speakers. For us “very few” means not enough to perform significance tests which allow estimating the probability that the result is replicable in another sample. However, we do not think that everybody in the world needs to work using statistical methods. Therefore, we could retreat from the expression “very few” in favour of “unspecific number of speakers” or “judgments obtained in an unsystematic” or “not documented way”. Of course, the judgments of two speakers might be perfectly in line with those of 102 speakers, but the question remains: how can we obtain confidence about that.



### 3.2 *From researchers' intuition to triangulation of methods*

and the same answer, which, of course, would be a waste of time and human resources (Buchstaller & Khattab 2013: 85). Already in the 1970s it was recognised that such an approach can be problematic: Sampson (1975: 74) stated that “[o]ne of the unfortunate consequences of Chomsky’s mentalist view of linguistics is that in recent years a number of younger linguists have indulged very heavily in arguments based on their intuitions about quirks of their personal idiolects”. Similarly later he claimed that “[w]e do not need to use intuition in justifying our grammars, and as scientists, we must not use intuition in this way” (Sampson 2001: 135).

The methodological literature mentions several problems with introspective data collection. One serious problem was identified by Labov (1972: 199) and more recently repeated by Schütze (2016). They point out that it might be dangerous to produce theory and data at the same time. Schütze (2016: 5) warns that if linguists continue to produce theory and data at the same time, what is to stop them from purposely or accidentally manipulating the introspection process in order to substantiate their own theories?

To sum up: intuition-based judgments can suffer from bias, unreliability, and narrowness (Schütze 2016). These problems are described in quite some detail in relation to data on CLs in BCS in our book. Specifically, data based on linguists’ informal judgments very often turned out to be contradictory and flawed.<sup>4</sup> In Section 3.2 we present how these problems can be overcome through triangulation of methods. A detailed discussion of the empirical approach chosen for this monograph together with a concise overview of corpus linguistic and psycholinguistic methods can be found in Section 3.3. The chapter ends with a presentation of the experiment chosen for our study.

## 3.2 From researchers’ intuition to triangulation of methods

### 3.2.1 Triangulation of methods

The pitfalls of studies based exclusively on intuition mentioned above can be overcome through triangulation of methods, an approach well-established in the social sciences. Following the definition of Yeasmin & Rahman (2012: 154), triangulation “is a process of verification that increases validity by incorporating

---

<sup>4</sup>To understand this problem fully, compare the data from the literature in Chapter 6 with the empirical data from spoken varieties in Chapters 7 and 8. The gap between data based on informal judgments and empirical data is even better illustrated in our Chapters 13–15.

several viewpoints and methods”. The importance of triangulation for linguistic studies has only recently been acknowledged. Hoffmann (2013: 100) and Ford & Bresnan (2013: 311) recommend triangulation of methods to provide corroborating evidence and to capture language usage most accurately. According to Angouri (2010: 33), mixed methods designs (i.e. combining or integrating quantitative and qualitative elements) arguably contribute to a better understanding of the various phenomena under investigation, since quantitative research is useful for generalising research findings, while qualitative approaches are particularly valuable in providing rich in-depth data.

Rosenbach (2013: 293) argues that the combination of different methods eliminates the restrictions which emanate from the limitations of a given single method. For instance in corpus-based data it is hard to control influencing factors, a problem which we can overcome when we use experimental elicitation of data. The problem of limited context and the lack of naturalness, i.e. ecological validity (see section below), which accompanies experimental data, can be avoided if we supplement it with corpus data.

However, combining different methods has disadvantages as well, since it generates higher overall costs than applying a single approach: it is more time consuming and requires expertise in both methods (Rosenbach 2013: 293, Ford & Bresnan 2013: 311).<sup>5</sup> Hence, the last disadvantage often means involving more researchers, which, can in itself be counted as an advantage, since independence of scholars improves research objectivity. Moreover, there is no standard methodology in the field of triangulation and incorrectly combined methods do not fulfil their assumed function, which can pose some additional problems. Finally, one should also be aware that repeating a study involving several methods is less likely to happen than if a single method had been applied.

#### 3.2.2 Research validity

As mentioned above, the main strength of triangulation of methods lies in providing robust evidence of real language use, and it is a reliable method for verifying results.<sup>6</sup>

In our study, ecological validity is supported in three ways. First, we retrieve fully uncontrolled material from web corpora, which guarantees observations

---

<sup>5</sup>Sometimes, as we argue in Section 14.4, the cost of combining different methods may be lower. This happens, for example, when retrieving rare structures from corpora is far more complicated than elicitation and may lead to the problem of negative evidence.

<sup>6</sup>As already explained in Chapter 1 under the expression “real language use” we understand observable language data as opposed to language data obtained by introspection.

from a fully natural environment, without any influence on language users from investigators. Secondly, the examples obtained from corpora are used as model sentences for acceptability experiment. Finally, we conducted a pilot study where we asked native speakers to evaluate our target sentences. The results of their feedback were used to improve stimuli to sound as natural as possible. This solution should ensure that constructed examples are not entirely artificial, and hence are likely to appear in real-life situations.<sup>7</sup>

When talking about research validity, Brewer (2000) also distinguishes internal validity, sometimes called construct validity – “the degree to which a study allows unambiguous causal inferences”, and external validity – “the degree to which a study ensures that potential findings apply to settings and samples other than the ones being studied” (Brewer 2000). These two types of validity rarely apply to a single study. This is because while for example words and sentences in corpora are not without their broader contexts, words and sentences in acceptability judgments are usually elicited in isolation (Myers 2017: 3). In our case, the findings obtained from laboratory experiment guarantee high internal validity, while structures retrieved from corpora support ecological and external validity.

## 3.3 Empirical approach in the current study

### 3.3.1 Chosen strategy

The current work is language-use oriented and we follow the scheme intuition/theory – observation – experiment. Many theoretical claims concerning CLs in BCS are contradictory (see: Chapters 2, 6, 10, and 11). Therefore, in these chapters we first verify them against empirical data collected from corpora – our first source of observation. Since corpus data can be analysed quantitatively, some hypotheses can be also verified at this stage.<sup>8</sup> This procedure is applied mainly in Part III of the book, which focuses on the understudied phenomenon of clitic climbing, but also in Part II, where we analyse the behaviour of CLs in spoken Bosnian.

Nevertheless, the high level of ecological validity typical of corpora is also their drawback, as internal validity in large collections of spontaneously produced texts is quite low. Very often the influence of extralinguistic factors cannot be ruled out, e.g. due to the lack of information about the social background of the authors. Nevertheless, hypotheses formulated on the basis of corpus material

---

<sup>7</sup>For more information see Section 15.3.3.2.

<sup>8</sup>The quantitative methods we use are discussed in Chapter 12.

### 3 Empirical approach to clitics in BCS

can be further tested in acceptability judgment experiments where the level of control on particular factors can be adjusted. Additionally, corpora as recordings of natural language production can be nicely supplemented with experimental data such as acceptability judgment data because while they both provide evidence about syntax, the kind of evidence differs. While corpora reflect language production, acceptability data primarily reflect language comprehension (Myers 2017: 3).

#### 3.3.2 Corpus studies

##### 3.3.2.1 Corpus linguistics

Generally speaking, in the current work we understand corpus linguistics as a language-use oriented research approach which utilises collections of texts produced in a natural communicative situation, called corpora, and applies quantitative and qualitative analytical tools and techniques to them. “Over the last few decades, corpus-linguistics methods have established themselves as among the most powerful and versatile tools to study language acquisition, processing, variation and change” (Gries & Newman 2013: 257). We decided to use corpus linguistics methods, as an alternative to intuitive acceptability judgments made by one person, only since they offer (more) objective, quantifiable, and replicable findings (cf. Gries & Newman 2013: 257). Contrary to what most works (Tummers et al. 2005) present, corpus linguistics has more to offer than a simple opportunity to extract authentic examples for the purpose of introspective research. In the following section we present the main approaches to corpus research and describe our own.

##### 3.3.2.2 Hybrid approach to corpus linguistics

Investigations incorporating corpus linguistic methods are traditionally divided into corpus-driven and corpus-based. Gries (2010: 328) provides three typical features of the CORPUS-DRIVEN approach:

1. building theory from scratch as an aim, without any theoretical assumptions,
2. basing theories exclusively on corpus data,
3. (often) rejecting corpus annotation.

### 3.3 Empirical approach in the current study

Scholars treat these three elements differently, so corpus-driven is still a rather fuzzy term. In its most extreme form the corpus-driven approach allows only the assumption of word forms and requires a purely distributional analysis of the corpus in order to identify any linguistic units (cf. Gries 2010: 329, Biber 2015: 201). Thus, we agree that “*truly* corpus-driven work seems a myth at best” (Gries 2010: 330). In contrast, the CORPUS-BASED approach is often understood as the reverse of the corpus-driven approach in the sense that here the corpus is treated as a source of examples and possibly frequency information needed to confirm or disprove some existing theory or hypothesis (cf. Meyer 2014: 15).

So the question is: which type does our study belong to? The use of corpus material in the present work is versatile. On the one hand, we confront the existing theoretical claims with empirical evidence, indicate counterexamples, and test hypotheses, which brings us close to the corpus-based approach. As it does to Biber (2015), corpus analysis offers us the perfect methodology for identifying the most frequent and most rare patterns in the given discourse variety, often counter to prior expectations.

We are not the first to observe that expressions labelled “ungrammatical” by linguists have been found to be used by native speakers (cf. Sampson 2001, Stefanowitsch 2007, Bresnan & Nikitina 2009) or to be accepted by non-linguists (cf. Wasow & Arnold 2005, Bresnan 2007).

On the other hand, as we explained in Chapter 1, our study is rather data- than theory-oriented. We do not limit ourselves to examining known patterns: we also aim to explore the occurrences which have not yet appeared in theoretical approaches to CLs. In that way, corpus-driven approach helps us, at least partially, to overcome the problem of false negatives (rejection of true hypothesis), the matter which is usually neglected in the studies dealing with the accuracy of introspective and experimental data (cf. Sprouse & Almeida 2012: 611–612). Furthermore, all the claims that we formulate are based on material from corpora and further tested by statistical methods and/or additionally verified in cautiously designed psycholinguistic experiment in order to achieve higher control of particular factors and to reject observations which occur due to error. In this sense, our study meets some criteria of a corpus-driven study.

Hence, instead of drawing a sharp border between the two approaches, we are in favour of a HYBRID APPROACH (Biber 2015) which on the one hand admits the validity of predefined grammatical categories and syntactic features (such as CLs and CC), but involves corpus-driven methods in the inductive analysis of corpora on the other.

### 3 Empirical approach to clitics in BCS

#### 3.3.2.3 Corpora as a source of authentic data

As shown in Chapters 11, 13, and in 14 on the one hand we encounter large disagreement among scholars concerning the possibility of CC in certain contexts, and on the other we see an absolute lack of empirical studies. For many studies on CLs in BCS the following statement applies: “you imagine examples of language used in this or that situation and ask yourself whether they are grammatical or ungrammatical, natural or unnatural, appropriate or inappropriate” (Clark & Bangerter 2004: 25). In contrast, we believe that authentic data can help form and test hypotheses as well as settle ongoing disputes. Our first source is corpus data, which mainly fulfil the observatory function. We use corpora to provide counterexamples to theoretical claims.

Since some of the syntactic constructions we wanted to investigate, such as CC out of *da*<sub>2</sub>-complements, tend to have extremely low absolute numbers of occurrences, we decided to turn to large web corpora {bs, hr, sr}WaC.<sup>9,10</sup> Such corpora are collections of texts extracted from the world wide web and include many spontaneously produced, unedited texts, which gives prospects for valuable findings unlikely to be encountered in literary texts, often reviewed by editors with respect to some “standard” of language (cf. Gries & Newman 2013: 259).

#### 3.3.2.4 Limitations of corpus linguistics

We have to be aware that corpus linguistic methodology has its limitations arising mainly from the nature of the data with which it deals. First, one well-known drawback is no possibility of providing evidence of absence. In other words, the lack of occurrence of a certain structure in the corpus is not proof of its unacceptability, as the reason for it may be purely accidental. Additionally, while statistical tests may show that a given construction is improbable, they cannot give a reason for this improbability (cf. Stefanowitsch 2006). Secondly, corpora contain records of speech, and therefore all tests concern language users’ performance but not their competence. Furthermore, it is hard to assess the acceptability of the occurring structures ad hoc. Corpora also include accidental forms (e.g. mispronunciations or typing/writing errors) which can be misinterpreted as rare but possible forms. The usual assumption in big data is that the most frequent structures are the most grammatical while noise is rather infrequent (Kilgarriff &

---

<sup>9</sup>For basic information on *da*-complements see Section 2.5.3 and for empirical data on CC out of *da*<sub>2</sub>-complements see Chapter 13.

<sup>10</sup>For a detailed description of corpora available for BCS and our reasons for choosing to work with those corpora over others, see Chapter 4. For the queries used in our corpus studies see Chapter 12.

Grefenstette 2003: 9). Retrieving rare and complex structures is nevertheless challenging, and in the case of web corpora problems related to information retrieval accuracy measures – precision and recall – are impossible to overcome.<sup>11</sup> In order to gather more high-quality data we give preference to acceptability judgment tasks over elicitation of naturally occurring data through trigger questions in interview-based corpora, because we assume that the latter will still not provide us with numerous occurrences of the relevant structure and will not include all the context we are interested in. Although interview-based corpora would provide more ecologically valid data, we decided to systematically collect data which would fulfil all conditions necessary for inferential statistical methods in order to be able to make more robust generalisations. Furthermore, acceptability judgment experiments as an empirical approach seem more appropriate to us since they enable us to generalise from many individual ratings. This provides more accurate answers to the research questions addressed than uncounterbalanced interview data or data from a single linguist would.

### 3.3.3 Psycholinguistic experiments

#### 3.3.3.1 Types of psycholinguistic tasks

Next to corpus data (i.e. observational data), other techniques of collecting empirical data are available, for instance psychological responses to linguistic stimuli. We can divide the many experimental tasks into *NON-SPEEDED* (non-chronometric) tasks where reaction or response times are not collected and analysed as data, and *SPEEDED* (chronometric) *TASKS* (cf. Derwing et al. 2009: 237). While the former reflect only the final outcome of the psychological processes, the latter can reflect the time course of language processes (Myers 2017: 3).

Reaction time was first introduced by Donders (1868), whose main idea was that more complex cognitive tasks take more time to complete. Donders believed that cognitive operations are additive, i.e. that more complex tasks take longer because more cognitive operations are recruited. In accordance with this belief, he proposed the famous method of subtracting reaction times in a series of tasks that differed in only one cognitive operation, in order to determine the time taken by the additional cognitive operation. Although the original hypothesis on the additive nature of cognitive operations has been abandoned, the idea of reaction time as the indicator of cognitive load or processing cost has survived. Reaction time is one of the most frequently used behavioural measures in psychology and psycholinguistics (Luce 1986).

---

<sup>11</sup>For example in our study this is the case of CC out of object-control CTPs (see Section 14.4 and Chapter 15).

### 3 Empirical approach to clitics in BCS

Non-chronometric tasks include:<sup>12</sup>

1. segmentation tasks,
2. rating tasks,
3. string manipulation tasks (or “experimental word games”),
4. manipulation of miniature artificial real language subsystems,
5. stimulus classification (or “concept formation”),
6. recall and recognition tasks (Derwing et al. 2009: 240).

One of the most popular rating/scaling experiments used in syntax is the acceptability judgment task (cf. Derwing et al. 2009: 244), which is used in order to indirectly access grammaticality.

In order to avoid problems related to obtaining linguistic data exclusively from informal acceptability judgments, which we discussed in sections above, we decided to conduct what we call an acceptability judgment experiment with non-linguists. In the literature this method is also referred to by the terms well-formedness, nativeness, naturalness and grammaticality (cf. Myers 2017: 2).<sup>13</sup> For the reasons given in the following subsection we use the term acceptability.

#### 3.3.3.2 What exactly does an acceptability judgment test measure?

Traditionally speakers’ reactions to sentences have been called “grammaticality judgments” (Schütze & Sprouse 2013: 27), but in our view this term is misleading. Based on Chomsky (1965: 4, 11f) linguists generally agree that grammaticality and acceptability are two distinct concepts.<sup>14</sup> The former refers to whether a sentence conforms to the rules of grammar, while the latter, to the degree to

---

<sup>12</sup>A detailed description of each task and some examples of the concrete experiments conducted can be found in Derwing et al. (2009).

<sup>13</sup>Wordlikeness is a term often used in morphology and lexical phonology research (Myers 2017: 2).

<sup>14</sup>Chomsky (1965) clearly distinguishes between competence (grammar knowledge) and performance (a decision based on grammar knowledge). “We thus make a fundamental distinction between competence (the speaker-hearer’s knowledge of his language) and performance (the actual use of language in concrete situations). Only under the idealisation set forth in the preceding paragraph is performance a direct reflection of competence. In actual fact, it obviously could not directly reflect competence” (Chomsky 1965: 4). “Acceptability is a concept that belongs to the study of performance, whereas grammaticalness belongs to the study of competence” (Chomsky 1965: 11). “The notion “acceptable” is not to be confused with “grammatical”” (Chomsky 1965: 11).



which a sentence is judged by native speakers to be permissible in their language. On the one hand, sentences which are perfectly grammatical can be evaluated as unacceptable because they violate some prescriptive or pragmatic rules. On the other hand, sentences which are ungrammatical can be evaluated as acceptable depending on the informants' ability to imagine necessary, though missing, context. Therefore, some scholars such as Featherston (2005: 701f) propose abandoning the mentioned difference between acceptability and grammaticality, and argue that grammaticality can be operationalised only in terms of acceptability (cf. Featherston 2005: 674, 701f, Riemer 2009: 624). Following the latter approach, we can prove or falsify existing and potential syntactic theories, since the results of carefully constructed, relative acceptability judgments used as empirical data approximate grammaticality, which normally is not directly accessible, possibly closely (Newmeyer 1983: 51, Schütze 2016: 26, Featherston 2007: 402f). In applying this experimental method we must not forget that informants' judgments are not influenced only by grammatical, but also by extragrammatical factors. In order to avoid or neutralise the influence of the latter, researchers take various steps. For instance, they try to balance stimuli for length, lexical content, processing difficulty, plausibility, etc. as much as possible (see Schütze 2016, Cowart 1997, Featherston 2005 for further discussion).<sup>15</sup>

#### 3.3.3.3 Different types of judgment tasks

Acceptability judgments involve explicitly asking speakers to “judge” whether a particular string of words or graphemes/phonemes is a possible utterance of their language (Schütze & Sprouse 2013: 28). Acceptability judgments can be divided into two main categories: non-numerical or qualitative tasks, and numerical or quantitative tasks. While the former group includes yes-no and forced choice tasks, the latter group comprises the magnitude estimation task, Likert scale task, and the thermometer task, which have been designed to give us information about the size of the difference between the structures of interest (cf. Schütze & Sprouse 2013: 33ff).

The acceptability of a sentence can be judged using the Likert scale task. Participants are given a numerical scale (usually from 1 to 5, from 1 to 7 or from -3 to +3) whose endpoints are labelled acceptable or unacceptable, and they are asked to rate each stimulus on the scale (cf. Schütze & Sprouse 2013: 33). In this kind of experiment, the researcher normally provides examples for the highest (ceiling) and lowest (floor) point of the scale, i.e. completely acceptable and completely

---

<sup>15</sup>To balance does not necessarily mean to suppress those factors: as Cowart (1997: 47) puts it, they can be controlled for if they are uniformly spread across all the stimuli.

### 3 Empirical approach to clitics in BCS

unacceptable, which helps participants to take decisions during the experiment (cf. Schütze & Sprouse 2013: 37).

In the magnitude estimation experiment a reference sentence (called standard) with an arbitrary value (called modulus) is presented to participants and they are asked to ascribe values to all other stimuli in comparison to the standard, so if the new stimulus is twice as good as the standard, it has to be assigned a number which is also twice as high as the modulus, etc. (cf. Schütze & Sprouse 2013: 34).<sup>16</sup> In order to be able to express all their judgments relative to the standard stimulus, participants must have access to the standard sentence and its value (modulus) throughout the whole time of the experiment (Hoffmann 2013: 101).

Featherston (2008, 2009) proposed the thermometer task, which combines the intuitive nature of point scales with the sensitivity of the magnitude estimation task. In this kind of experiment, participants are presented with two reference sentences and their values (ceiling and floor of acceptability). Afterwards the values ascribed to the stimuli are plotted on a line relative to those two points.

The fourth solution is to let participants evaluate stimuli on a binary scale: acceptable vs unacceptable. In these so-called yes-no tasks, it is important for participants to be exposed to polarised sentences; therefore, besides target sentences, they should get target-like incorrect sentences. Fillers have to be polarised as well, otherwise the participants will start to evaluate acceptable sentences as unacceptable.<sup>17</sup>

The fifth possibility is the forced-choice task in which participants are faced with two (or more) sentences, and they are asked to select the most (or the least), in their opinion, acceptable sentence (cf. Schütze & Sprouse 2013: 31).

While considering which type of acceptability judgment task to choose, we had to bear in mind the following advantages and disadvantages of each of them. For instance, the magnitude estimation task is more sensitive to fine contrasts between different types of structures and the results can be statistically evaluated with parametric tests (cf. Dąbrowska 2010: 8).<sup>18</sup> Furthermore, the magnitude estimation task allows participants to rate stimuli on their own scales and not on

---

<sup>16</sup>The other possibility is to give a reference stimulus and ask participants to assign a number to it themselves (Hoffmann 2013: 100).

<sup>17</sup>Filler items are items (i.e. words or sentences) which are not related to the research question. Their main purpose is to reduce the chances of participants figuring out which sentence type is being tested, i.e. to avoid conscious response strategies (Schütze & Sprouse 2013: 39).

<sup>18</sup>“Parametric tests involve statistical approximations and rely on the sampled data being distributed in a particular way” (Gries 2013: 322). “There are differences between the inferences licensed by parametric and non-parametric tests. For example, when all of the assumptions are met, parametric tests can be used to make inferences about population parameters from the samples in the experiment. Non-parametric tests, which do not assume random sampling, can only be used to make inferences about the sample(s) in the experiment itself” (Schütze & Sprouse 2013: 44).

the scale provided by the researcher, i.e. artificial limitation of rating is avoided (Hoffmann 2013: 103). Compared to the Likert scale task, magnitude estimation is more time consuming and less intuitive. Namely, participants have to decide how many times better the stimulus is than the standard rather than deciding if a particular stimulus is closer to the “good” or “bad” end in the Likert scale (cf. Dąbrowska 2010: 8, Schütze & Sprouse 2013: 33, 35). Additional argument against such a time-consuming task comes from recent studies which showed that even in the case of magnitude estimation which should allow insight into fine differences between various kinds of structures, participants use a small set of numbers repeatedly instead of rating every stimulus differently. Thus it seems that they treat the magnitude estimation task as a type of Likert scale task (cf. Schütze & Sprouse 2013: 34f). Although some researchers object to the use of parametric tests in the case of Likert scale tasks, others argue that parametric tests are quite robust and that violations of the intervalness assumption have relatively little impact on the results. Thus, the use of parametric tests with data obtained using the Likert scale has become standard (cf. Blaikie 2003, Pell 2005).<sup>19</sup> Yes-no and forced-choice tasks were designed to qualitatively compare at least two conditions, but they do not catch the fine-grained differences between acceptable and borderline structures. On the other hand, they allow both the participants and the researchers to work quickly, which is important in the case of complex experiment design and shortage of participants (cf. Schütze & Sprouse 2013: 31ff).<sup>20</sup> Finally, it is worth pointing out that since in all judgment tasks participants are asked to do the same cognitive task, the data yielded by different kinds of tasks are likely to be very similar, especially in the case of large sample size (e.g. twenty-five participants or more), so the choice of task is relatively inconsequential (cf. Schütze & Sprouse 2013: 36).

#### 3.3.3.4 Acceptability thresholds for different types of judgment tasks

First of all, we need to state that acceptability is not a categorical, but a graded phenomenon (Lau et al. 2017). Data from acceptability tasks with various modes of presentation converge to form such a conclusion. If speakers are presented with an acceptability judgment scale, their average ratings will be distributed across the scale values. If the speakers are presented with a binary acceptability

---

<sup>19</sup>Furthermore, z-score transformation has been suggested as a possible solution, since it allows each participant’s response to be expressed on a standardised scale (cf. Schütze & Sprouse 2013: 34, 43).

<sup>20</sup>It seems that forced-choice tasks are much easier to develop and later on, conduct as an experiment since they do not need fillers (cf. Schütze & Sprouse 2013: 32).

### 3 Empirical approach to clitics in BCS

judgment (yes-no), a single speaker will always either accept or reject a sentence, but the proportion of speakers who accepted (or rejected) a sentence will differ.

Having in mind the continuous nature of acceptability, we face the problem of interpreting acceptability data. Extreme values are clearly easily interpreted as acceptable and unacceptable. However, the problem remains of how to interpret the middle ground. To the best of our knowledge, there is no established linguistic strategy that we could rely upon. Therefore, we look at the practices that are firmly established in empirical psychology, namely the measurement of sensation.

In psychology, if the task is to detect a stimulus (so-called detection task; e.g. “press yes if you hear something”), a stimulus is at the threshold value if it is detected in 50% of the trials (Weber 1834, Fechner 1860, Smith 2008, Jang et al. 2009, Goldstein 2010). If the task is to choose between two alternatives (so-called “two forced choice task”, where participants are presented with two alternatives – two stimuli and the task to pick one), 75% is taken as the threshold, as in this case 50% denotes guessing. Given that binary acceptability judgments cannot be treated as two forced choice tasks, as only one stimulus is presented at a time, 50% acceptance should be interpreted as the threshold. The definition of the threshold which applies is that it is the smallest intensity of stimulation for which 50% of participants declare that they were able to detect it (Smith 2008, Goldstein 2010).

With all this in mind we decided to adopt a 50% acceptance rate (i.e. acceptance by 50% of the speakers) as the threshold of acceptability. It is important to note that we do acknowledge the fact that acceptance is a graded phenomenon (as demonstrated by Lau et al. 2017) and we do not imply that there is a strict line between acceptable and unacceptable sentences. We intend to use this threshold only for the purposes of orienting.

#### 3.3.3.5 Pros and cons of judgment data

According to the literature, judgment data can provide *NEGATIVE DATA* and data which cannot be collected otherwise, i.e. on infrequent structures that fail to appear even in a very large corpus (such as web corpus) (Hoffmann 2013: 117, Krug & Sell 2013: 92, Rosenbach 2013: 280, Schütze & Sprouse 2013: 29). In other words, introspection experiments such as acceptability judgments allow rare phenomena to be investigated and negative data to be obtained (Hoffmann 2013: 100). Moreover, judgment data can be used whenever there is no corresponding corpus at all or to complement corpus data (Hoffmann 2013: 117). Furthermore, if we compare judgment data with spontaneous usage data, we should emphasise that the

latter include some proportion of production errors (slips of the tongue/pen/keyboard) which can later be misinterpreted as evidence for rare structures (Schütze & Sprouse 2013: 29). Another advantage is that researchers can influence and control the kind and amount of data which is being collected and later evaluate it relatively quickly (Krug & Sell 2013: 92). Additionally, we should underline that the accumulation of many informants' judgments produces supra-individual, less erratic intuition-based ratings, i.e. this kind of introspective data is claimed to be objective (Hoffmann 2013: 117, Krug & Sell 2013: 92).

However, there are also disadvantages to such an approach; for instance, experiment and stimuli preparation can be time-consuming as experiments have to be carefully designed (Hoffmann 2013: 117, Krug & Sell 2013: 92). Furthermore, researchers do not collect natural speech/writing, and stimuli are not usually taken from spontaneously produced language material (Krug & Sell 2013: 92). Acceptability judgments rely on informants' ratings and intuition and are not a direct investigation of actual language use; moreover, generalisations are limited to the specific conditions (combinations of observed factors) which were tested (Krug & Sell 2013: 92, Rosenbach 2013: 282).

#### 3.3.3.6 Outlook: production experiments

Finally, we would like to note that “in an ideal world” without human and funding restrictions we would have obtained naturalistic production data, which are versatile and have high ecological validity, like the data from WaC corpora.<sup>21</sup> Production experiments in the narrow sense as standardised procedures are an ideal case for researchers who want to be able to systematically manipulate some variables and control for the effect of others in order to collect data suitable for quantitative analysis (cf. Eisenbeiss 2010: 11). Such experiments can be non-speeded or speeded. Widely used tasks include:

1. elicited imitation experiments – participants are asked to imitate/repeat/paraphrase spoken sentences,
2. elicited production experiments – involve prompts to produce target forms,
3. speeded production experiments – participants must produce target forms,
4. syntactic priming – it is observed whether participants repeat syntactic structures of unrelated utterances,

---

<sup>21</sup>For more information on BCS WaC corpora see Section 4.4.

### 3 Empirical approach to clitics in BCS

5. the input/feedback experiment – participants get input or both input and feedback on correct form.<sup>22</sup>

One of the most used elicited production experiments in syntactic research is a paper-pencil task in which participants are asked to fill in gaps with target items. We will leave production tests for future research.

#### 3.3.4 Experiment chosen for our study

Since the magnitude estimation task can show fine-grained differences between the tested items and conditions, it is often considered the most appropriate measure of acceptability. Because of the assumption that grammaticality is gradient, it seems important to measure acceptability either with tasks like magnitude estimation or at least with Likert scales with many levels of measurement which would allow insights into this gradience.

However, Weskott & Fanselow (2011: 253) accurately point out that a certain degree of gradience may also be captured with binary yes-no scales. They emphasise that if each experimental condition is tested with at least four items, even the resulting mean values of the binary measures exhibit variability to some degree: a mean of four binary judgments can take on five different possible values (0, 0.25, 0.50, 0.75 and 1) (Weskott & Fanselow 2011: 253). Thus, it seems that even fixed-scale judgments with a small number of points like binary scales can, depending on the number of observations gathered, exhibit a certain range of variability, and are not per se less suited to represent gradient acceptability than for instance magnitude estimation (Weskott & Fanselow 2011: 253).

Since several studies (e.g. Bader & Häussler 2010, Weskott & Fanselow 2011, Fukuda et al. 2012) showed that acceptability judgment tasks with different response types give very similar results and since binary scales can capture gradience in a similar way to numerical scales, we decided to use speeded yes-no acceptability judgment.

As we showed in the previous subsections, the discussion of the best acceptability judgment task boils down to a trade-off between ease of application by the participants, statistical power, and time consumed by preparation and data processing. In respect of this, although the yes-no task is more demanding for the researcher in terms of data collection (as it requires more participants and more items per condition), it was our task of choice due to its advantages from the perspective of participants and the methodological advantages related to eliminating strategic responding. We refer to the ease with which participants can grasp

---

<sup>22</sup>For more details and examples of tasks used in each of the mentioned methods including descriptions of procedures see Eisenbeiss (2010).

### *3.3 Empirical approach in the current study*

the basic idea behind the task, i.e. what is expected from them. By informing the participants that the time allowed for each trial, although more than enough for their decision, will nevertheless be limited, we additionally strengthen the explicit instruction to reply intuitively, without overthinking. This way we also reduce the possibility of the participants building some kind of strategy while responding. In addition to being less likely to involve overthinking of each response, making a simple yes-no decision is also less time consuming, thus allowing a larger number of responses to be collected during the same total time. Although the need to collect data on more items per condition (which is more strongly recommended for the yes-no task compared to some other tasks) may seem a disadvantage of this task, it can also be viewed as an advantage, or even as an obligation. As Clark (1973) noticed, the peculiarity of psycholinguistic research is double-sampling. While sampling from the population of speakers, researchers also sample from the population of language items. In other words, the researchers' aim is to be able to generalise their conclusions to all speakers, but also to all items of a chosen type (as opposed to relating conclusions only to the specific examples presented in the experiment). Therefore, as well as including multiple speakers in the experiment, one must also include multiple items per condition.

Finally, the yes-no task (as a simple form of binary choice for participants) enables us to record response time, i.e. the time taken by participants to categorise each item as acceptable or as unacceptable. The long history of empirical research in psychology has demonstrated that complex tasks incur longer response latencies. In terms of language research – items that are rarely encountered, unusual, or complex take longer to process. Therefore, we expect items that are accepted by more participants to also elicit shorter response latencies, and vice versa (those that are rarely accepted should elicit longer reaction time). Having two measures (acceptance and response latency) for each token of interest we obtain two indicators of the same underlying speakers' disposition, thus increasing the reliability of our research. Also, it should be noted that whereas participants' responses could potentially be affected by response strategies, it is hard to imagine how speakers could build a strategy to control their processing time.





# 4 Corpora for Bosnian, Croatian, and Serbian

## 4.1 Introduction

The goal of this chapter is to explain the choice of corpora used to extract linguistic evidence, formulate further hypotheses and find examples of the language structures in focus. As explained in the previous chapter our approach to CLs in BCS is primarily empirical and not oriented towards any particular working grammatical framework. Therefore, the role of data in our research is not limited to extracting examples confirming or contradicting the existing theories: we principally use corpora inductively to identify patterns which form regularities and exceptions concerning the behaviour of CLs.

The rest of this chapter is structured as follows: Section 4.2 discusses the double meaning of the term *corpus* in linguistics and briefly summarises types of electronic text sources. In Section 4.3 we present an overview of the most important corpora for BCS with a special focus on web corpora in Section 4.4. Section 4.5 discusses available corpora of spoken language. Section 4.6 presents some concluding remarks.

## 4.2 Some remarks on corpus types

### 4.2.1 The meanings of the term *corpus* in modern linguistics

In order to analyse the advantages and disadvantages of the available corpora we should first review the term corpus and discuss types of corpora, as this kind of collection of data forms a very heterogeneous group. Let us start from a very broad characteristic of the term corpus given by McEnery & Wilson (1996: 21):

In principle, any collection of more than one text can be called a corpus: the term ‘corpus’ is simply the Latin for ‘body’, hence a corpus may be defined as any body of text. It need imply nothing more. But the term ‘corpus’ when used in the context of modern linguistics tends most frequently to

have more specific connotations than this simple definition provides for. These may be considered under four main headings:

- sampling and representativeness,
- finite size,
- machine-readable form,
- a standard reference.

Thus, a corpus is a collection of naturally-occurring language documentation, gathered with respect to some particular framework. This framework can be either oriented towards characterising a particular type of language Sinclair (1991: 171) and results in both general reference corpora and small, specialised corpora, or towards studying a particular linguistic phenomenon, in which case often only particular types of structures are stored.

In order not to confuse these two approaches to data collection, we call the former type corpus<sub>1</sub>, whereas a corpus constructed in order to test a research hypothesis will be called corpus<sub>2</sub>.

These two approaches to data collection should be kept separate, as a corpus<sub>1</sub> can be a potential source of material for a corpus<sub>2</sub>. A corpus<sub>2</sub>, however, can rarely be used in the function of a corpus<sub>1</sub>, unless the linguistic phenomenon under study is a certain variety of a language as a whole.

Hence, in the present chapter we focus on the available corpora in the broader sense of corpus<sub>1</sub> in order to find out which of them can best serve the extraction of a representative collection of naturally occurring utterances, relevant to our project on CLs in BCS.

#### 4.2.2 Types of corpora as text collections

When describing corpora as text collections, it is important to cover several parameters. First, whether the corpus contains written or spoken texts. Secondly, whether the corpus is monolingual or multilingual. Among multilingual varieties there are parallel corpora – where one piece of semantic content is represented in several languages (source language and one or more translations), and comparable corpora – where texts with similar characteristics (register, lexical content, style, genre) have been collected from several languages. In the case of comparable corpora it is important for the proportions of features according to which the stratification takes place to be preserved in all languages.

In the present study we focus on the microvariation of CLs in BCS, which excludes the use of parallel corpora as a source because of the interference from

### 4.3 Overview of traditionally compiled corpora for BCS

the source text. This is because “phenomena pertaining to the make-up of the source text tend to be transferred to the target text” (Toury 1995: 275). Therefore, in the next part we will focus on the description of monolingual corpora of BCS, including both written and spoken varieties.

The features important for linguistic studies are size, content and sampling principles, which allow to assess for what language varieties the given corpus is representative and which research questions can be studied with its help. In this respect modern corpora can be divided into traditional sources which follow certain priorly defined principles and criteria (stratified sample), and opportunistic collections compiled from what is easily available and accessible (convenience sample). Stratified samples are typical of reference and monitor corpora, whose task is to reflect the “real-life” state of language, and of small corpora compiled for specific research purposes.

However, most publicly available corpora are nowadays collected for purposes of computational linguistics and here corpus size is the deciding factor. As Manning & Schütze (1999: 120) point out, “in Statistical NLP, one commonly receives as a corpus a certain amount of data from a certain domain of interests, without having any say how it is constructed.” The most popular solution is crawling the internet. We elaborate on this approach in Section 4.4.

In the overview below, we will describe both the smaller traditional and the more impressive in size opportunistic sources, and discuss their utility for studying microvariation of CLs in BCS. The overview represents the state-of-the-art for the period 2015–2018 when the study data were retrieved. Due to rapid technological development, mainly the increase of storage possibilities and computational capacities, new resources appear very quickly, so some sources available now are not mentioned.

## 4.3 Overview of traditionally compiled corpora for BCS

### 4.3.1 Bosnian corpora

Unfortunately, even now the range of available text collections for the Bosnian language is rather narrow.

The first widely available digital corpus of Bosnian is the Oslo Corpus of Bosnian Texts (OCBT, Santos 1998). It was the only larger corpus of Bosnian for a long time until recently, when bsWaC and SETimes (Ljubešić & Klubička 2014: 30) were compiled.

OCBT was created as a joint project of the Department for East European and Oriental Studies and the Text Laboratory of the University of Oslo. The main

## 4 Corpora for Bosnian, Croatian, and Serbian

Table 4.1: Bosnian corpora

Corpus	Words	Text age	Text types	Lemm.	PoS	MSD
OCBT	1,500,000	1989–1997	fiction, essays, newspapers, children’s books, Islamic texts, legal texts, folklore	yes	no	no

goal was to make Bosnian texts from the period 1989–1997 available for linguistic research (Santos 1998). The corpus is accessible for online search after registering for a free account. Table 4.1 summarises the most important facts about the OCBT.<sup>1</sup> The OCBT contains written texts belonging to different genres and its estimated size is 1,500,000 words (Santos 1998). It is searchable online through the Corpus Query Processor (CQP). The interface is rather simple, as it provides only concordances of words, phrases, suffixes, prefixes, or their combinations (Santos 1998). Functionalities often considered a standard, such as sorting and filtering, are not available. As to metainformation, the origins of texts are provided, which is a great advantage, but no morphosyntactic annotation is applied. Moreover, the OCBT is useful mostly for studies of standard language, as can be read from the content description in Table 4.1.

Summing up, it appears that the only traditionally compiled, monolingual source of Bosnian is the OCBT and therefore this language is definitely under-resourced. The Oslo Corpus of Bosnian Texts is certainly quite diversified with respect to the functional styles which it includes, but the texts are quite old, as they originate from the first development phase of standard Bosnian. However, the biggest objection to using this corpus in our study is that morphosyntactic annotation, which would allow efficient searching, is not available.

### 4.3.2 Croatian corpora

The two publicly available corpora of standard Croatian are *Hrvatski nacionalni korpus* (Tadić 2002, 2020), that is, the Croatian National Corpus (CNC) and *Hrvatska jezična riznica* (Čavar & Brozović-Rončević 2011, 2012), that is, the Croa-

<sup>1</sup>MSD stands for “morphosyntactic descriptions”.

### 4.3 Overview of traditionally compiled corpora for BCS

tian Language Corpus (Riznica). Table 4.2 gives basic information about these sources.

Table 4.2: Croatian corpora

Corpus	Words	Text age	Text types	Lemm.	PoS	MSD
CNC	2,559,160	1990–	Croatian literature, journals, and newspapers, booklets, official letters	yes	yes	yes
Riznica	84,536,657	since 2 <sup>nd</sup> half of the 19 <sup>th</sup> century	Croatian literature, non-fiction: scientific publications, online journals, and newspapers	yes	yes	yes

CNC was the first widely available digital corpus of contemporary Croatian language. The most current, third version comprises 216.8 million tokens. It is available via a NoSketchEngine interface, which allows complex queries to be constructed using the syntax of Corpus Query Languages (CQL).

The main goal of the project initiated at the Department of Linguistics (University of Zagreb) was to construct a corpus which would be big enough to cover the whole scope of standard Croatian in order to generate a primary source of linguistic data for lexicographical, orthographical, morphological, syntactic, and semantic research on contemporary Croatian (Tadić 1998: 339). During compilation, special attention was paid to the desired ideal corpus structure, according to which reference corpora such as the British National Corpus are built. This aim has not been achieved yet, mainly because a spoken corpus has yet to be created (Tadić 1998: 346, 2002: 446).

#### 4 Corpora for Bosnian, Croatian, and Serbian

The sampling frame was based on a variety of media, text types, genres, fields, and topics (Tadić 2002: 442) according to the standards for text typologies (EAGLES 1996). It is important to emphasise that only texts from 1990 on were incorporated into the corpus since the Croatian language could only develop without any obstructions starting from that period (Tadić 2002: 442).

The CNC is lemmatised and morphosyntactically annotated. However, from the user perspective we have to make the objection that neither is it easy to find a description of the morphosyntactic tagset in use, nor to follow how attributes should be built.<sup>2</sup> In many cases the theoretically available opportunities fail, because for example overly long concordances in CQL seem to be too complex for the corpus.

Riznica was compiled at the Institute of Croatian Language and Linguistics in Zagreb.<sup>3</sup> The goal was to produce a publicly available linguistic resource on the Croatian language and to provide crucial information about the Croatian language standard (Čavar & Brozović-Rončević 2012: 51). The collection covers texts written in various functional domains and genres and dated from the second half of the 19<sup>th</sup> century onwards.<sup>4</sup> It includes essential Croatian literature, including poetry, scientific publications from various domains, online journals, and newspapers. In contrast to the CNC, Riznica also contains translated literature from outstanding Croatian translators.<sup>5</sup> Because of its rigorously selected texts, Riznica as a corpus could be an interesting object of linguistic research as long as the intention was to explore how the desired structures should behave in proper, standardised Croatian. Riznica only became an attractive source of standardised language in 2018, that is in the last phase of our study, when its new release allowed for part-of-speech searches as well as for queries concerning morphosyntactic structures.<sup>6</sup>

---

<sup>2</sup>For a thorough insight into tagset visit <http://nl.ijs.si/ME/V4/msd/html/msd-hr.html>.

<sup>3</sup>Over the period of the current project, Riznica went through an incredible change concerning meta-information and the corpus manager.

<sup>4</sup>The official description of the corpus states that it is compiled from texts from the period of the standardisation of Croatian (Čavar & Brozović-Rončević 2012: 52, <http://riznica.ihjj.hr/dokumentacija/index.en.html>). However, texts from previous centuries, such as *Planine* by P. Zoranić, or *Judita* by M. Marulić, may be found during querying.

<sup>5</sup>For more information visit <http://riznica.ihjj.hr/dokumentacija/index.en.html>.

<sup>6</sup>According to the official description of the corpus, the first release should be annotated for lemma and word-class (Čavar & Brozović-Rončević 2012: 52); however, when queries are made through <http://riznica.ihjj.hr/index.hr.html> that kind of functionality is not available. The newest version was annotated with ReLDI tagger (Ljubešić & Erjavec 2016a), and is available via CLARIN-SI at [https://www.clarin.si/noske/run.cgi/corp\\_info?corpname=riznica&struct\\_attr\\_stats=1](https://www.clarin.si/noske/run.cgi/corp_info?corpname=riznica&struct_attr_stats=1).

### 4.3.3 Serbian corpora

Even though the range of available corpora of Serbian is not that narrow, in the literature we often find statements that Serbian is an under-resourced language with respect to the availability of electronic corpora (Dobrić 2012: 685, Balvet et al. 2014: 4106). Table 4.3 gives an overview of the best-known available digital corpora of the Serbian language.

*Korpus savremenog srpskog jezika*, that is the Corpus of contemporary Serbian (SrpKor2003), has been accessible online since 2002 (Utvić 2011: 41a). Its first version, NETK, lacked information about text sources, which was incorporated into the newer version. Both corpora are still available online; nevertheless, authorisation is required to use them. NETK and SrpKor2003 are monolingual corpora of raw texts written in the 20<sup>th</sup> century and belonging to different functional styles (Krstev & Vitas 2005).<sup>7</sup> They contain 22,000,000 words. The concepts which were important for the development of these corpora are described in Vitas et al. (2000).

The latest version, SrpKor2013, was released in 2013. It can be queried through an interface available for non-commercial purposes after registration. SrpKor2013 contains 122,000,000 words and it is the largest corpus of Serbian compiled in a traditional way. SrpKor2013 is lemmatised, and annotated for parts-of-speech. The markup scheme contains 16 tags (Utvić 2011: 43a). The existing documentation is very limited, so it is hard to draw many conclusions about the existence of further, for instance morphosyntactic, markup. SrpKor2013 is composed exclusively of written language and its texts can be divided into five functional styles (literary, scientific, publicistic, administrative, and others) (Utvić 2011: 42a). Bibliographic metainformation is provided for all texts and searches may only be performed separately for individual styles. From the description of the corpus it may be concluded that it also contains translation and some texts from online portals. Nonetheless, local experts consider the existing version of the Corpus of contemporary Serbian to be insufficient. In their opinion, in a new release more attention should be paid to achieving a balance between registers (Utvić 2011: 42a).

One part of SrpKor2013 has been extracted as a separate corpus under the name SrpLemKor. It is a lemmatised and PoS annotated corpus which consists of 3,763,352 words. This is the only part of the corpus for which the proportions of texts from particular registers are given in the documentation.

Korpus srpskog jezika (KSJ) is well described in Kostić (2003). It comprises 11,000,000 words. Dobrić (2009: 47) emphasises its diachronic dimension, which

---

<sup>7</sup>It is important to emphasise that both corpora are available only without an annotation layer.

#### 4 Corpora for Bosnian, Croatian, and Serbian

Table 4.3: Serbian corpora

Corpus	Words	Text age	Text types	Lemm.	PoS	MSD
NETK	22,000,000	since 1920	literature, scientific, and journalistic texts	no	no	no
SrpKor2003	22,000,000	since 1920	literature, scientific, and journalistic texts	no	no	no
SrpKor2013	122,000,000	from 20 <sup>th</sup> and 21 <sup>st</sup> century	literature, scientific, publicistic, administrative, and other texts	yes	yes	no
SrpLemKor	3,763,352	from 20 <sup>th</sup> and 21 <sup>st</sup> century	general fiction, literature, scientific, and legislative texts	yes	yes	no
KSJ	11,000,000	from 12 <sup>th</sup> up to 20 <sup>th</sup> century	literature, scientific, and legislative texts	yes	yes	yes



is reflected in the inclusion of texts dating back to the 12<sup>th</sup> century. On the other hand, KSJ does not cover spoken language or many contemporary texts. Additionally no clear line can be drawn between Croatian, Serbian, and Serbo-Croatian where texts originating from the second half of the 20<sup>th</sup> century are concerned. Undoubtedly the biggest advantage of KSJ is its detailed annotation, consisting of the grammatical status of each word, number of graphemes, syllable division and phonological structure, which was completed manually. Nevertheless, the main problem with KSJ is its accessibility. In personal communication with Dušica Filipović Đurđević and Aleksandar Kostić, son of the corpus compiler Dorđe Kostić, we found out that querying the corpus is possible only indirectly. One has to contact Aleksandar Kostić with a precise description of the data necessary and then the members of the Department of Psychology of the University of Belgrade extract concordances and send them back. Needless to say, first, such a mode of work is not very convenient and secondly, it is very likely that only simple queries are possible.

Summing up, considerable resources exist for Serbian, but similarly to other BCS corpora, they suffer from certain drawbacks. First, the annotation and searchability do not fulfil current standards. In this respect, the biggest problem seems to be the extremely limited possibility of using morphosyntactic annotation in queries. Secondly, many sources also contain diachronic data, or possibly include other varieties of BCS which remain unannotated.

## 4.4 {bs,hr,sr}WaC

### 4.4.1 The concept of the Web as a Corpus

{bs,hr,sr}WaC (Ljubešić & Klubička 2014, Ljubešić & Erjavec 2016b,c,d) belong to the Web as Corpus family of corpora, first popularized by WaCky (Baroni et al. 2009). The idea of using the internet as a source of linguistic data was controversial at first and generated a discussion about the content of web pages, since in such cases the acquisition of material is less controlled than in the case of traditional corpora. However, within the last decade the concept has become more and more popular, in particular because it is faster and cheaper in comparison to the traditional way of compiling a corpus (Benko 2017: 43).

The lack of resources for most of the South Slavic languages, which we hope we managed to demonstrate above, has also been recognised by the group of linguists behind the Regional Linguistic Data Initiative ReLDI. Furthermore, for smaller languages we do not have the luxury of text sampling, since the amount of data written in these languages is limited by their population, in comparison

#### 4 Corpora for Bosnian, Croatian, and Serbian

to, for example, English or Spanish. On the other hand, this can be a point in favour of web data since a large part of all writings are available online and can be turned into a language corpus (Ljubešić & Erjavec 2011). Therefore, treating web corpora as fully-fledged language resources is certainly appropriate in the case of South Slavic languages.

We will now provide the key data about {bs,hr,sr}WaC, and discuss the problems and limitations of these three corpora.

##### 4.4.2 {bs,hr,sr}WaC in a nutshell

The {bs,hr,sr}WaC corpora are undoubtedly the largest existing corpora for each of the three languages. Some key statistical data are presented below in Table 4.4.

Table 4.4: {bs,hr,sr}WaC corpora

Corpus	Words	Tokens	Documents	Compiled
bsWaC	248,478,730	286,865,790	896,059	10/28/2017 17:23:26
hrWaC	1,210,021,198	1,397,757,548	3,611,090	10/28/2017 01:27:08
srWaC	476,888,297	554,627,647	1,353,238	10/28/2017 04:17:28

Numerical data show that hrWaC is definitely the biggest of them, as it is more than double the size of srWaC and nearly five times bigger than bsWaC. We can identify several reasons for this state of affairs. First, the size of the Croatian economy and market. Second, the proportion of content written in closely related languages which appears in the Bosnian web and which had to be eliminated. And last but not least, the fact that the authors of the project are Croatians, and therefore may be more dedicated to the development of tools for studying Croatian.

{bs,hr,sr}WaC are a family of top-level-domain corpora of Bosnian, Croatian, and Serbian, which are available for download and online work via the No-SketchEngine concordancer.<sup>8</sup> They are currently accessible through the same platform as the latest version of Riznica. Like Riznica, they have been automatically lemmatised and morphosyntactically annotated with the unified tagset pattern according to MULTEXT-East Morphosyntactic Specifications.<sup>9</sup> The tagsets for Croatian and Serbian are identical on the morphosyntactic level, apart from

<sup>8</sup>Other top-level-domains are: .ba, .hr, .rs, .biz, .com, .eu, .info, .net.

<sup>9</sup>For more information visit <http://nl.ijs.si/ME/V5/msd/html/msd-hr.html>. The newest version has been released in 2019, see <http://nl.ijs.si/ME/V6/msd/html/msd-hbs.html#msd.msds-hbs>.

one additional subset of tags for the synthetic future tense in Serbian (Ljubešić & Klubička 2014: 31).

The morphological annotation was performed automatically with an accuracy estimated at 92.5%, which fulfils the current standards in NLP (Ljubešić et al. 2016: 4269)

#### 4.4.3 WaC content

The main problem of corpora compiled from the web is the lack of metadata on corpus composition. This applies to all possible categories which are used to characterise traditionally compiled corpora (sociolinguistic information, text age, style, genre, and register). Web corpora can be characterised with technical information (domain, URL, date of update or upload, and size), and, if additionally processed, with internal linguistic factors such as size of lexicon and frequency of grammatical features. Such analyses are nevertheless time-consuming and can usually, due to the massive volume of data, explain only part of variance. Additionally, in order to evaluate web corpora as a source, similar analysis must be performed on traditionally compiled, representative sources, which, as stated above, barely exist in BCS.

Benko (2017) shows that, regardless of problems with characterising their exact content, web corpora should not be treated as an inferior type of data, but simply a different one. Furthermore, experiments conducted for English (Biber & Egbert 2016) and Czech (Cvrček et al. 2018) show that as far as internal linguistic features are in question, web data and traditional corpora overlap to a large extent.

Gato (2014: 62f) observes that although web corpora cannot cover all possible registers, they provide quite a wide spectrum, starting with formal legal texts on the one hand, and ending with informal blogs, and chat rooms on the other. It seems that the web contains both traditional genres adapted to the new medium, like newspaper and academic articles, and entirely new ones, such as tweets or Facebook entries, rarely included in traditionally compiled sources.

Finally, Schäfer & Bildhauer (2013: 106), authors of the German web corpus, come to the conclusion that web corpora do not generally perform noticeably worse than traditional ones of the same size. In addition, since size matters, it has to be said that large web corpora frequently outperform smaller traditional corpora (Schäfer & Bildhauer 2013: 106). In other words, although the contents of web corpora cannot be described in traditional terms, there are good reasons to assume that with respect to linguistic structure a massive corpus is better than a small one.

#### 4.4.4 Sources of noise

##### 4.4.4.1 Closely related languages

Another point of critique towards web corpora is noisy data. The creators of South-Slavic WaC corpora indicate two main sources of noise: first, documents written in other, closely related languages and secondly, texts of low quality (Ljubešić & Klubička 2014: 29).

In order to solve the problem of closely related languages, the creators used two classifiers: a blacklist classifier and unigram-level language models (Ljubešić & Klubička 2014: 32). Table 4.5 shows what share of documents in each corpus was identified as written in a closely related language (cf. Ljubešić & Klubička 2014: 33). The authors used a ternary classifier in bsWaC, where the share of foreign documents was the highest, and assumed that a binary classifier for hrWaC and srWac, which distinguishes only between Serbian and Croatian, is sufficiently informative.

Table 4.5: Distribution of identified languages through the three corpora

	bs	hr	sr
bsWaC	78.0%	16.5%	5.5%
hrWaC	-	99.7%	0.3%
srWaC	-	1.3%	98.7%

Nonetheless, although most documents in {bs,hr,sr}WaC are classified correctly, one should be aware that single paragraphs in closely related languages might still appear. This is mostly the result of reader comments, where the content of the document is generated by many users. Still, we think that such appearances should not affect our results because all unexpected occurrences can be checked manually before they are included in the data set.

An issue that is linked to content written in closely related languages is the occasional appearance of lexical elements from other South Slavic varieties. We must point out that even strictly monitored corpora such as the CNC contain words which, according to handbooks, do not belong to the Croatian standard, such as: *opšti* ‘general’ (*opći* in Croatian), *januar* ‘January’ (*siječanj* in Croatian), *sveštenik* ‘priest’ (*svećenik* in Croatian), *tačka* ‘dot, point’ (*točka* in Croatian). Although the authors of the CNC tried to minimise this phenomenon by selecting texts written after 1990, such word forms are present, for instance because

academic texts which discuss differences between Croatian and Serbian have been included. Similar evidence of non-Croatian word forms can be found also in Riznica, where the lemma *tačka* typical of Serbian is attested not only in academic texts, but also in texts by 19<sup>th</sup> century Croatian writers. In the same manner, Croatian word forms such as *nogomet* ‘football’ (*fudbal* in Serbian), *glazba* ‘music’ (*muzika* in Serbian) and *zrakoplov* ‘aircraft’ (*vazduhoplov*, *avion* in Serbian) can be found in SrpLemKor. A similar problem applies to web corpora, but on a larger scale.

#### 4.4.4.2 Non-standard language use and low quality data

The authors of the corpora approach the problem of low quality data with the assumption that most of the content of each web corpus can be qualified as good (Ljubešić & Klubička 2014: 33). In order to easily detect low quality text the most frequent types of deviation must be identified and classified. Above all, non-standard usage of the upper case, lower case and punctuation, and usage of non-standard language, understood as slang and dialects, belong here (Ljubešić & Klubička 2014: 33).

Not all these problems are easy to solve, but during the procedure of noise removal from the first release of {bs,hr,sr}WaC, it was postulated that a low percentage of diacritic characters should reflect less standard language usage and this assumption was used as a very simple estimate of text quality (Ljubešić & Klubička 2014: 34). In the second release, the REDI tool was used to restore diacritics, so that the texts could be correctly lemmatised and part-of-speech annotated.<sup>10</sup>

To this we can add the problem of avoiding standard punctuation, which can be partly related to specific writing style. It is commonly known that some texts, for instance those written on discussion fora, are characterised by a relaxed approach to punctuation and the use of symbols so even when those texts are built of several sentences they can contain hardly any periods at all (Schäfer & Bildhauer 2013: 90). Gato (2014: 43) also points out that online texts often contain misspelled words and grammatical mistakes, or include improper usage by non-natives.

Non-standard data could be an interesting object of CL research as they represent a very spontaneous, non-planned channel of communication, but they must be approached cautiously. Hence, results arising from non-standard language used online must always be checked manually to decide whether a particular divergence is caused by the relaxed use of language or carelessness of the language user, or else whether its source is non-native language use.

<sup>10</sup>The REDI tool is available at <https://github.com/clarinsi/redi>.

The second kind of low quality data typical of web documents are URLs, automatic translations, words split into fragments, and emoticons, but they do not affect our research much as they can be easily filtered out.

### 4.5 Corpora for spoken BCS

#### 4.5.1 Bosnian

In the area of spoken varieties the availability of resources is even lower than for written varieties. Building spoken corpora is related to higher costs understood in terms such as time, money, and manpower. Moreover, the workflow is more complicated since compiling a corpus of spoken data requires the same steps as in the case of written varieties, but additionally recordings must be obtained and transcribed. Spoken data is also further from theoretical language models and normative description, so many structures not included in normative descriptions, such as (dis)fluencemes, occur.<sup>11</sup> This poses a challenge for both human annotators and automatic taggers and lemmatisers. It comes as no surprise that only a handful of spoken corpora, compiled mainly for specific research purposes, are currently available.

The only corpus of Bosnian that we found was a corpus of narrative interviews compiled within the DFG-funded project *Corpus-based analysis of local and temporal deictics in (spontaneously) spoken and (reflected) written language*. The corpus is called Bosnian Interviews (Stevanović 1999) and was mainly transcribed and annotated by Slavica Stevanović. It used to be available for searches through an online interface, but currently access to its XML files can be obtained only on request. The data consist of 13 narrative conversation-situations with Bosnian refugees. The corpus is neither PoS annotated nor lemmatised, but tagging of v/t/n-deictics was performed for purposes of the above-mentioned research goal. An additional meta-layer of regional pronunciation is also featured. The formal description of the corpus is, nevertheless, very vague as, for example, the size of the corpus is not stated. We provide more details on this corpus in Chapter 8.

#### 4.5.2 Croatian

The Croatian Adult Spoken Language Corpus HrAL (Kuvač Kraljević & Hržica 2016) was built by sampling spontaneous conversations of 617 speakers from all Croatian counties, and it comprises over 250,000 tokens and over 100,000 types

---

<sup>11</sup>For more information on such structures see Section 8.5.

in 165 transcripts annotated with the ages and genders of the speakers, as well as the location of the conversation. It was compiled in three periods: 2010–2012, 2014–2015 and 2016. Croatian speakers from different parts of Croatia with access to groups of speakers (friends and families) were recruited and trained to collect samples of spoken language. Sampling was performed in informal situations, predominantly spontaneous conversations among friends, relatives or acquaintances during family meals, informal gatherings, and socialising. Thus the corpus contains rather short, often interrupted utterances.

### 4.5.3 Serbian

We are not aware of any publicly available, electronically stored corpus of spoken Serbian. Nevertheless, this gap might be filled in the near future, as some efforts towards building both Serbian and Bosnian spoken sources are being made, e.g. in a project at Humboldt-Universität zu Berlin.<sup>12</sup>

## 4.6 Discussion

### 4.6.1 The scope of available data

This section compares the properties of traditionally compiled corpora and web corpora for BCS with the goals of the study. As shown above, sources for corpus analysis in BCS are certainly limited. On the one hand, we have at our disposal rather sparse traditionally compiled corpora. They mostly represent language strongly influenced by normative prescription. Additionally, the languages are not equally represented if we compare size and type of data and the extent of annotation, which implies an individual approach to working with each corpus. The worst situation is in the area of spoken language and dialects, where little or no data can be identified.<sup>13</sup>

On the other hand, large data sets of unknown composition obtained from the web can be easily accessed and processed in a comparable way for all three South Slavic varieties in question. Although the language of web corpora cannot be described in traditional terms, a considerable share of the language represented in web corpora is not influenced by normative prescription, but is probably not worse as concerns linguistic richness than traditional data, as we hope we have shown in Section 4.4.3.

---

<sup>12</sup>For more information visit <https://www.slawistik.hu-berlin.de/de/fachgebiete/suedslawsw/colabnet/projects/spoc/spoc>.

<sup>13</sup>Some corpora such as Vuković (2021) were developed after our project was finished.

#### 4 Corpora for Bosnian, Croatian, and Serbian

Additionally, the analysis of url domain lists shows that web corpora not only cover texts typically included in corpora of standard language such as literary, journalistic, administrative, academic, and popular scientific texts, but also contain very new registers and genres that appear in user-generated content such as blogs and fora which are much closer to spontaneous language, even though written and not spoken (Schäfer & Bildhauer 2013: 4). This type of data is a valuable source of colloquial language and as such certainly relevant for studies of microvariation. Next to the available meta-information (allowing to track where the texts come from), size, and accessibility, such variety of data is a great advantage of web corpora which, at the same time, is hard to obtain from traditional sources.<sup>14</sup>

The question of the extent to which the available data can be considered representative appears. Following Biber (2005: 243), “representativeness refers to the extent to which a sample includes the full range of variability in population”. Ironic as it may seem, ideal representativeness is not possible to achieve. This is because however much corpus constructors try, they can only create a corpus which is the representation of itself, Kilgarriff & Grefenstette (2003: 1) claim.<sup>15</sup> Furthermore, the representativeness criterion seems useless nowadays, because web corpora do not contain that sort of metadata. Therefore, neither is it possible to check the range of text types they cover, nor can one be sure about the population of text types themselves, since the web covers a considerable, but not systematised, share of texts. Nonetheless, its variety and particularly its size counterbalance the limited information about its representativeness (Gato 2014: 45).

Manning & Schütze (1999: 120) argue that “having more training data is normally more useful than any concerns of balance, and one should simply use all the text that is available”. Furthermore, we agree with what Kilgarriff already noticed, namely that “it is the web that presents the most provocative questions about the nature of language” (cf. Kilgarriff 2001: 344).

---

<sup>14</sup>We are aware that the internet is often criticized for poor quality of texts, which includes numerous spelling errors, omission of diacritic signs and non-standard use of the upper and lower case and that this critique also pertains to texts from {bs, hr, sr}WaC. However, we would like to point out that for us these corpora are a source of authentic, spontaneously produced written texts, which were not under strict influence of the norm or externally corrected to look like prescribed standard Bosnian, Croatian or Serbian.

<sup>15</sup>Kilgarriff & Grefenstette (2003: 8f) list several reasons why corpora fail to represent real language usage. They draw attention to the arbitrariness which dominates in the text sampling, i.e. it is literally impossible to include all text types and topics (cf. Kilgarriff & Grefenstette 2003: 9).



Therefore, we follow Manning & Schütze (1999) and try to use a possibly broad scope of available corpora according to our goals. Apart from providing naturally-occurring, non-externally normativised and proofread language, corpora benefit our work on two topics. The first is CL placement and inventory in spoken varieties (see Part II). The second is an empirical approach to CC (see Part III), a controversial topic which has so far been studied exclusively in terms of theoretical syntax. In the two sections below we explain our choices as to the studied sources.

It is important to remember that currently digital sources develop very dynamically. During the period when the current project was conducted, some significant changes could be observed, such as the improvements in the morphosyntactic tagger for BCS and the new release of Riznica. Due to practical reasons, primarily time constraints, we could not benefit from all the advancements. Some parts of study were conducted on the old versions. Some corpora were rejected due to their poor quality at the time.

#### 4.6.2 Variation in spoken BCS

As presented in Section 4.5, the most under-resourced area of BCS is spoken sources. Thus, making statements about CL behaviour in spoken varieties and dialects based on corpus data is barely possible at the moment. The available spoken sources neither meet the standards applied to written corpora with regard to morphosyntactic annotation level, nor are they preprocessed with regard to transcript standards.

Work with both the Bosnian Interviews and HrAL corpora would require performing a high load of additional preprocessing. Importantly, the two corpora are not comparable. While Croatian transcripts contain mostly conversations, Bosnian Interviews are rather spoken narratives. As a consequence studying CLs is more feasible in the case of Bosnian data. Therefore, we decided that the first attempt to study the behaviour and distribution of CLs would be in spoken Bosnian, in particular concerning the influence of discourse structuring elements and disfluencemes on CL placement.<sup>16</sup> The results of this study, as well as a more detailed description of the corpus, based on our own explorations, are described in Chapter 8.

Given the lack of sufficient spoken dialectological corpora, we decided to work with the written sources described in detail in Chapter 7.

---

<sup>16</sup>For more information on discourse structuring elements and disfluencemes see Section 8.5.

### 4.6.3 Clitic climbing in BCS

In order to study the variation in constructions featuring verbal embeddings in the three South-Slavic varieties, a similar kind of data should be acquired for each variety. In that respect {bs,hr,sr}WaC are superior to other sources because, as explained above, a quite similar type of language variety is represented in all three web corpora. Additionally, the tagset and the query syntax are identical, so the results of searches are also comparable. Comparison in that respect across standard varieties on the basis of traditionally compiled corpora is barely possible, mainly due to the very limited searchability.

Web corpora are also unbeatable in terms of size. This increases the chances that even very rare variants of studied constructions will occur. For this reason they provide the best environment for examining the possibilities of CC from *da*<sub>2</sub>-complements in Serbian, as described in Chapter 13 and in Jurkiewicz-Rohrbacher, Hansen & Kolaković (2017).

Finally, as already mentioned, web corpora include user-generated content which represents spontaneous, non-edited, and thus very authentic language typical of ordinary users, present in fora, blogs, and reader comments. This type of language is not represented in the traditionally compiled corpora of BCS.<sup>17</sup> Since WaC are in a sense anonymous, as we rarely have access to sociolinguistic metadata, the possibilities for in-depth study of sociolinguistic variation are extremely limited. On the other hand, because Riznica has been available on the same online platform as hrWaC since spring 2018 and since it uses the same tagset as WaC, some conclusions can be drawn as to the factor of standard vs colloquial variety. Therefore in Chapter 14 we study CC from infinitive complements in Croatian in the forum.hr URL domain and in Riznica.

---

<sup>17</sup>In some languages, e.g. Czech, this type of language is steadily coming to be incorporated also in monitor corpora, e.g. Koditex (Zasina et al. 2018).

## **Part II**

# **Parameters of variation**



## 5 Parameters of variation: Inventory, internal organisation of cluster, and position

Part II is dedicated to variation in the diatopic and the diaphasic dimensions. It is based on the first two steps of our methodological approach, that is, on intuition/theory and observation. The first step always involves a thorough analysis of the existing research literature, independently of the relevant theoretical framework.<sup>1</sup>

We start with the literature-based Chapter 6, which compares and sums up the treatment of CLs in the three standard varieties. First we review the principal prescriptive handbooks. This information on CLs is complemented by related theoretical studies on CLs. Although in the latter studies BCS is usually treated as one abstract system, they, like the prescriptive handbooks, help us detect differences in BCS standard varieties, mostly through contradictory statements on the acceptability of certain structures.<sup>2</sup> Furthermore, in Chapter 6 we compile information that some authors give on variation in the CL system with respect to different registers, i.e. diaphasic variation.

Chapter 7 which follows provides complementary information. As there are no studies dealing specifically with CLs in dialects, in this chapter, like in the preceding one, we apply only the first two steps of our methodological approach: intuition/theory and observation. We summarise and synthesise data from the extensive dialectological literature, which usually consists of holistic descriptions of the grammatical and lexical systems of small local idioms. We consider these data to be highly valuable as they provide insights into the spoken idioms which might influence not only the colloquial varieties but also the standard norms. The CL system in Kajkavian and Čakavian differs significantly from the CL system in Štokavian dialects. Furthermore, some Štokavian dialects served as the

---

<sup>1</sup>A detailed description of our methodological approach can be found in Section 3.3.

<sup>2</sup>Although the authors of these works treat BCS as one abstract system when discussing whether certain structures are possible or not, i.e. acceptable or not, they usually use their own sense of language, their own dialects or idiolects as a baseline for comparison. Consequently, readers can easily find contradictory statements on the CL system when comparing such works.

## 5 *Parameters of variation*

base for the contemporary BCS standard varieties. We therefore focus mainly on Štokavian dialects. Nevertheless, we sometimes touch on the CL system in the Kajkavian and Čakavian dialects, mostly to comment on features which appear in contact dialects.

Chapter 8 goes one methodological step further as it describes an empirical study on the usage of CLs in spoken variety based on corpus data. We would like to emphasise that this is the first ever, pilot study on CLs in spoken BCS. In it we develop an annotation scheme for spoken data which takes into consideration the peculiarities of the syntax of spoken language. These are for instance disfluency phenomena which make it more difficult to establish clause boundaries and consequently to determine the position of the CL or the CL cluster in the clause. This study brings to the fore not only interesting findings concerning the CL inventory, internal organisation of the cluster, and morphological processes within it, but also insights into usage-based patterns of CL placement and, most importantly, into the heaviness of the host (or in the case of DP, of the host and the initial phrase) in spoken variety. This is the very first study on the heaviness of phrases preceding CLs in BCS that goes beyond linguists' intuitive judgments. It is based on Kosek et al.'s (2018) approach to measuring heaviness.

In Chapter 9, we summarise the findings presented in the previous three chapters and identify some global patterns of microvariation. We present new findings concerning language prescription in the normative handbooks, which is based on the conscious selection of only some of the features attested in non-standard varieties.

# 6 Clitics and variation in grammaticography and related work (Bosnian, Croatian, Serbian)

## 6.1 Introduction

The goal of this literature-based chapter is to present the current state of the art on CL systems in grammaticography and related work, with reference to variation. Furthermore, we compile information some authors give concerning different registers, i.e. diaphasic variation. It should be noted that there are actually no works on CLs focusing specifically on variation; neither are there any empirical variationist studies on CLs.<sup>1</sup>

Since our aim is a deeper empirical investigation of CLs, at this stage the most important goal is to detect possible instances of microvariation in the CL system with the help of the parameters of variation outlined in Chapter 2. Afterwards the selected CL phenomena recognised as a source of variation can be thoroughly investigated. It goes without saying that we must discuss here approaches authors adhere to. Some of them favour the phonological one whereas others share a more formal syntactic orientation.

In the next section, we explain our strategy, which allowed us to gather rather the scattered data on variation. Subsequent sections of this chapter follow the order in which we presented the parameters of microvariation in Chapter 2. Section 6.3 gives an overview of the inventory of CLs. In Section 6.4 we present the rules of CL clusterisation and morphonological changes which occur within CL clusters. Section 6.5 on position of the CL or CL cluster follows, in which we focus on the placement of CLs after breaks, heavy phrases, conjunctions, and complementisers. In the same section, we discuss in detail 2P, phrase splitting and DP of CLs. Special attention is given to the concept of 2P and different views on it.

---

<sup>1</sup>Information on this topic is scattered around the quoted works. Furthermore, we were able to find some information on CLs in respect of diachronic variation, but none of the authors mentioned diastatic variation, i.e. differences in the language use of individual social groups.

The final subsection addresses the problem of phrase splitting and the syntactic contexts which allow it.

## 6.2 Detecting variation

This chapter follows the first step of our empirical approach presented in Chapter 1 and described in some detail in Chapter 3. We approach this goal by inspecting grammar books for Bosnian, Croatian, and Serbian written by native authors, i.e. first we try to detect both systemic and sociolinguistic microvariation in each standard variety. A comparison of the existing descriptions of CLs in the grammar books, which are widely used in Bosnia, Croatia, and Serbia, enable us to detect the first level of variation in the CL system: diatopic variation between standard varieties of BCS on the level of prescribed language usage.<sup>2</sup> We chose grammar books which are currently in use as handbooks in schools or at universities. We believe that the grammars on our list reflect the current state of the CL system in standard varieties of the mentioned three countries.

Since there is no long tradition of Bosnian grammaticography, we chose the first post-Yugoslavian grammar book, *Gramatika bosanskoga jezika* by Jahić et al. (2000).<sup>3</sup> In addition to this work, we decided to take into consideration the descriptions of CLs in *Bosnian for foreigners with a comprehensive grammar* (Ridjanović 2012) written in English.

Among Croatian grammar books we thoroughly examined Katičić's (1986) highly influential *Sintaksa hrvatskoga književnog jezika*. Further, we observed how CLs are presented in *Gramatika hrvatskoga jezika*, a handbook used in elementary education (Težak & Babić 1996). We also took into consideration *Gramatika hrvatskoga jezika: za gimnazije i visoka učilišta* (Silić & Pranjković 2007) and *Hrvatska gramatika* (Barić et al. 1997), which are used in high school education and by students of the Croatian language at Croatian universities.

As to Serbian grammars, we deliberately started the analysis from an older book, namely Stevanović's (1975) *Savremeni srpskohrvatski jezik*, in order to see if there have been any changes in the CL system or in the norm. Next we included the high school grammar handbook *Gramatika srpskog jezika* written by Stanojčić & Popović (2002). Since CLs are a phenomenon which lies at the intersection of syntax with other disciplines, we analysed *Sintaksa savremenoga*

---

<sup>2</sup>We must point out that although the grammar books used imply the description of some kind of standard variety (the examples which authors provide are definitely neither colloquial nor dialectal), they are not labeled as normative grammar books, except Piper & Klajn (2014).

<sup>3</sup>The first grammar book which has the word Bosnian in its title was *Gramatika bosanskoga jezika za srednje škole* (Vuletić 1890).



*srpskog jezika* (Piper et al. 2005) as well. In addition, we took into consideration the recent *Normativna gramatika srpskog jezika* by Piper & Klajn (2014). Like for Bosnian, we analysed descriptions of CLs in one grammar book for foreigners, *Gramatika srpskog jezika za strance* by native authors Mrazović & Vukadinović (2009).

Besides variation between BCS standard varieties, browsing the grammar books revealed some diaphasic or diatopic variation within individual varieties. We were especially interested in so-called instances of common mistakes or deviations from standard language use. To us this actually indicates that there is some kind of variation within one variety, since native speakers of BCS normally have to learn the standard because they usually speak a dialect or some other kind of non-standard variety at home. We also inspected *jezični savjetnici* ‘language guidebooks’ for the same reason – they helped us detect both diatopic variation between BCS varieties, and diaphasic and diatopic variation within one variety.

However, since, as we already emphasised, we are not only interested in the CL system in the prescriptive norms of standard varieties of BCS. After grammar books and language guidebooks we consulted some other papers on CLs in which we could find scattered information on diatopic variation between standard varieties of BCS and on diaphasic variation (i.e. in different registers) in the CL system of BCS. It should be pointed out that most theoretical works on CLs in BCS presented in Chapter 2 do not mention variation or are not interested in it at all. Here, we refer not only to works addressing this topic like Radanović-Kocić (1988) and Milićević (2007), but also to works which do not address it directly. We found the latter especially interesting if they contained statements on CLs which contradict the statements of other scholars.

## 6.3 Inventory

### 6.3.1 Inventory of pronominal clitics in BCS standard varieties

The following table lists the CL and full forms of all personal pronouns in the genitive, accusative, and dative (cf. Barić et al. 1997: 208f, Mrazović & Vukadinović 2009: 363f).<sup>4</sup>

It is worth noting that the pronominal CLs for the third person singular have two CL variants in the accusative: *ga* and *nj* ‘him’ for masculine and neuter, and *ju* and *je* ‘her’ for feminine (cf. Mrazović & Vukadinović 2009: 366). The CL *nj* differs

---

<sup>4</sup>Although Mrazović & Vukadinović (2009: 363f) list only forms with a short rising accent in the paradigm, they admit that a short falling accent is also possible.

6 Clitics and variation in grammaticography and related work

Table 6.1: Pronominal CLs in BCS and their corresponding full forms

		Genitive/accusative		Dative	
		CL form	Full form	CL form	Full form
singular	I	<i>me</i>	<i>měne, mène</i>	<i>mi</i>	<i>měni, mèni</i>
	you	<i>te</i>	<i>těbe, tèbe</i>	<i>ti</i>	<i>těbi, tèbi</i>
	he, it	<i>ga/ga, nj</i>	<i>njěga, njèga</i>	<i>mu</i>	<i>njěmu, njèmu</i>
	she	<i>je/ju, je</i>	<i>njê/njû</i>	<i>joj</i>	<i>njôj</i>
plural	we	<i>nas</i>	<i>nâs</i>	<i>nam</i>	<i>nâma</i>
	you	<i>vas</i>	<i>vâs</i>	<i>vam</i>	<i>vâma</i>
	they	<i>ih</i>	<i>njîh</i>	<i>im</i>	<i>njîma</i>

syntactically from other CLs since it follows prepositions, does not clusterise and is not associated with the second position. Mrazović & Vukadinović (2009: 128) emphasise that it is felt to be archaic in Serbian.

We noticed some diatopic variation concerning the accusative CL *ju* between the three BCS standard varieties. Radanović-Kocić (1988: 56) states that “The accusative singular feminine CL *ju* is completely replaced by the genitive *je* in the Eastern variant, while it is still used in the Western variant.”<sup>5</sup> Whereas in the Yugoslavian period Croatian grammar authors of the post-Maretić era tried to allow wider use of the CL *ju*, for instance after verbs and words ending with *-e/-je* (cf. Mamić 1995: 187), Serbian linguists were not keen to accept such an idea.<sup>6</sup> The Serbian linguist Stevanović (1975: 306) claimed that the pronominal CL *ju* ‘her’ could only be used in combination with the verbal CL *je* ‘is’, arguing that all other uses of *ju*, those in which it does not follow the verbal CL *je*, are dialectal. He criticised the Croatian authors Brabec, Hraste & Živković who introduced the broader use of *ju* (cf. Stevanović 1975: 306). The Serbian linguists Piper & Klajn (2014: 97) explicitly state that the usage of CL *ju* (apart from instances of suppletion in the combination with the verbal CL *je* ‘is’, *nije* ‘is not’ and other *-je* ending

<sup>5</sup>It is not completely clear what the term “Eastern variant” actually means since Radanović-Kocić does not define it. From her observations on the usage of the accusative CL *ju*, we infer that she uses this term for some kind of standard Serbian since her statement definitely cannot apply to all Serbian/Eastern varieties. In Chapter 7 we provide data on the usage of the pronominal CL *ju* in Eastern Štokavian dialects.

<sup>6</sup>In 1899 Maretić published two grammar books which became highly influential (ten adapted editions for use in schools until 1928) in which he argued for the usage of the CL *ju* only in suppletion contexts.

verbs) is not correct in standard Serbian. The usage of *ju* beyond these contexts of suppletion is to be considered a foreign, i.e. Croatian, construction (cf. Piper & Klajn 2014: 97).<sup>7</sup> Accordingly, the stance of Croatian linguists Frančić & Petrović (2013: 143) is that both *ju* and *je* are acceptable in standard Croatian; however, in certain contexts they recommend only one of them. If the host ends in *-ju* (like *skrivaju* ‘are hiding’), they advise the use of *je*, whereas if it ends in *-je* (as *prije* ‘before’ and *nije* ‘is not’), they recommend the use of *ju* (cf. Frančić & Petrović 2013: 143). However, native speakers in BCS standard varieties do not always follow those recommendations and there is a great deal of diaphasic and diatopic variation within Croatian. Frančić & Petrović (2013: 143) estimate that deviations from the above-mentioned rules are very frequent in the standard Croatian language. As we already pointed out in Section 6.2, we assume that this is caused by other Croatian varieties, which are spoken by native speakers whose language differs from the standard in respect of the mentioned rule.

To sum up, the distribution of the pronominal CLs *ju* and *je* is sometimes attributed to the diatopic differences in inventory and sometimes to suppletion.

### 6.3.2 Inventory of verbal clitics in BCS standard varieties

Most authors (e.g. Težak & Babić 1996: 72, Barić et al. 1997: 72, Popović 2004: 284, Silić & Pranjković 2007: 21, Piper & Klajn 2014: 28) agree that there are three types of verbal CLs: unstressed present tense and aorist forms of the verb *biti* ‘be’ and unstressed present tense forms of the verb *ht(j)eti* ‘will’.<sup>8</sup>

#### 6.3.2.1 Present tense clitics of *biti* ‘be’

Table 6.2 contains the forms of the present tense CLs *biti* ‘be’ as presented in the grammar books of all three standards (cf. Barić et al. 1997: 271, Mrazović & Vukadinović 2009: 128). There are no diatopic differences between CL systems of BCS standard varieties.<sup>9</sup>

<sup>7</sup>For more information on suppletion see Section 2.4.2.2.

<sup>8</sup>Unlike other authors, Stevanović (1975: 350), Stanojčić & Popović (2002: 97), Mrazović & Vukadinović (2009), Jahić et al. (2000: 271) consider the forms *sam, si, je...* to be unstressed present tense forms of the verb *jesam*. The question whether the present tense forms *sam, si, je...* and *budem, budeš, bude...* are different forms of one verb *biti* or whether there are two verbs, *biti* with the present tense *budem, budeš, bude...* and *jesam* with the present tense *sam, si, je...* is not relevant to our study.

<sup>9</sup>In Table 6.2 we can see that there is diatopic difference between the BCS standard varieties in the third person singular full form of verb *biti* ‘be’. While *jěst* is proposed for standard Croatian, *jěste* is proposed in standard Bosnian and Serbian. However, it seems that the current state is a consequence of diachronic variation within standard Serbian. Piper & Klajn (2014: 198) claim

Table 6.2: Clitic and full forms of the present tense of *biti* ‘be’

		singular		plural	
	CL form	Full form		CL form	Full form
be.1SG	<i>sam</i>	<i>jèsam</i>	be.1PL	<i>smo</i>	<i>jèsmo</i>
be.2SG	<i>si</i>	<i>jèsi</i>	be.2PL	<i>ste</i>	<i>jèste</i>
be.3SG	<i>je</i>	<i>jé</i> <i>jèst, jèste</i>	be.3PL	<i>su</i>	<i>jèsu</i>

The only difference we noticed is in views on whether *je* ‘is’ can bear an accent or not. Mrazović & Vukadinović (2009: 128) claim that *je* is always unstressed, but unlike other CLs it can be used at the beginning of a sentence together with the question particle *li*.<sup>10</sup> A slightly different explanation is offered by Jahić et al. (2000: 272), who say that *je* at the beginning of a sentence is used as an accented word. In contrast, Katičić (1986: 496) and Piper & Klajn (2014: 198) argue that there are two forms: one is the CL *je*, which is unstressed, and the other, the full form *jé*, which is stressed and used in questions. In other words, they claim that in the question expression *jé li* the form *jé* bears its own accent, and can therefore be placed at the beginning of a sentence (cf. Katičić 1986: 496, Piper & Klajn 2014: 198).

### 6.3.2.2 Aoristal/conditional clitics of *biti* ‘be’

As may be seen in Table 6.3, there is no diatopic variation between the three standard varieties with respect to the CL aoristal forms of *biti* ‘be’ (cf. Barić et al. 1997: 271, Mrazović & Vukadinović 2009: 128).<sup>11</sup>

that in previous periods of Serbian language the full third person singular form of the verb *biti* was *jèst*, while today the form *jèste* is more common. It is interesting to note that Mrazović & Vukadinović (2009: 128) provide only the form *jèst* in the complete present tense paradigm, without mentioning *jèste*. However, all the examples which they provide on the same page contain only *jèste* as the full form third person singular present tense and none include the form *jèst*.

<sup>10</sup> Although they claim that the unstressed form of *je* is an exception and is the only present tense CL which can take the first position in a sentence, in the example they provide on the same page there is an accent symbol on *jé*, i.e. *Jé li ovo dobro?* ‘Is this ok?’ (cf. Mrazović & Vukadinović 2009: 128).

<sup>11</sup> The conditional is formed with the third person plural form *bi*, the full form *bīše* is only part of the aorist tense.

Table 6.3: Clitic and full forms of aorist of *biti* ‘be’

singular		plural			
CL form	Full form	CL form	Full form		
be.1AOR	<i>bih</i>	<i>bīh</i>	be.1AOR	<i>bismo</i>	<i>bīsmo</i>
be.2AOR	<i>bi</i>	<i>bī</i>	be.2AOR	<i>biste</i>	<i>bīste</i>
be.3AOR	<i>bi</i>	<i>bī</i>	be.3AOR	<i>bi</i>	<i>bīše, bīšē</i>

Again, the only difference is in the interpretation: some authors believe that aoristal forms are not stressed, even when they take 1P in the sentence, while others claim that aoristal forms can be stressed. With respect to this Piper & Klajn (2014: 28) state that the CLs *bih*, *bi*, etc. and *li* do not have stressed counterparts. In contrast, many other authors of Bosnian, Croatian, and Serbian grammar books (e.g. Katičić 1986: 498, Težak & Babić 1996: 246, Jahić et al. 2000: 272, Popović 2004: 284, Mrazović & Vukadinović 2009: 124, Ridjanović 2012: 264, 565) explicitly state that stressed counterparts of aoristal forms of the verb *biti* do exist.<sup>12</sup> Katičić (1986: 498) claims that the auxiliary forms of *biti*, with which the conditional is compounded, are stressed forms when followed by *li*. Furthermore, in one Croatian grammar the following example (1) of the stressed *bih* can be found in a contrasting conditional subordinative sentence:

- (1) Ako ti<sub>1</sub> nè bi<sub>1</sub>,      jâ<sub>2</sub> bīh<sub>2</sub>  
 if you NEG COND.3SG I COND.1SG  
 ‘If you would not, I would.’ (Cr; Barić et al. 1997: 72)

Although there is no diatopic variation between the BCS standard varieties, diaphasic variation does exist, since in the colloquial language of all three varieties the invariable form *bi* is used for all persons (cf. Francić et al. 2006: 106, Mrazović & Vukadinović 2009: 159).

### 6.3.2.3 Present tense/future clitics of *ht(j)eti* ‘will’

In comparison with the full and CL forms of the verb *biti*, the picture of *ht(j)eti* ‘will’ seems quite clear.<sup>13</sup> *Ht(j)eti* has full and CL forms in the present tense (Piper

<sup>12</sup>Barić et al. (1997: 271) list *bi* in the paradigm for stressed forms of the aorist for second and third person singular, but without any accentuation symbol.

<sup>13</sup>Only *htjeti* is used in standard Croatian and standard Bosnian, while both *htjeti* and *hteti* are allowed in standard Serbian, although the latter clearly dominates in the Serbian WaC corpus.

& Klajn 2014: 199). The CL forms are used to build the future tense. The full forms have a different meaning, i.e. ‘wish’, but they can be used in the future tense as well, for instance to form questions (Piper & Klajn 2014: 199). Table 6.4 presents CL and full forms of *ht(j)eti* (cf. Barić et al. 1997: 272, Mrazović & Vukadinović 2009: 125).

Table 6.4: Clitic and full forms of the present tense of *ht(j)eti* ‘will’

	Singular		Plural		
	CL form	Full form	CL form	Full form	
will.1SG	<i>ću</i>	<i>hòću</i>	will.1PL	<i>ćemo</i>	<i>hòćemo, hòćemo</i>
will.2SG	<i>ćeš</i>	<i>hòćeš</i>	will.2PL	<i>ćete</i>	<i>hòćete, hòćete</i>
will.3SG	<i>će</i>	<i>hòće</i>	will.3PL	<i>će</i>	<i>hòće</i>

The only diatopic variation attested in the BCS standard varieties with respect to *ht(j)eti* is in the accent in the full forms, i.e. *hòćemo* (Croatian) vs *hòćemo* (Serbian) (cf. Barić et al. 1997: 272, Mrazović & Vukadinović 2009: 125). Apart from this, there is purely orthographic variation between BCS standard varieties: while in standard Serbian the CL forms of *ht(j)eti* merge together with the *-ti* infinitive, i.e. *písaću, písaćeš*, this is not the case in standard Croatian, i.e. *písat ću, písat ćeš* ‘I will write, you will write’ (cf. Mrazović & Vukadinović 2009: 155, Barić et al. 1997: 241). There is no difference in pronunciation.

### 6.3.3 Reflexive markers *se* and *si* in BCS standard varieties

There is some diatopic difference in the inventory of reflexive CLs between BCS standard varieties. None of the Serbian authors list  $\text{REFL}_{2\text{ND}}$  *si* as a CL dative form, and Radanović-Kocić (1988: 56) explicitly says, “The dative clitic of reflexive pronoun has been completely lost in the Eastern variant, but it is still used in Western variant [...]”. While the past 60 years have seen some vacillation in Croatian grammaticography regarding the standardness of the  $\text{REFL}_{2\text{ND}}$  CL *si*, the stance of Serbian authors on the issue was rather stable: they (e.g. Stanojčić & Popović 2002: 97) either did not list the reflexive CL form *si* or they (e.g. Mrazović & Vukadinović 2009: 367) explicitly stated that there is no such form in standard Serbian.<sup>14</sup>

<sup>14</sup>In the former Yugoslavia some Croatian grammarians (e.g. Brabec et al. 1963: 96), like all Serbian grammarians, did not consider the reflexive CL form *si* to be part of the standard language

Table 6.5: Reflexive *se* in BCS and its corresponding full forms

	Genitive		Accusative		Dative	
	CL form	Full form	CL form	Full form	CL form	Full form
REFL	<i>se</i>	<i>sěbe, sèbe</i>	<i>se</i>	<i>sěbe, sèbe</i>	<i>si</i>	<i>sěbi, sèbi</i>

However, it is clear from Table 6.5 that the authors (e.g. Barić et al. 1997: 208, Mrazović & Vukadinović 2009: 367) ascribe case to the reflexive markers. While Mrazović & Vukadinović (2009: 367) say that there is only one CL *se* which is in the accusative case, Barić et al. (1997: 208) claim that there are three CL forms: in the accusative, genitive and dative. Without discussing this issue here, we can say that BCS grammarians traditionally distinguish between the reflexive pronoun *se* (REFL<sub>2ND</sub> in the terminology proposed in Section 2.5.4.2) and the reflexive particle *se* (REFL<sub>LEX</sub>). Ridjanović (2012: 558f) is the only author who explicitly claims that the REFL<sub>LEX</sub> *se* has only the CL form and no corresponding full form. As we already pointed out in Chapter 2, we do not use the term reflexive pronoun. Instead we use the term reflexive marker.

The diatopic variation in the inventory of BCS standard varieties is nicely summed up by Ridjanović (2012: 440) who claims that the REFL<sub>2ND</sub> CL *si*, which is widely used in Croatian, can hardly be found elsewhere in BCS territory.<sup>15</sup>

### 6.3.4 Polar question marker *li*

Finally, we would like to observe that there is no variation in respect of the CL particle *li* in BCS varieties. This is why it will not be part of our research focus, as we already pointed out in Chapter 2.

---

and they did not list it with other CLs. In their publications from the former Yugoslavian period, other Croatian authors (e.g. Barac-Grum et al. 1971: 364) tried to defend the standardness of the reflexive CL *si*. Barac-Grum et al. (1971: 364) claimed that the CL dative form *si* is correct and necessary in the language. The latter view prevailed in Croatian grammaticography since all grammar books published after 1990 list the CL reflexive form *si*.

<sup>15</sup>As we show in Section 7.4.3, this claim is not completely true; while it may apply to standard varieties, there is undeniable diatopic/dialectal variation – we provide data on the usage of the REFL<sub>2ND</sub> CL *si* outside Croatian territory. Moreover, the form in question is present also in the spoken Bosnian variety, for more details see Section 8.7.4.

## 6.4 Internal organisation of the clitic cluster

### 6.4.1 Clitic ordering within the cluster in BCS standard varieties

In this section we will summarise the main information on CL clusters found in literature. Several authors (e.g. Jahić et al. 2000: 471, Popović 2004: 284, 289, Piper & Klajn 2014: 451) observe that if there is more than one CL in the same simple clause, CLs will group and linearise. Piper & Klajn (2014: 451f) point out that the CL cluster usually consists of two or three elements, rarely four, while groups of five or more CLs are quite infrequent. Similarly, Ridjanović (2012: 558) claims that the maximum number of CLs in one cluster is five, commenting that such clusters are very rare. We will start our discussion with the cluster ordering presented in Section 2.4.2.1:

$li > \text{VERBAL}^* > \text{PRON}_{\text{DAT}} > \text{PRON}_{\text{ACC}} > \text{PRON}_{\text{GEN}} > \text{REFL} > je$   
 \* except *je* = PRS.3SG of *biti* ‘be’

While reviewing the grammar books, we encountered divergent information concerning the ordering of pronouns in the accusative and genitive. Note that these are homophonous forms. Some authors (e.g. Težak & Babić 1996: 246, Barić et al. 1997: 596, Jahić et al. 2000: 472, Mrazović & Vukadinović 2009: 659) propose the order dative > genitive > accusative. Barić et al. (1997: 596f) support this order with the following two examples (presented in (2) and (3)) which contain  $\text{REFL}_{\text{LEX}}$ .

- (2) On *joj*      *se*      nasmiješio.  
       he her.DAT REFL smile.PTCP.SG.M  
       ‘He smiled at her.’ (Cr; Barić et al. 1997: 596)
- (3) Djeca *su*      *ga*      *se*      nagledala.  
       children be.3PL him.GEN REFL look.at.PTCP.PL.N  
       ‘The children saw enough of him.’ (Cr; Barić et al. 1997: 597)

However, we would like to point out that this argumentation hinges solely on the doubtful interpretation that *se* is marked for accusative case. With respect to CL order in the cluster, some authors (e.g. Piper et al. 2005: 104, Ridjanović 2012: 564, Piper & Klajn 2014: 451) even claim that the dative comes first and then the accusative or genitive, which basically means that one of the two can be expected. However, the Serbian linguists Piper & Klajn (2014: 452) later explicitly state that the pronominal accusative CL stands before the genitive one, like in the permuted example (4a) in which the genitive complement *svoje pažnje* ‘their own attention’ of the verb *lišiti* ‘deprive’ has been replaced by the genitive CL *je* ‘her’.



- (4) a. Lišili                    *su ih*                    svoje pažnje.  
 deprive.PTCP.PL.M be.3PL them.ACC own attention  
 ‘They deprived them of their attention.’
- b. Lišili                    *su ih je*.  
 deprive.PTCP.PL.M be.3PL them.ACC her.GEN  
 ‘They deprived them of it.’ (Sr; Piper & Klajn 2014: 452)

In contrast two other Serbian linguists, Mrazović & Vukadinović (2009: 659), claim that there are no patterns combining accusative and genitive CLs in standard Serbian. They emphasise that the CL form of the genitive is never used together with the accusative: for example, they say that in standard Serbian the sentence presented in (5a) cannot be paraphrased as (5b) but only as (5c) (cf. Mrazović & Vukadinović 2009: 659).

- (5) a. Vlasti            *su lišile*                    narod slobode govora.  
 authorities be.3PL deprive.PTCP.PL.F folk freedom speech  
 ‘The authorities deprived the people of freedom of speech.’
- b. \* Vlasti            *su ga je*                    lišile.  
 authorities be.3PL him.ACC her.GEN deprive.PTCP.PL.F  
 Intended: ‘The authorities deprived him of it.’
- c. Vlasti            *su ga lišile*                    nje / toga /  
 authorities be.3PL him.ACC deprive.PTCP.PL.F her.GEN that  
 slobode govora.  
 freedom of.speech  
 ‘The authorities deprived him of it/that/freedom of speech.’  
 (Sr; Mrazović & Vukadinović 2009: 659)

The Bosnian linguist Ridjanović (2012: 565) shares the latter point of view that genitive and accusative CLs never occur together. An interesting discussion of this problem is offered by Milićević (2007: 104ff), who presents an account of the constitution of the CL cluster within the framework of Meaning-Text Theory. She claims that in most cases genitive and accusative CLs may come in either order. For her either of the variants presented in (6a–6d) is acceptable:

- (6) a. Lišili                    *su ih ga*.  
 deprive.PTCP.PL.M be.3PL them.ACC him.GEN  
 ‘They deprived them of it/him.’
- b. ? Lišili                    *su ih ga*.  
 deprive.PTCP.PL.M be.3PL them.GEN? him.ACC?  
 ‘They deprived him of them.’

6 *Clitics and variation in grammaticography and related work*

- c. ? Lišili                    su    ga    ih.  
 deprive.PTCP.PL.M be.3PL him.ACC them.GEN  
 ‘They deprived him of them.’
- d. ? Lišili                    su    ga    ih.  
 deprive.PTCP.PL.M be.3PL him.GEN? them.ACC?  
 ‘They deprived them of it/him.’ (Sr; Milićević 2007: 105)

She claims that this sentence can also be read either way. Due to this ambiguity, she argues that the order accusative > genitive seems to be “the default case”.

The Serbian linguist Popović (2004: 291) is the only one who notes that the presented CL order within the cluster stays the same even if the cluster consists of CLs which are governed by two different verbs, i.e. the CL order in simple and mixed clusters is the same.<sup>16</sup> Mrazović & Vukadinović (2009: 660) are the only ones who claim that permutations of CL order in the cluster are not possible and that no other element can be inserted into the CL cluster in standard Serbian.

At the end of this subsection we would like to point out that we have not observed any diatopic variation between BCS standard varieties in respect of CL order within the cluster. Any observed differences actually arise from obviously disparate interpretations, primarily with respect to the realisation of clusters with both genitive and accusative CLs.

There is no information on diaphasic variation in the grammar books. We have only come across a short article by Ondrus (1957), who claims that in the Serbian colloquial register the verbal CL *je* ‘is’ can precede pronominal CLs (cf. Ondrus 1957: 517f). He provides the following example (7) of the reversed order from colloquial Serbian:<sup>17,18</sup>

- (7) Žao joj je ga.  
 sorry her.DAT be.3SG him.ACC  
 ‘She is sorry for him.’ (Sr; Ondrus 1957: 518)

<sup>16</sup>For the difference between simple and mixed clusters see Section 2.4.2.1.

<sup>17</sup>This reversed order in CL clusters can be quickly verified in srWaC and hrWaC. In srWaC v1.2 we found 152 (0.3 per million) examples with *je*<sub>be.3SG</sub> *ga*<sub>him.ACC</sub> and 183 (0.3 per million) examples with *je*<sub>be.3SG</sub> *mu*<sub>him.DAT</sub> word order. It seems that the mentioned word order is even more frequent in Croatian, since *je*<sub>be.3SG</sub> *ga*<sub>him.ACC</sub> is attested by 681 (0.5 per million) and *je*<sub>be.3SG</sub> *mu*<sub>him.DAT</sub> by 531 (0.4 per million) examples.

<sup>18</sup>This kind of reversed CL order in which pronominal CLs are preceded by the verbal CL *je* is also attested in dialects. For more information see Section 7.5.1.

## 6.4.2 Morphological processes within the cluster

### 6.4.2.1 Suppletion

Most authors (e.g. Stevanović 1975: 306, Barić et al. 1997: 210, 597, Jahić et al. 2000: 472, Mrazović & Vukadinović 2009: 366, Ridjanović 2012: 434, Piper & Klajn 2014: 28, 97) generally agree that if the pronominal CL *je* ‘her’ precedes the verbal CL *je* ‘is’, the former will be replaced by its alternative form *ju*. However, the Bosnian linguist Ridjanović (2012: 434) claims that this dissimilated form is a feature of deliberate speech. Furthermore, he insists that in everyday colloquial language Bosnians use just one *je* instead of two CLs, meaning they prefer haplology (cf. Ridjanović 2012: 434).<sup>19</sup> Unfortunately, he does not provide any examples for this, but Piper & Klajn (2014: 97) observe a similar phenomenon among Serbian native speakers and label it as a common mistake in standard Serbian, see (8):

- (8) \*Teorija nosi ime naučnika koji je prvi  
 theory carry.3PRS name scientist which be.3SG first  
 formulisao.  
 formulate.PTCP.SG.M  
 Intended: ‘The theory carries the name of the scientist who first  
 formulated it.’ (Sr; Piper & Klajn 2014: 97)

It seems that the mentioned feature, which is not accepted in standard Bosnian and Serbian, is an actual piece of evidence for diaphasic variation, since it occurs in non-standard varieties spoken by native speakers.

Suppletion also occurs in both standard Serbian and standard Croatian if the third person feminine accusative CL stands after *nije* ‘is not’ or another verb which ends with *-je* (cf. Barić et al. 1997: 210, 597, Mrazović & Vukadinović 2009: 366, Piper & Klajn 2014: 97), as in example (9).

- (9) Ne smije ju ni vidjeti.  
 NEG may.3PRS her.ACC NEG see.INF  
 ‘He may not even see her.’ (Cr; Barić et al. 1997: 597)

It seems that in Serbian the suppletion in the last two cases mentioned is the result of a change in the norm. The rule that in standard Serbian the accusative

<sup>19</sup>It seems that Ridjanović’s observation might be correct. While analysing language material for Chapter 7 we could not find examples from local idioms (dialects) in which suppletion does take place. Moreover, our colleagues from Croatia and Serbia who specialise in dialectology could not provide us with examples of suppletion from their transcripts either. However, there are local idioms (dialects) in which the string *ju je* occurs, but not as a consequence of suppletion since the CL *ju* is the only third person singular feminine accusative CL form available.

CL *je* must be replaced with its counterpart CL *ju* after verbs which end in *-je*, or after *nije* ‘is not’ emerged during recent decades, since in the 1970s Stevanović (1975: 306) claimed that *ju* could be used only in combination with *je*. As we have already pointed out in Section 6.3, Stevanović (1975: 306) argued that all uses of *ju* which are not the result of its placement after the verbal CL *je* are dialectal.

At the end of this section we would like to point out that we did not observe any diatopic variation between the BCS standard varieties with respect to suppletion, which as a phenomenon exists in all BCS standard varieties. However, we have demonstrated that BCS standard varieties do differ as to the range of contexts in which suppletion is recommended.

#### 6.4.2.2 Haplology of unlikes

Several BCS grammar books (e.g. Težak & Babić 1996: 246, Barić et al. 1997: 596, Jahić et al. 2000: 471, Ridjanović 2012: 302, 333, Piper & Klajn 2014: 450) mention haplology of unlikes, i.e. that the verbal CL *je* ‘is’ is deleted if it would follow the reflexive CL *se*.<sup>20</sup> The Croatian linguist Katičić (1986: 497) claims that such deletion usually occurs, but it is not necessarily a general rule. He thinks that keeping the reflexive CL *se* is a feature of a pedantic and explicit style of expression (cf. Katičić 1986: 497). Similar statements can be found in descriptions of standard Serbian. Namely, Piper & Klajn (2014: 452) are even stricter when it comes to the *se je* cluster: they explicitly mark example (10a) as incorrect, but consider example (10b) to be correct in standard Serbian.

- (10) a. \* On *se je* obradovao.  
           he REFL be.3SG gladden.PTCP.SG.M  
           Intended: ‘He was gladdened.’
- b. On *se* obradovao.  
           he REFL gladden.PTCP.SG.M  
           ‘He was gladdened.’ (Sr; Piper & Klajn 2014: 452)

Regarding the omission of the verbal CL *je*, which is in contact position with the reflexive CL *se*, the Bosnian linguist Ridjanović (2012) offers a syntactic explanation. He claims that not every verbal CL *je* is omitted in the combination with the reflexive CL *se* (cf. Ridjanović 2012: 564). According to him, in standard Bosnian omission is possible as long as the verbal CL *je* is a past tense auxiliary (cf. Ridjanović 2012: 564). However, if the verbal CL *je* is a copula like in example (11), it will not be omitted (cf. Ridjanović 2012: 564).

<sup>20</sup>For more information on the haplology of unlikes see Section 2.4.2.2.

- (11) Dobro *se je* nadati.  
 good REFL be.3SG hope.INF  
 ‘It is good to hope.’ (Bs; Ridjanović 2012: 564)

While the combination of the reflexive CL *se* and auxiliary CL *je* in the simple CL cluster (10a) leads to the deletion of the auxiliary CL *je* (10b), Ridjanović’s example (11) with the CL copula *je* is a case of a mixed cluster.<sup>21</sup> It seems that whereas the auxiliary CL *je* is regularly omitted in simple CL clusters in BCS standard varieties, in standard Bosnian the CL copula *je* is preserved in mixed clusters if it co-occurs with the reflexive CL *se*.<sup>22</sup> Težak & Babić (1996: 246) add that the verbal CL *je* is also often omitted after CLs *me* ‘me’ and *te* ‘you’ in Croatian – see the example presented in (12).

- (12) Gizela *me čekala* u posječenom parku.  
 Gizela me.ACC wait.PTCP.SG.F in trimmed park  
 ‘Gizela was waiting for me in the trimmed park.’ (Cr; Barić et al. 1997: 596)

This usage should be examined in the context of the so-called truncated perfect (Serbian *krnji perfekat*). In headlines or certain contexts of spoken language the auxiliary can be omitted for all persons irrespective of the presence of other CLs. Meermann & Sonnenhauser (2016: 98f) who analysed this usage in spoken Serbian claim that it involves a distancing mechanism. The truncated perfect indicates a lack of anchoring to the point of speech and serves as a means for the speaker to distance himself or herself from what has been said. While Meermann & Sonnenhauser (2016: 99f) claim that the truncated perfect produces effects of surprise or indignation, some Croatian authors (e.g. Katičić 1986: 41, 52, 55, Barić et al. 1997: 404, 596) argue that omitting the auxiliary CLs *je* and *su* brings a stylistic value of greater brevity and expressivity.

<sup>21</sup>For more information on the simple and mixed CL cluster see Section 2.4.2.

<sup>22</sup>It is interesting to note that in the grammar books we did not find any information about co-occurrence of the auxiliary CL *je* in mixed clusters with the reflexive CL *se*. However, in bsWaC we found both examples with hapology of unlikes (i) and without it (ii).

- (i) Uspio *se* izvući.  
 Manage.PTCP.SG.M REFL get.out.INF  
 ‘He managed to get out.’ [bsWaC v1.2]
- (ii) Baš ovdje uspio *se je* pokazati i kao pjesnik.  
 exactly here manage.PTCP.SG.M REFL be.3SG show.INF and as poet  
 ‘Exactly here he managed to show himself as a poet.’ [bsWaC v1.2]

While the omission of the verbal CL *je* after the reflexive CL *se* and pronominal CLs *me* and *te* is phonologically motivated, i.e. to avoid duplication of the same vowel, according to several Croatian authors (e.g. Težak & Babić 1996: 129, Katičić 1986: 41, 52, 55, Barić et al. 1997: 404, 596), not only the verbal CL *je* but also *su* as an auxiliary can be omitted, even if there is no pronominal or reflexive CL in the sentence.

Here we would like to emphasise that we did not observe any diatopic variation between the BCS standard varieties with respect to haplology. The only observed discrepancies concern the interpretation of whether the haplology is obligatory or optional.

## 6.5 Position of the clitic or the clitic cluster

### 6.5.1 General remarks on clitic placement in BCS standard varieties

Many grammarians comment on the peculiarity of CL placement. In comparison to conjunctions and prepositions, CLs do have greater freedom of positioning (cf. Težak & Babić 1996: 246). Therefore, it can be said that the place in the sentence which CLs take is relatively free (cf. Piper & Klajn 2014: 450). However, CLs cannot be placed in any position in a sentence (cf. Težak & Babić 1996: 246). Jahić et al. (2000: 470) also emphasise that obligatory word order in Bosnian, which determines the place of clitics (proclitics and enclitics) in a sentence, is controlled only by prosodic rules. In this vein, the Croatian linguist Katičić (1986: 495) claims that the positioning of CLs is strictly and mechanically determined, which makes it stylistically neutral (Katičić 1986: 495).

Taking into consideration all the above-mentioned factors, below we present the treatment of CL placement in standard BCS varieties in detail. We discuss the following interrelated factors: breaks (or punctuation), conjunctions and complementisers, 2P, DP and phrase splitting, including the limits of the latter. Phrase splitting will receive the most attention because it is a major source of micro-variation in CL placement and it has been studied and discussed in quite some detail.

### 6.5.2 Placement with respect to breaks in BCS standard varieties

With regard to CL placement, many authors emphasise that CLs cannot follow a break, which is in line with the so-called phonological and mixed formal approaches to 2P.<sup>23</sup> A physiological break is a pause needed for normal breathing,

<sup>23</sup>For more information on phonological and mixed formal approaches to 2P see Section 2.4.3.2.

and the shortest is realised after a prosodic unit (e.g. Težak & Babić 1996: 243). Bosnian, Croatian, and Serbian linguists (cf. Težak & Babić 1996: 246, Jahić et al. 2000: 471, Stanojčić & Popović 2002: 371, Popović 2004: 283, 303, Piper et al. 2005: 105, Piper & Klajn 2014: 450) agree that CLs cannot follow a break. Therefore the sentence provided in (13) should be considered incorrect (a break is marked by |)

- (13) \* Ivica | je potrčao.  
 Ivica be.3SG run.PTCP.SG.M  
 Intended: ‘Ivica started running.’ (Cr; Težak & Babić 1996: 243)

In some cases in written language breaks are visible in orthography (comma, full stop, colon, etc.), but not always, since orthography only partially correlates with prosody. It seems that there is no diatopic variation between BCS standard varieties in respect of CL position after a break. Specifically, all scholars (e.g. Težak & Babić 1996: 246, Stanojčić & Popović 2002: 371, Popović 2004: 303, Piper et al. 2005: 105, Piper & Klajn 2014: 450) agree that CLs cannot be placed directly after physiological breaks, orthographically represented by a comma (14b) or bracket (15b), for example, or after any other kind of insertion and/or punctuation.

- (14) a. Taj profesor, poštovana koleginice, napisao je  
 that professor respected colleague write.PTCP.SG.M be.3SG  
 udžbenik za svaki predmet koji je predavao.  
 textbook for every subject which be.3SG teach.PTCP.SG.M  
 ‘That professor, respected colleague, wrote the textbook for every  
 subject he taught.’  
 b. \* Taj profesor, poštovana koleginice, je napisao [...].  
 that professor respected colleague be.3SG write.PTCP.SG.M  
 (Sr; Piper & Klajn 2014: 450)
- (15) a. Genitiv partitivni (piti vode) uzmanje je dijelova od  
 genitive partial drink.INF water taking be.3SG parts of  
 cjeline.  
 whole.  
 ‘The partial genitive (drink water) is used for parts of a whole.’  
 b. \* Genitiv partitivni (piti vode) je uzimanje [...].  
 genitive partial drink.INF water be.3SG taking  
 (Cr, Težak & Babić 1996: 246)

The Serbian linguist Popović (2004: 307) notes that people tend to place CLs after commas although it is against the norms of standard language. One of the Serbian

grammar books (e.g. Piper et al. 2005: 105) recommends the use of full instead of CL forms directly after a break, like in example (16) below.

- (16) Tobožnji snimak, uprkos svim uveravanjima, *jeste* falsifikat.  
 supposed recording despite all assurances be.3SG counterfeit  
 ‘The supposed recording, despite all assurances, is counterfeit.’  
 (Sr; Piper et al. 2005: 105)

Heavy phrases are also recognised as a factor influencing CL placement. The Croatian authors Težak & Babić (1996: 246) emphasise that CLs cannot follow longer syntagms in standard Croatian. As an illustration, they provide examples with a heavy phrase presented in (17a) and (17b).

- (17) a. “Tri dana kod sina” mala je, ali značajna pripovijetka.  
 three day at son small be.3SG but important story  
 ‘Three days with the son is a small, but important story.’  
 b. \* “Tri dana kod sina”, je mala [...].  
 three day at son be.3SG small  
 (Cr; Težak & Babić 1996: 246)

Similarly, the Serbian authors Piper & Klajn (2014: 450) point out that in standard Serbian, CLs do not follow a long initial phrase after which the break is – as they say – more expressed. In such cases, CLs follow the next stressed word, like in example (18) provided below (cf. Piper & Klajn 2014: 450).

- (18) Profesor uvoda u lingvistiku dobar je čovek.  
 professor introduction in linguistics good be.3SG man  
 ‘The Introduction to Linguistics professor is a good man.’  
 (Sr; Piper & Klajn 2014: 450)

Both Croatian and Serbian authors offer a prosodic explanation (physiological break) for this particular case of CL placement, which is supported by syntactic arguments (long initial phrase). However, it is important to emphasize that neither the Croatian nor the Serbian authors specify how to determine longer syntagms. From the treatment of DP in the grammars we can only infer that there must be language-internal variation (within one standard variety).<sup>24</sup>

---

<sup>24</sup>For more information on this see Section 6.5.4.



### 6.5.3 Placement with regard to different types of hosts in BCS standard varieties

CL positioning after conjunctions and complementisers varies. CLs cannot be placed either after the negative particle *ne* ‘not’ or after the conjunctions *a* and *i* ‘and’, but they can directly follow the conjunctions *pa* and *te* ‘so, and, then’ (cf. Barić et al. 1997: 595, Jahić et al. 2000: 471, Stanojčić & Popović 2002: 371, Popović 2004: 297, Piper & Klajn 2014: 451).<sup>25</sup> The negative coordinative conjunction *ni* ‘neither/nor’ cannot as a proclitic be a host for enclitics, whereas *niti* ‘neither/nor’ can (cf. Ridjanović 2012: 537, 562, Popović 2004: 297, Piper & Klajn 2014: 451). The coordinative conjunction *no* behaves ambiguously: when synonymous with *ali* ‘but’ it cannot host CLs, but when synonymous with *nego* and *već* ‘than’, it can (cf. Stanojčić & Popović 2002: 371, Popović 2004: 298). The bookish particle *pak* ‘however’ can only follow CLs (cf. Piper & Klajn 2014: 451), as in example (19) presented below.

- (19) *Ona je htela da im pomogne, on joj [pak]*  
 she be.3SG want.PTCP.SG.F that them.DAT help.3PRS he her.DAT but  
*to nije dozvolio.*  
 that NEG.be.3SG allow.PTCP.SG.M  
 ‘She wanted to help them, but he did not allow her to.’

(Sr; Piper & Klajn 2014: 451)

CLs can follow the coordinating conjunctions *ali* and *ili*, and according to Bosnian and Serbian literature (e.g. Jahić et al. 2000: 471, Stanojčić & Popović 2002: 371, Popović 2004: 284, 298f, Ridjanović 2012: 562, Piper & Klajn 2014: 451) they must follow all complementisers.<sup>26,27</sup> In example (20) presented below the verbal CL *je* ‘is’ directly follows the complementiser *koji* ‘which’.

<sup>25</sup>Examples with CLs placed after the conjunctions *a* and *i* can be found in *Šumadijsko-vojvođanski*, *Istočnohercegovački*, and *Srednjobosanski* dialects. For more information see Section 7.6.4.

<sup>26</sup>Croatian authors Barić et al. (1997: 595) only claim that CLs follow question and relative complementisers, but from their statement it is not clear whether these complementisers are the only complementisers which CLs follow, nor whether CLs must or can follow them.

<sup>27</sup>This may be true for standard Bosnian and standard Serbian. However, as dialectological data show for the *Šumadijsko-vojvođanski* dialect, CLs do not always follow the complementiser *da*. Moreover, our data from srWaC with CC out of *da*<sub>2</sub>-complements also show that CLs do not always follow the complementiser *da*. For more information and examples see Section 7.6.2 and Chapter 13.

- (20) Čita roman koji je napisao jedan mladi pisac.  
read.3PRS novel which be.3SG write.PTCP.SG.M one young writer  
'He is reading a novel written by a young writer.'  
(Sr; Piper & Klajn 2014: 450)

The Serbian author Popović (2004: 300f) is the only one who notes the difference between the two kinds of *jer* 'because'. He claims that CLs follow the causal complementiser *jer*, while *jer* as a connector (*nadovezivački veznik*) is not followed by CLs since a break can be felt after it (cf. Popović 2004: 300f).

We can sum this subsection up as follows: there is no diatopic variation between standard BCS varieties as regards CL placement in the case of the negative particle *ne* and the conjunctions *a*, *i*, and *ni*. They cannot host CLs. By contrast, all South-Slavonic grammarians recognise that the conjunctions *pa*, *te*, *niti*, *ali*, and *ili* can be hosts to CLs. However, since they do not state that the mentioned conjunctions must host CLs, we can expect variation within each standard variety. Furthermore, Bosnian and Serbian authors emphasise that CLs must follow all complementisers. In contrast, Croatian authors Barić et al. (1997: 595) only state that CLs follow question and relative complementisers, but they do not specify whether such complementisers are obligatory hosts to CLs. Hence, here we can expect variation within one standard variety and possibly diatopic variation between different BCS standard varieties.

## 6.5.4 Second position, second word and delayed placement

### 6.5.4.1 Second position vs second word in BCS standard varieties

While above we presented the problem of CL placement in BCS in a broader context, i.e. with respect to elements which can host CLs, this section deals specifically with the vague treatment of 2P.<sup>28</sup> Croatian and Serbian authors have different approaches to what 2P actually is. As we demonstrate in the following, Serbian linguists tend to interpret 2P as the position after the first phrase, which can but does not need to be compound, while Bosnian and Croatian authors take it to mean the position after the first word (the 2W solution).<sup>29</sup>

All the analysed grammar books of BCS standard varieties (e.g. Katičić 1986: 495, Jahić et al. 2000: 471, Popović 2004: 17, Piper & Klajn 2014: 29, 450) state that CLs attach to the preceding stressed word, and that they are consequently placed in the second position in the sentence, in principle after the first stressed

<sup>28</sup>For basic information and theoretical discussion on 2P and DP see Sections 2.4.3.1, 2.4.3.2, and 2.4.3.3.

<sup>29</sup>The term "compound phrase" refers to a phrase which consists of at least two content words.

word. However, in some grammar books (e.g Piper et al. 2005: 105) it is emphasised that the rule in question refers to the simple clause. Moreover, as we show, there is continuous discussion on the insertion of CLs after initial compound phrases, even after those which consist of only two stressed content words like *moj prijatelj* ‘my friend’ in example (21) below.<sup>30</sup>

Even in Yugoslavian times it was clear that the 2P after the first stressed word, i.e. 2W, was typical of the Western part of Serbo-Croatian language territory (cf. Pešikan 1958: 307). The Croatian linguist Babić (1964: 154f), for instance, rejected any possibility of CL insertion after a syntagm, even one containing only two stressed content words. However, it seems that not all Croatian linguists of the time agreed with Babić. Barac-Grum et al. (1971: 434) and Brabec (1964: 146f), for instance, allowed sentences in which a CL follows a two-word phrase, like in the example provided below where the verbal CL *je* ‘is’ attaches to the initial compound phrase *moj prijatelj*.

- (21) [Moj prijatelj]<sub>phrase1</sub> *je* jučer došao k nama.  
 my friend be.3SG yesterday come.PTCP.SG.M to us  
 ‘My friend came to us yesterday.’ (Cr; Barac-Grum et al. 1971: 434)

Half a century later, Croatian linguists still disagree on the question whether CLs can follow compound phrases of two stressed content words or not. While Raguž (1997: 344) allows it, Frančić et al. (2006: 182) reject such a possibility. Katičić (1986: 496f) claims that placing the CL directly after the first compound phrase bears the hallmarks of a substandard colloquial expression.<sup>31</sup> However, unlike Croatian linguists, the Serbian authors Piper & Klajn (2014: 29, 450) and Ivić et al. (2011: 161) underline that the 2P rule should not be taken literally, because it is possible to place a CL after a compound phrase in standard Serbian. Moreover, Serbian linguists provide examples with CLs which follow initial phrases, which contain more than two content words, such as in examples (22) and (23) provided below.

- (22) Manji deo takmičara *je* iz Beograda.  
 smaller part contestants be.3SG from Beograd  
 ‘The smaller part of the contestants is from Beograd.’  
 (Sr; Ivić et al. 2011: 161)

<sup>30</sup>In idioms of the Neo-Štokavian *Istočnohercegovački* dialect, which was one of the dialects that served as a base for standard Croatian, CLs can follow phrases which consist of two content words, for more information see Section 7.6.1.

<sup>31</sup>This does not mean that instead of 2P, Croatian authors prefer DP of CLs. As we have already pointed out in the previous lines, Pešikan (1958: 307) observed that placing CLs after the first stressed word was typical of the Western part of Serbo-Croatian language territory. In the next sections it becomes obvious that in standard Croatian 2W is not less preferred than DP of CLs.

- (23) Moj prijatelj s trećeg sprata je došao.  
 my friend from third floor be.3SG come.PTCP.SG.M  
 ‘My friend from the third floor came.’ (Sr; Piper & Klajn 2014: 450)

#### 6.5.4.2 Heavy initial phrases

The Serbian linguist Pešikan (1958: 308) admits that it is difficult to give concrete examples of longer initial phrases which cannot be followed by CLs in Serbian; in his opinion CL placement also depends on word length.<sup>32</sup> In a later work, Radanović-Kocić (1988) explains this in more detail. She claims that CLs usually do not directly follow an initial phrase longer than two words, but if the first long phrase is the subject, CLs can lean on it optionally (cf. Radanović-Kocić 1988: 108ff, Radanović-Kocić 1996: 435).<sup>33</sup> She supports her claims with examples (24a) and its permutation (24a-i):

- (24) a. Kolutovi plavičastog dima penjali su se [...].  
 rings blueish smoke climb.PTCP.PL.M be.3PL REFL  
 Kolutovi plavičastog dima su se penjali [...].  
 rings blueish smoke be.3PL REFL climb.PTCP.PL.M  
 ‘Rings of blueish smoke were climbing [...].’  
 (BCS; Radanović-Kocić 1988: 110)
- b. Svoje probleme i dileme lingvistika će rešavati [...].  
 own problems and dilemmas linguistics FUT.3SG solve.INF  
 ‘Linguistics will be solving its problems and dilemmas [...].’
- c. \* Svoje probleme i dileme će lingvistika rešavati [...].  
 own problems and dilemmas FUT.3SG linguistics solve.INF  
 (BCS; Radanović-Kocić 1988: 119f)

Radanović-Kocić (1996: 435) is the only author who claims that long initial objects cause DP of CLs – compare examples (24b) and (24c) provided above. Unlike her, Mrazović & Vukadinović (2009: 658) and Piper & Klajn (2014: 450) do not explicitly distinguish between long initial subjects and other kinds of phrases. According to them, in Serbian the 2P reserved for CLs is after the first stressed unit, i.e. if there is a compound phrase at the beginning of a sentence, the CL does not follow the first word, but the first phrase (Mrazović & Vukadinović

<sup>32</sup>For theoretical discussion and the definition of the term heavy initial phrase in this work see Section 2.4.3.3.

<sup>33</sup>In the examples by Ivić et al. (2011) and Piper & Klajn (2014) provided in (22) and (23) the subject is the first long phrase and therefore it can host CLs.

2009: 658, Piper & Klajn 2014: 450). Since Serbian authors differ in their opinion, we inevitably expect variation within the Serbian language.

#### 6.5.4.3 Delayed placement

In contrast to Mrazović & Vukadinović (2009) and Piper & Klajn (2014), Alexander (2009: 48) claims that CLs do not have to follow either the first word or the first phrase: their placing can be delayed. If the language user does not want to split the first phrase or splitting is not appropriate to that particular register, CL placement can be delayed (cf. Alexander 2009: 48). See the Croatian example provided below in (25).

- (25) Vaša naklapanja zaista *mi*      dosađuju.  
 your ramblings really me.DAT bore.3PRS  
 ‘Your ramblings really bore me.’ (Cr; Katičić 1986: 496)

Moreover, if the second phrase is a compound, it can be split as well, which results in DP of the CL combined with phrase splitting. See the examples from Croatian provided below in (26) and (27).

- (26) Psunj, Papuk i Krndija tvrdo *su*      eruptivno gorje.  
 Psunj Papuk and Krndija hard be.3PL eruptive mountains  
 ‘Psunj, Papuk and Krndija are hard volcanic mountains.’  
 (Cr; Barić et al. 1997: 597)

- (27) Od toga doba mnogo *je*      vode proteklo.  
 from that time much be.3SG water flow.PTCP.SG.N  
 ‘Much water has flowed (much time has elapsed) from that time.’  
 (Cr; Katičić 1986: 496)

Similarly, Bosnian authors (e.g. Jahić et al. 2000: 471) claim that if CLs do not follow the first stressed word in a sentence, they will directly follow the predicate, i.e. placement will be delayed. But, as we can see in the examples provided above, the second phrase is not always a predicate.

During past decades, Croatian authors (e.g. Weber 1859: 150–152, Jonke 1953: 150, Frančić et al. 2006: 182) considered the DP of CLs to be fully acceptable when one did not want to separate syntactically or semantically tightly bounded words, i.e. if one did not want to split compound phrases. Conversely, the Serbian linguist Pešikan (1958: 308) claimed that it is better to place CLs after a two-word phrase than to use DP of CLs. Popović (2004: 364) sees the Croatian tendency to

delay CL placement as a major factor in the growing divergence between Serbian and Croatian writers.

We can recapitulate this subsection with the following observations: Serbian grammarians differ from Bosnian and Croatian grammarians in their comprehension of the 2P; consequently we can expect diatopic variation between the BCS standard varieties. Furthermore, while Bosnian and Croatian authors recommend delaying the placement of CLs as a better alternative to placing CLs after compound phrases, as we saw Serbian authors propose quite the opposite. Therefore, we can assume that there is diatopic variation between BCS standard varieties in respect of CL placement. These tendencies are corroborated by Reinkowski's (2001) diachronic corpus study, which analysed newspaper articles from the years 1905, 1935, 1965 and 1995. She found that in both Serbian and Croatian journalistic registers the DP is dominant (cf. Reinkowski 2001: 183, 202).<sup>34</sup> Furthermore, she established that phrase splitting had also been present in the Croatian journalistic register for over the previous 100 years, although it reached its lowest point in 1965 (cf. Reinkowski 2001: 191–195). In contrast, phrase splitting has been slowly disappearing from the Serbian journalistic register over the analysed period of 90 years (Reinkowski 2001: 191–195).

### 6.5.5 The limits of phrase splitting in BCS standard varieties

As phrase splitting has attracted the attention of both normativists and formal linguists, we would like to give an account of the data discussed in the literature. Reinkowski (2001: 81) claims that Meillet & Veillant (1924: 289) were the first to notice and mention the fine diatopic variation between Croatian and Serbian: in their own words, the language of Belgrade prefers no splitting of the subject phrases. Likewise Alexander (2008: 11) emphasises that Croatian and Serbian differ as to phrase splitting and that this difference was already noticeable well before the break up of Serbo-Croatian. Even those Croatian linguists who were not determined apologists of phrase splitting and 2W CL placement like Brabec (1964: 145) admit that phrase splitting has always been common in Croatian texts from all periods and all regions. The Serbian linguist Pešikan (1958: 309) finds the Croatian tendency to insert CLs after the first stressed word and to split phrases is an exaggeration. Radanović-Kocić (1988: 111) claims that there is an important difference between initial two-word subject phrases and non-subject phrases: in

---

<sup>34</sup>In following years, in her own independent study, Alexander (2008: 14) proved and verified Reinkowski's results.

her dialect only subject phrases can be split, whereas others cannot.<sup>35</sup> Furthermore, she believes that the placement of CLs after the first word of a two-word initial subject or after the whole phrase depends on the structure of the subject phrase (cf. Radanović-Kocić 1988: 112). However, Alexander (2009: 52–55) notes that not only is phrase splitting more frequent in Croatian than in Bosnian and Serbian, it is also found in more contexts, i.e. in more types of phrase structures in Croatian. Frančić et al. (2006: 182) give precise examples of phrases which can be split by CLs in Croatian: CLs can split an adjective, numeral, pronoun or noun from a noun, and a forename from a family name. Regarding Bosnian, Čedić (2001: 196f) admits that phrase splitting does occur, but he sees it rather as a Croatian import than a Bosnian feature.

As will become evident in this subsection, Franks & Peti-Stantić (2006: 4) correctly noticed that “there is a high degree of variation in judgments about the acceptability of different kinds of splitting, both across speakers and across languages”. Therefore, in the following we compare cases in which phrase splitting is possible in all three standard varieties. Before further elaborating on this phenomenon we must point out that the Serbian linguist Popović (2004: 306) is the only one to notice that CLs can be inserted only after the first word in a compound phrase.<sup>36</sup> In both Croatian and Serbian, CLs can split an adjective and a noun: compare examples (28)–(30) (cf. Težak & Babić 1996: 246, Piper & Klajn 2014: 450).

- (28) Motovunske *su* ulice vrvjele pukom.  
 Motovunian be.3PL streets buzz.PTCP.PL.F people  
 ‘Motovunian streets were buzzing with people.’

(Cr; Težak & Babić 1996: 246)

- (29) Dobra *se* roba brzo proda.  
 good REFL wares quickly sell.3PRS  
 ‘Good wares sell quickly.’

(Sr; Piper & Klajn 2014: 450)

<sup>35</sup>She does not explain what “my dialect” actually means. It could mean the Eastern variant of Serbo-Croatian with all its varieties, only the standard Serbian variety or the *Istočnohercegovački* dialect, since she was born where this dialect is spoken. The problem is that her thesis is called *The grammar of Serbo-Croatian clitics*, so on the one hand she examines Serbo-Croatian as one abstract system, while on the other hand she admits that variation exists, but ultimately she uses her own language feeling and her own dialect as a baseline of comparison when she claims that something is ungrammatical.

<sup>36</sup>As our data from spoken Bosnian indicate, this is not completely true. For examples of phrase splitting in which CLs are not inserted after the first stressed word in a phrase see Section 8.9.5.2.

- (30) Anina *mi ga je* sestra poklonila.  
 Ana's me.DAT it.ACC be.3SG sister gift.PTCP.SG.F  
 'Ana's sister gave it to me as a present.' (BCS; Progovac 1996: 419)

Radanović-Kocić (1988: 112) claims that two-word initial subject phrases with an attribute-noun structure represent the only real case of microvariation, because CLs can follow the first word or first phrase in both Croatian and Serbian variants. Although as a rule the majority of formal linguists tend to discount both syntactic microvariation and sociolinguistic variation, Radanović-Kocić (1988: 135) assumes that in the case of the adjective attribute and noun, phrase splitting is more frequent in Croatian. She argues that in her dialect CLs can be placed after the adjective only if the adjective carries the phrasal stress, while in other dialects this condition is not necessary (cf. Radanović-Kocić 1988: 134). However, in a later paper (Radanović-Kocić 1996: 435) she states that such examples are marginal and that more complex CL clusters are not allowed in that position.<sup>37</sup> She provides the example presented in (31a) and its permutation (31b).<sup>38</sup>

- (31) a. ? *Moj ga se* brat sjeća.  
 my him.GEN REFL brother remember.3PRS  
 'My brother remembers him.'
- b. \* *Moj ti ga se* brat sjeća.  
 my you.DAT him.GEN REFL brother remember.3PRS  
 Intended: 'My brother remembers him, you know.'
- (BCS; Radanović-Kocić 1996: 435)

Radanović-Kocić (1988: 114) claims that unlike subjects with an adjective-noun structure, subjects with other structures are rarely split by CLs. Furthermore, from the perspective of her dialect she evaluates Katičić's (1986) sentences with a non-subject initial split phrase presented in (32) as grammatically questionable.

- (32) ? *Takvoj se* definiciji može staviti prigovor.  
 such REFL definition can.3PRS put.INF complaint  
 'Such a definition is subject to complaint.'
- (Cr; Radanović-Kocić 1988: 111)

<sup>37</sup>As we show later in this section, Progovac (1996) also comments that sentences become worse if CL clusters, and not single CLs, split phrases. However, dialectal and spoken data speak against those claims made by theoretical syntacticians – see Sections 7.6.3 and 8.9.5.1.

<sup>38</sup>This dative CL *ti* 'you' in this example is called the ethical dative (not an argument) and is not easily translatable into English. It is used in spoken language, in directed speech and signals closeness (Silić & Pranjković 2007: 220).



From the discussion presented above it seems that insertion of a CL between an adjective and a noun is less restricted in standard Croatian than in standard Serbian. As we can see from the literature, other kinds of attributes can be split from their head noun by a CL as well. We find examples of a CL splitting a demonstrative pronoun and a noun in both standard Croatian (33) and standard Serbian (34) (cf. Barić et al. 1997: 597, Piper & Klajn 2014: 450).<sup>39</sup>

- (33) Taj *će* *se* režim prije ili kasnije naći pod ruševinama  
 that FUT.3SG REFL regime sooner or later find.INF under ruins  
 svoje nasilne politike.  
 own violent politics  
 ‘This regime will sooner or later find itself under the ruins of its own  
 violent politics.’ (Cr; Barić et al. 1997: 597)
- (34) Taj *nam* predlog ne odgovara.  
 that us.DAT suggestion NEG answer.3PRS  
 ‘That suggestion does not suit us.’ (Sr; Piper & Klajn 2014: 450)

In the previous century the Serbian linguist Pešikan (1958: 306f) claimed that in Serbian CLs cannot separate a noun from its pronoun attribute. As we saw above, contrary to him, currently the Serbian authors Piper & Klajn (2014: 450) allow such a possibility in standard Serbian.

Furthermore, CLs can split adverbial phrases in both Croatian and Serbian standard language – see the example provided below in (35).

- (35) Vrlo *su* hrabro to uradili.  
 very be.3PL bravely that do.PTCP.PL.M  
 ‘They did that very bravely.’ (Sr; Piper & Klajn 2014: 450)

Piper & Klajn (2014: 450) state that CLs can directly follow the modifiers *samo* and *jedino* ‘only’. It is important to emphasise that in example (36) provided by Piper & Klajn (2014: 450), CLs are inserted into a prepositional phrase just like in the Croatian example (37). Unlike Piper & Klajn (2014), Radanović-Kocić (1988: 114, 1996: 436) believes that there are very few cases in which a CL can split a head noun and its modifier. She adds that examples in which CLs are placed between a noun and its modifier in a PP are ungrammatical in her dialect (cf. Radanović-Kocić 1996: 436).<sup>40</sup>

<sup>39</sup>We found this kind of split phrase in dialectological data. For more information see Section 7.6.3.

<sup>40</sup>As we already emphasised, she does not really explain what “my dialect” means. Since it could mean the *Istočnohercegovački* dialect, we would like to point out that data from dialectological literature show that these kinds of splitting are possible in the *Istočnohercegovački* dialect; for more information see Section 7.6.3.

- (36) Samo / Jedino *se* iz Niša nije niko javio.  
 only only REFL from Niš NEG.be.3SG no.one reply.PTCP.SG.M  
 ‘Only from Niš no one replied.’ (Sr; Piper & Klajn 2014: 450)
- (37) Neki *su* od njih sada čučali na pragu, drugi *se*  
 some be.3PL from them now crouch.PTCP.PL.M on doorstep others REFL  
 razvalili u hladovini.  
 lay.down.PTCP.PL.M in shade  
 ‘Some of them were now crouching on the doorstep, others were lying in  
 the shade.’ (Cr; Barić et al. 1997: 597)

Ridjanović (2012: 560) claims that phrase splitting in Bosnian has its limits: namely, a NP which consists of a noun and an adverbial, nominal complement or modifier cannot be split by a CL. Therefore in such cases the CL will follow that phrase (Ridjanović 2012: 560) – see example (38) provided below.

- (38) Avion u letu *je* slikalo nekoliko turista.  
 airplane in flight be.3SG photograph.PTCP.SG.N several tourists  
 ‘The airplane in flight was photographed by several tourists.’  
 (Bs; Ridjanović 2012: 560)

Only for standard Croatian did we find information that a CL can split a noun from its postmodifying genitive attribute – see example (39) provided below. In contrast, according to Pešikan (1958: 306f) in standard Serbian CLs cannot separate a noun and its postmodifying genitive attribute – see (40).

- (39) Kontrast *je* ovih fakata očigledan.  
 contrast be.3SG these facts obvious  
 ‘The contrast between these facts is obvious.’ (Cr; Barić et al. 1997: 597)
- (40) \*Čovek *će* visoke inteligencije uvek uspeti.  
 man FUT.3SG high intelligence always succeed.INF  
 Intended: ‘A man of high intelligence will always succeed.’  
 (Sr; Pešikan 1958: 306f)

The problem of phrase splitting in the context of a noun and its modifier in case has been thoroughly discussed in the theoretical literature. Franks & Progovac (1994: 70) and Mišeska Tomić (1996: 522) have stated that examples with CLs inserted between a noun and its modifier in the genitive are incorrect – compare the examples provided below in (41a), (41b) and (42). Two years later, however, Progovac (1996: 419) admitted that the example presented in (41b) is possible, but the phenomenon is extremely marginal.<sup>41</sup>

<sup>41</sup>The very same example was deemed unacceptable in Franks & Progovac (1994).

- (41) a. Prijatelji moje sestre *su* upravo stigli.  
 friends my sister be.3PL just arrive.PTCP.PL.M  
 'My sister's friends just arrived.'
- b. \*Prijatelji *su* moje sestre upravo stigli.  
 friends be.3PL my sister just arrive.PTCP.PL.M  
 (BCS; Mišeska Tomić 1996: 522)

Progovac (1996: 418) rejects examples in which more than one CL splits a noun and its modifier in a case as ungrammatical. Furthermore, she claims that the examples become worse when more than one CL is inserted (cf. Progovac 1996: 419) – compare her examples (42) and (43) below. However, she observes that conversely, an insertion of more CLs in a possessive phrase makes no difference, like in her example (30) above.

- (42) \*Roditelji *su se* uspešnih studenata razišli.  
 parents be.3PL REFL successful students separate.PTCP.PL.M  
 Intended: 'Parents of successful students separated.'  
 (BCS; Progovac 1996: 418)
- (43) ?\*Prijatelji *su mi ga* moje sestre poklonili.  
 friends be.3PL me.DAT it.ACC my sister gift.PTCP.PL.M  
 Intended: 'My sister's friends gifted it to me.' (BCS; Progovac 1996: 419)

Similarly but going into less detail, Radanović-Kocić (1988: 114, 1996: 436) believes that there are very few cases in which a CL splits a head noun and its modifier in the genitive. She even emphasises that in most cases, in her dialect CLs have to follow such phrases and adds that sentences in which CLs are placed between a noun and its modifier in the genitive are ungrammatical in her dialect (cf. Radanović-Kocić 1996: 436). However, Alexander (2009: 54) uses examples from *Hrvatska gramatika* (Barić et al. 1997) to argue that such restrictions do not apply to standard Croatian, i.e. CLs can be inserted between a head noun and its modifier in the genitive, as we already demonstrated in example (39) provided above.

Only for standard Croatian do we find information that CLs can split an apposition from a noun – see the example provided below.

- (44) Gospoja *ih se* Olivija naprosto plašila.  
 madam them.GEN REFL Olivia simply afraid.PTCP.SG.F  
 'Madam Olivia was simply afraid of them.' (Cr; Barić et al. 1997: 597)

6 *Clitics and variation in grammaticography and related work*

A CL can separate parts of indefinite pronouns and adverbs (cf. Katičić 1986: 496, Barić et al. 1997: 207, Popović 2004: 295f). However, it seems that in both Croatian and Serbian the version without splitting presented in (45a) is more frequent than the version with splitting presented in (45b) (cf. Katičić 1986: 496, Barić et al. 1997: 207, Popović 2004: 295f, 323).

- (45) a. Tko god *je* vidio njegove slike, bio *je*  
 who ever be.3SG see.PTCP.SG.M his paintings be.PTCP.SG.M be.3SG  
 zadržan.  
 impressed  
 ‘Whoever saw his paintings was impressed.’
- b. Tko *je* god vidio njegove slike [...].  
 who be.3SG ever see.PTCP.SG.M his paintings  
 ‘Whoever saw his paintings [...].’ (Cr; Barić et al. 1997: 207)

In both Bosnian and Croatian scholarly literature and textbooks (e.g. Babić 1963: 64, Katičić 1986: 496, Barić et al. 1997: 598, Jahić et al. 2000: 471, Frančić et al. 2006: 182, Frančić & Petrović 2013: 195) it is claimed that even forenames can be separated from family names by CLs – see example (46) below. Furthermore, it is claimed that such positioning, which strictly follows the rule of CL placing, is a hallmark of stylistically polished expression (cf. Katičić 1986: 496).

- (46) Luka *bi* Šušmek polazio u šetnju da namigne kojoj  
 Luka COND.3SG Šušmek depart.PTCP.SG.M in walk that wink.3PRS which  
 curi.  
 girl  
 ‘Luka Šušmek would go on walks to wink at some girl.’  
 (Cr; Barić et al. 1997: 598)

Unlike Bosnian and Croatian scholars, Serbian linguists (e.g. Popović 2004: 319) do not allow splitting of forenames and family names in contemporary Serbian, although they admit that such occurrences were possible in earlier periods of Serbian.<sup>42</sup> Pešikan (1958: 306) provides the following example presented in (47) which was previously acceptable in Serbian.

- (47) Matija *je* Benadić čovek sasvim star.  
 Matija be.3SG Benadić man very old  
 ‘Matija Benadić is a very old man.’ (Sr; Pešikan 1958: 306)

<sup>42</sup>Not only was this kind of phrase splitting possible in earlier periods of Serbian, but it is also now present in dialects spoken on the Serbian territory, see Section 7.6.3.

Čavar & Wilder (1994: 37) believe that cases in which a verbal CL splits a fore-name and a family name are marginal for most speakers. Bošković (2001: 3) claims that splitting a first and last name by a CL is an eccentricity generally possible in South Slavic. Following Franks (1997: 116), Bošković (2001: 16f, 29) argues that a CL can split the first name and the last name only when both names are inflected for structural case. According to Franks (1997: 116), splitting of proper names can only occur when both first and last name are treated as separate heads.<sup>43</sup> A structurally similar case is compound geographical names and terms, which according to Pešikan (1958: 306f) and Radanović-Kocić (1988: 116) cannot be split either.

The Serbian linguists Pešikan (1958: 307), Radanović-Kocić (1988: 116), and Mišeska Tomić (1996: 523) observe that conjoined NPs in general are never split by CLs: compare the examples presented in (48) and (49b) with the example in (49a).

- (48) \* Petar će i Marko doći.  
 Petar FUT.3SG and Marko come.INF  
 Intended: 'Petar and Marko will come.' (Sr; Pešikan 1958: 307)

<sup>43</sup>We completely agree that *Lav* in example (i) is not a head, which is probably one of the elements contributing to the unacceptability of the example in question.

- (i) \* Lav sam Tolstoja čitala.  
 Leo.NOM be.1SG Tolstoj.ACC read.PTCP.SG.F  
 Intended: 'I read Leo Tolstoj'. (BCS, Bošković 2001: 17)

Note, however, that Franks (1997: 116) and Bošković (2001: 16) admit that declining only one part is, in fact, marginally possible in BCS and that this marginality is independent of splitting. Moreover, we would like to point out that the "case test" cannot be applied in the straightforward manner assumed by Bošković (2001) and Franks (1997), since examples of phrase splitting like (ii) in which seemingly only one part of the proper name phrase is inflected can be attested in corpora.

- (ii) S Jadrankom je Kosor godinama radio i  
 with Jadranka.INS be.3SG Kosor.NOM years.INS work.PTCP.SG.M and  
 suradivao [...].  
 cooperate.PTCP.SG.M  
 'He worked and cooperated with Jadranka Kosor for years.' [hrWaC v2.2]

Moreover, we believe that in future, theoretical assumptions on the range and limits of proper name splitting should be verified against robust empirical data. For instance, Franks (1997: 116) considers that examples with splitting of proper names in which both parts are in the nominative case are not completely acceptable, although as we show in this section, such examples appear in the normative and descriptive BCS literature: see examples in (46) and (47).

6 *Clitics and variation in grammaticography and related work*

- (49) a. O Veri i Jani si mi govorio.  
 about Vera and Jana be.2SG me.DAT speak.PTCP.SG.M  
 ‘You told me about Vera and Jana.’  
 b. \*O Veri si mi i Jani govorio.  
 about Vera be.2SG me and Jana speak.PTCP.SG.M  
 (BCS; Mišeska Tomić 1996: 523)

However, it seems that not all facts are this clear cut, even in the case of Serbian. Progovac (1996: 418f), for instance, claims that examples of conjoined NPs with one inserted CL (as in (50)) are marginal, and those with an inserted CL cluster (as in (51)) are outright ungrammatical in her Serbian.

- (50) ?? Sestra će i njen muž doći u utorak.  
 sister FUT.3SG and her husband come.INF in Tuesday  
 Intended: ‘My sister and her husband will come on Tuesday.’  
 (BCS; Progovac 1996: 418)  
 (51) \* Sestra će mi ga i njen muž pokloniti.  
 sister FUT.3SG me.DAT it.ACC and her husband gift.INF  
 Intended: ‘My brother remembers him.’ (BCS; Progovac 1996: 418f)

In her opinion, the examples become worse when more than one CL is inserted (cf. Progovac 1996: 419). However, in contrast to all the theoretical linguists who reject phrase splitting in the case of conjoined NPs (e.g. Radanović-Kocić 1988: 116f, Schütze 1994: 66ff, Progovac 1996: 418f;), Franks & Peti-Stantić 2006: 5, 11 claim that splitting of conjoined NPs is perfectly fine for many native speakers of Croatian. Popović (2004: 320) claims that inserting CLs into conjoined phrases is very rare in contemporary Serbian, but he does not call it ungrammatical.<sup>44</sup>

According to Ridjanović (2012: 458) the most frequent cases of phrase splitting in Bosnian are those with interrogative pronouns and their adjective postmodifiers: see the example in provided in (52).

- (52) Čemu se dobrom možemo nadati?  
 what REFL good can.1PRS hope.INF  
 ‘What good can we hope for?’ (Bs; Ridjanović 2012: 458)

If there is a relative or question pronoun (question word) in a Bosnian sentence, CLs can be placed directly after it, but this is not obligatory (Ridjanović 2012: 563).

<sup>44</sup>The mentioned structure, controversial from the theoretical point of view, is according to dialectological data widespread in the *Istočnohercegovački* dialect and has been attested on Serbian territory, for more information see Section 7.6.3.

This kind of phrase splitting is not considered controversial in Serbian literature (e.g. Pešikan 1958: 307, Popović 2004: 294), moreover it is claimed to be quite common. Compare examples (53a), (53b) and (54) provided below.

- (53) a. *Koji ste grad posjetili?*  
 which be.2PL city visit.PTCP.PL.M  
 b. *Koji grad ste posjetili?*  
 which be.2PL city visit.PTCP.PL.M  
 ‘Which city did you visit?’ (Bs; Ridjanović 2012: 563)
- (54) *Koja ga je žena pozdravila?*  
 which him.ACC be.3SG woman greet.PTCP.SG.F  
 ‘Which woman greeted him?’ (Sr; Pešikan 1958: 307)

As we have shown above, not all phrase splitting possibilities are mentioned in grammar books of all standard varieties. We can infer that there is some microvariation with respect to phrase splitting. Moreover, Piper et al. (2005: 105) also claim that phrase splitting in Serbian is possible, but they clearly favour the examples presented in (55a) and (55b) and judge them to be far better than the phrase splitting version presented in (55c).

- (55) a. *Novi igrači su bili uspješniji.*  
 new players be.3PL be.PTCP.PL.M more.successful  
 b. *Novi igrači bili su uspješniji.*  
 new players be.PTCP.PL.M be.3PL more.successful  
 c. *Novi su igrači bili uspješniji.*  
 new be.3PL players be.PTCP.PL.M more.successful  
 ‘New players were more successful.’ (Sr; Piper et al. 2005: 105)

In contrast, Stanojčić & Popović (2002: 371) also mention the possibility of inserting CLs into a syntagm, but they do not specify in what cases it is possible and they do not state whether splitting or not splitting is better in Serbian.

Before we conclude this section, we would like to point out one more interesting fact. In most cases phrases are split by verbal CLs (compare examples presented in this section). This is also noted by Peti-Stantić (2005) for Croatian. She observes that verbal CLs very often split phrases in standard Croatian, which is not the case for pronominal ones (cf. Peti-Stantić 2005: 174f).<sup>45</sup>

<sup>45</sup>In this respect dialects and spoken data do not differ much from standard BCS varieties. For more information, see Sections 7.6.3. However, Peti-Stantić (2005) uses absolute and not relative values. For more discussion see 8.9.5.

We can conclude this section on splitting by referring to Alexander (2009: 50), who notices that there is still great need to investigate to what extent 2P after long phrases, and different types of phrase splitting are acceptable.

## 6.6 Summary

### 6.6.1 Clitic inventory

This part can be summed up as follows. Undoubtedly, from the descriptions of Serbian and Croatian linguists it can be seen that there is one important difference in the CL inventory of BCS standard varieties, since only Croatian grammarians accept the standardness of the reflexive CL *si*. The only scholar who clearly spells out this difference is the Bosnian author Ridjanović (2012: 440).

Moreover, Croatian and Serbian authors differ in their recommendations for the usage of the third person singular feminine accusative CL *ju* and *je*. According to some Croatian authors, *ju* can be treated as a separate unit of the inventory (and not only as the result of suppletion, for which see below).

### 6.6.2 Clitic cluster and morphonological processes within it

We would like to reiterate the following facts from this part. First of all, the linearisation of pronominal CLs presented in Bosnian, Croatian, and Serbian grammar books differs from the one shown in Franks & King (2000: 29) since in the former the authors claim that genitive precedes accusative. Further, there is some disagreement among Serbian authors regarding the realization of the hypothetically possible combination of genitive and accusative pronominal CLs within the CL cluster. While Piper & Klajn (2014: 451) provide an example of this, Mrazović & Vukadinović (2009: 659) strongly refuse such a possibility. It might be relevant to point out that the CLs in question are homophones.

Regarding morphonological processes, it is not very clear if haplology of unlikes is obligatory or not in standard Croatian. The assertions that the auxiliary CL *je* can be deleted and that it is deleted after the reflexive CL *se* are found in Težak & Babić (1996: 246) and Barić et al. (1997: 596). Unlike them, Katičić (1986: 497) does not consider haplology to be the rule. In opposition to Croatian authors, Piper & Klajn (2014: 452) are explicit in considering the sequence *se je* to be incorrect in standard Serbian. However, Ridjanović (2012) observes that haplology does not always occur in standard Bosnian: the exception to haplology is cases in which the verbal CL *je* has the function of a copula.



The third person feminine accusative pronominal CL *ju* can be used in standard Serbian only in the case of suppletion: direct contact with the verbal CL *je*, verbs ending in *-je* and *nije* (cf. Piper & Klajn 2014: 97). The usage of the CL *ju* in standard Croatian is not restricted to contexts of suppletion.

### 6.6.3 Position of clitics or clitic cluster: second position

This part can be summed up as follows. BCS authors emphasise that CLs cannot follow breathing breaks, i.e. they cannot follow punctuation symbols, brackets, inserted parts of a sentence, inserted sentences and listing. Težak & Babić (1996: 246) also underline that CLs cannot follow a longer syntagm. Using full forms instead of the CL ones directly after a breathing break is recommended (cf. Piper et al. 2005: 105). CLs can follow the coordinative conjunctions *pa*, *te*, *niti*, *ali*, and *ili* and they can never directly follow *a*, *i*, and *ni*. In all three standard varieties CLs are posterior to the subordinating conjunctions. Serbian authors (e.g. Piper & Klajn 2014: 450, Stanojčić & Popović 2002: 371) claim that the right-most position of CL pronouns and reflexives is after their governing verb. The most interesting facts we found in Croatian and Serbian grammar books are those which concern different types of variation. Silić & Pranjković (2007: 374) ascribe the differences in CL placement to the spoken and written language register, i.e. diamesic variation. Similarly, Piper & Klajn (2014: 452) consider that the variation in the placement of CLs depends on the type of CL, the sentence structure and the functional register in use, i.e. diaphasic variation.

### 6.6.4 Second position, delayed placement and phrase splitting

We would like to highlight the following facts in this part. While it seems that in Croatian and Bosnian the second position rule is understood as 2W, in Serbian literature it is emphasised that 2P is normally understood as the position posterior to the first phrase. However, even some Serbian authors acknowledge that it is possible to split a phrase by CL insertion, but this is less preferred. Piper & Klajn (2014: 450) specify the conditions under which CLs can be inserted into the first phrase in standard Serbian.

In contrast to Serbian, in which phrase splitting is uncontroversial only in cases of adjective attributes, adverbs, and the words *samo* and *jedino* (cf. Piper & Klajn 2014: 450), Croatian and Bosnian standards allow the insertion of CLs in far more contexts. For instance, CLs can be inserted between a head noun and its noun attribute, even if the latter is a PP, between an apposition and a noun, between parts of an indefinite compound pronoun and an adverb, and between a

## *6 Clitics and variation in grammaticography and related work*

question pronoun and a noun. Only in Bosnian and Croatian grammar books is it stated that CLs can split a forename from a family name. The data show some disagreement among scholars as to whether a phrase can be split by more than one CL.

We would like to conclude this chapter by pointing out that in the works analysed above, CC, diaclisis and pseudodiaclisis have almost completely escaped the attention of the normativist authors, i.e. they were touched upon in only very few cases and superficially. This might be explained by the fact that these concepts are not established in traditional grammaticography.

# 7 Clitics in dialects (Bosnian, Croatian, Serbian)

## 7.1 Introduction

Research on BCS clitics in the theoretical literature is mainly based on standard varieties. To the best of our knowledge there are no earlier studies devoted to CLs in BCS dialects. Data from dialects are included quite rarely, mainly in descriptions of phenomena which do not occur in standard languages like CL doubling (see Section 7.9). Moreover, in dialectological studies CLs are only sporadically mentioned in the sections dedicated to morphology and syntax. Thus, this study is most probably the first attempt to give an overview of the CL system in BCS dialects.

In Chapters 6, 8, and 14 we show that even the data from standard varieties display a certain degree of variation with respect to CLs. We strongly believe that this variation has its source in dialects. Namely, standard varieties emerge via a complex process of selection and normativisation of specific dialect(s) which are chosen as a basis for the standard. Standard varieties are therefore inseparable from local idioms. Additionally, standard varieties are learned at school. Moreover, as we already demonstrated in Chapter 6, some native speakers do not completely acquire certain rules in respect of CLs in the standard variety. Looking into the data from dialects can give us a clearer picture of the detected variation and can help understand why native speakers make what normativists consider “mistakes” when they use CLs in standard varieties.

Due to the mentioned lack of previous studies focusing on CLs in BCS dialects, we start with the first step of the research strategy presented in Section 3.3.1. Chosen strategy: intuition/theory. We summarise and critically synthesise explicit findings from dialectological literature. Since CLs behave completely differently in Kajkavian and Čakavian, and human and other resources are limited, we concentrate only on the data from Štokavian dialects. This dialect is our focus because it is more widespread than Čakavian and Kajkavian. Moreover, some Štokavian dialects serve as the base for the three standard varieties of BCS. However, since in Croatia Kajkavian and Čakavian dialects are in contact with Štokavian,

sometimes we could not neglect data from Kajkavian and Čakavian. Furthermore, data from Kajkavian and Čakavian are sometimes used to show parallels or divergences in the CL systems.

The second step of our empirical approach, i.e. observation, was applied only partially. First, as explained in Section 7.3, only some dialectological works contain transcripts and only some transcripts are valuable sources of data. Second, the lack of transcripts in digital form slowed us down. Since no quantitative analysis was possible, the transcripts of Štokavian dialects were analysed only qualitatively and the exact data on the distribution of certain structures are lacking. In our qualitative analysis we focused on interesting examples with phenomena detected as parameters of variation in Section 2.4.

Since we assume that not all readers are familiar with basic dialectological concepts, Section 7.2 presents a compact introduction to BCS Štokavian dialects which is followed by a short overview of available data in Section 7.3. The following sections bring a comprehensive account of parameters of CL variation. Section 7.4 gives exhaustive data on variation in the CL inventory, while Section 7.5 introduces interesting findings concerning the internal organisation of the CL cluster, which includes divergent patterns in CL cluster formations and morphological processes within the cluster. The position of the CL or CL cluster (1P, 2P, DP, phrase splitting, endoclitics) is discussed in Section 7.6. Sections 7.7 and 7.8 present data on CC and diaclisis in dialects. We attempt to describe clitic doubling in Štokavian dialects in Section 7.9. For the sake of comparison, the status of each parameter in standard varieties is thoroughly described in Chapter 6. An empirical study of parameters of microvariation carried out on material from spoken Bosnian is presented in the next chapter.

## 7.2 An overview of BCS Štokavian dialects

Three main groups of Slavonic dialects are spoken on the territory of Bosnia and Herzegovina, Croatia, Serbia, Kosovo and Montenegro: Štokavian, Čakavian and Kajkavian. The latter two are used only in Croatia, while Štokavian is used in all the abovementioned countries.

An overview of Štokavian dialects based on three classifications is presented in Table 7.1. The reader has to bear several things in mind. First of all classifications differ as to the number of Štokavian dialects. For instance Ivić et al. (2001) and Lisac (2003: 160f), who use modified versions of Brozović & Ivić (1988) map, list the *Smederevsko-vršački* dialect among Štokavian dialects, unlike Okuka (2008: 318f). Moreover, Lisac (2003: 160f) and Okuka (2008: 318f) include *Istočnobosanski*,

i.e. *Srednjobosanski*, whereas according to the map by Ivić et al. (2001) such a dialect does not exist.

Besides the differences in the number of Štokavian dialects, there are certain differences in terminology. Very often one and the same dialect is differently labelled by different authors: compare for instance the names for the *Zapadni* dialect in Table 7.1 (page 130). Throughout this chapter we will use the terms which are in small caps format in the table.

In this book we mainly use the terminology proposed by Okuka (2008), with three exceptions: for the sake of brevity, instead of *Istočnohercegovačko-krajiški* we use *Istočnohercegovački*. The term *Zapadnohercegovačko-primorski* is also replaced by the shorter term *Zapadni* dialect. And instead of using *Zetsko-raški*, whose second element refers to the medieval name of the region, we will use *Zetsko-južnosandžački* because we believe that the latter term is more transparent for those who are not that familiar with the medieval history of the region now called Sandžak. While many dialects have names corresponding to the regions in which they are used, the reader has to bear in mind that the names of dialects do not always completely overlap with the names of the regions. For instance, *Istočnohercegovački* is not used only in Eastern Herzegovina but also in Western Bosnia, North-Eastern Montenegro, Western Serbia and Eastern Croatia. Moreover, in Eastern Herzegovina not only is *Istočnohercegovački* spoken, but also *Srednjobosanski*. For the sake of clarity we would like to state that we use region names only to refer to regions, and dialect names only to refer to dialects.

The spatial distribution of Štokavian dialects from Table 7.1 is shown in Figure 7.1.

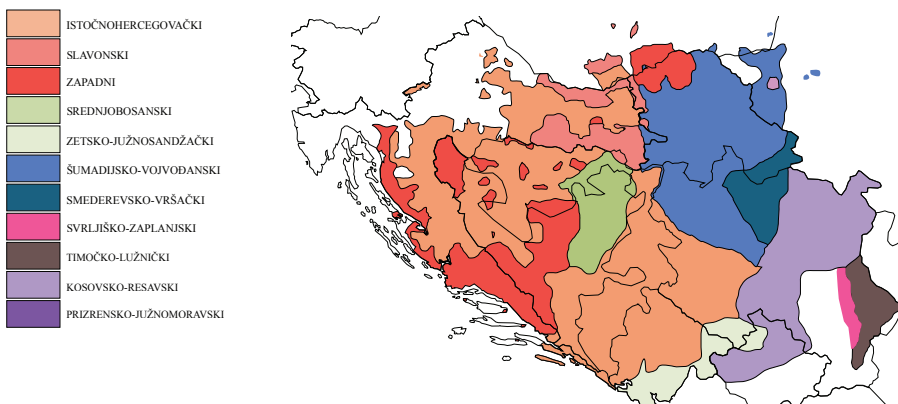


Figure 7.1: Štokavian dialects. Author: Dr. sc. Branimir Brgles

Table 7.1: An overview of Štokavian dialects

Countries	Ivić (2001)	Lisac (2003)	Okuka (2008)
		Neo-Štokavian dialects	
Ba, Hr, Me, Rs	Istočnohercegovački	Istočnohercegovačko-krajiški	Istočnohercegovačko-krajiški
Ba, Hr, Me, Rs	Šumadijsko-vojvođanski	Šumadijsko-vojvođanski	Šumadijsko-vojvođanski
Ba, Hr, Rs	Mlađi ikavski	ZAPADNI	Zapadnohercegovačko-primorski
		Old Štokavian dialects	
Me, Rs, Kos	Zetsko-sjениčki	ZETSKO-JUŽNOSANDŽAČKI	Zetsko-raški
Kos, Rs	KOSOVSKO-RESAVSKI	KOSOVSKO-RESAVSKI	KOSOVSKO-RESAVSKI
Rs	SMEDEREVSKO-VRŠAČKI	SMEDEREVSKO-VRŠAČKI	—
Ba, Hr	—	Istočnobosanski	SREDNJOBOSANSKI
Ba, Hr, Rs	Ekavski slavonski, Posavski ikavski	SLAVONSKI	SLAVONSKI
		Middle Štokavian dialects/Torlac dialects	
Rs, Kos	PRIZRENSKO-JUŽNOMORAVSKI		PRIZRENSKO-JUŽNOMORAVSKI
Rs	SVRLJIŠKO-ZAPLANJSKI	Torlački	SVRLJIŠKO-ZAPLANJSKI
Rs	TIMOČKO-LUŽNIČKI		TIMOČKO-LUŽNIČKI

As may be seen in Figure 7.1, Štokavian is spoken in nearly half of Croatia, and in all of Bosnia and Herzegovina, Montenegro, and Serbia (Lisac 2003: 15). Additionally, Štokavian is spoken in Italy (Molise), Austria (Vlahija in Burgenland), Hungary (various settlements), Romania (Rekaš), Slovenia (Bojanci and Marindol), Kosovo and Macedonia.

The internal classification of Štokavian dialects is usually done according to the following criteria: accents, reflex of the vowel [ě] (jat), and šćakavism or štakavism in some words.<sup>1,2,3</sup> Apart from features which are important as factors which help us distinguish different Štokavian dialects, there are common Štokavian features which clearly differentiate them from Čakavian and Kajkavian dialects. Lisac (2003: 17f) lists the following as the main features of Štokavian:

1. *što/šta* ‘what’ as an interrogative pronoun;
2. differentiation between two short and two or three long (with rising and falling intonation) types of accent;
3. non-accented long syllables are preserved (but not equally in all idioms);
4. morphological alternations: *jt* → *ć* (e.g. *pojti* → *poći* ‘go’), *jd* → *đ* (*pojdem* → *pođem* ‘I go’);
5. new iotation of dental and labial consonants with a lot of exceptions especially in Bosnia and Slavonia (*djevojka* → *đevojka* ‘girl’);
6. the loss of the phoneme [h] (with some exceptions);
7. the ending *-a* in the genitive plural for feminine and masculine nouns, with a lot of exceptions ;

<sup>1</sup>Neo-Štokavian idioms have the same accent system as the standard varieties, with four kinds of accent based on the combination of two features: pitch accent (rising or falling) and length (short or long). The combination of these two features results in four accents (long rising<sup>˘</sup>, long falling<sup>˘</sup>, short rising<sup>˘</sup>, short falling<sup>˘</sup>). In Neo-Štokavian dialects non-accented long syllables (marked with <sup>˘</sup>) appear only after accented syllables. In contrast, in Old Štokavian dialects non-accented long syllables can appear also before accented syllables (cf. Lisac 2003: 23).

<sup>2</sup>The Proto-Slavonic vowel [\*ě] was probably a long open front vowel. It is possible that even in this early period of the Slavonic languages’ development the pronunciation of this vowel varied greatly between dialects. In further development the vowel underwent changes and in the BCS area it was replaced by vowels [e], [i] or [ie] (the last is a diphthong). Thus the Proto-Slavonic \**děto* ‘child’ became *dete* in ekavian, *dite* in ikavian and *dijete* in ijekavian.

<sup>3</sup>Šćakavian dialects use *šč* (*ognjišće*) and *žd* (*zviždi*) while in štakavian dialects the same words have *št* (*ognjište* ‘fireplace’) and *žd* (*zviždi* ‘s/he whistles’).

## 7 *Clitics in dialects*

8. the ending *-u* for masculine and neuter nouns in the locative singular;
9. the infix *-ov/-ev-* for masculine nouns in the plural, with a lot of exceptions, especially between the Neretva and Dubrovnik;
10. syncretism of the dative, locative and instrumental plural for nouns, with many exceptions;
11. preservation of the aorist, with some exceptions around Dubrovnik;
12. numerous borrowings from Turkish.

For more information on the distinctive phonological, morphological and syntactic features of Štokavian dialects see Lisac (2003: 19–26).

## 7.3 Available data

### 7.3.1 Types of available data

Dialectology is an important research field at universities and research institutions in Bosnia, Croatia and Serbia: for instance, it is an obligatory part of the study programme for students who decide to study Croatian language and literature in Zagreb, Osijek, Rijeka and Split. The leading Institut za hrvatski jezik i jezikoslovlje (Institute of Croatian language and linguistics) has a separate department of dialectology. Similarly the Institut za srpski jezik Srpske akademije nauka i umetnosti (Institute for the Serbian language of SASA) conducts dialectological projects. This tradition goes back to pre-Yugoslavian times and for instance the anthology series *Srpski dijalektološki zbornik* was first published in 1905 and *Hrvatski dijalektološki zbornik*, in 1956. Besides these anthologies there are special dialectological journals such as *Čakavska rič*, *Kaj* etc. Moreover, many PhD dissertations are dedicated to idioms of individual villages or regions.

Published studies are usually based on data collected during fieldwork carried out by the researcher. In most cases these are interviews with NORM (non-mobile old rural male) speakers and the focus is on phonetics and phonology. Morphology and syntax are not studied thoroughly and the parts which are dedicated to these subdisciplines mainly consist of incompletely structured observations on archaic or innovative forms for cases, tenses, word order and sometimes agreement. Lexis is also usually poorly studied in general studies of some idioms. Scholars normally limit themselves to observations on archaic vocabulary, and German, Hungarian, Italian, Romanian or Turkish lexical borrowings. Nonetheless, some specialised studies such as Plotnikova (1997), Marinko & Baščarević



(2005), Marasović-Alujević & Knezović (2018), Vuletić & Skračić (2018), Horvat (2018), and Filipan-Žigniċ (2013) focus exclusively on lexis.

Studies published in journals are rarely accompanied by transcripts of interviews in the appendix, while some studies published in *Srpski dijalektološki zbornik* and *Hrvatski dijalektološki zbornik* provide only excerpts from interviews. PhD dissertations usually do not even provide excerpts and since in most cases researchers have to finance their fieldwork themselves, they tend to be unwilling to share their materials. Some excerpts can be found in Lisac (2003, 2009) and Okuka (2008) dialectological handbooks and in Menac-Mihalić & Celinić (2012) dialectological reader. Importantly, as far as we know no dialectological transcripts have been digitalised. Menac-Mihalić & Celinić (2012) are the only who attach an audio CD.

While investigating the literature, we focused first of all on three dialectology handbooks of Štokavian: Ivić et al. (2001), Lisac (2003) and Okuka (2008). Next, we concentrated on the most extensive and popular sources of dialectological studies, *Hrvatski dijalektološki zbornik* and *Srpski dijalektološki zbornik*. In this case we decided to take into account only the anthologies published after 1950. The reason is very simple: if the study was published before 1950, then the fieldwork was conducted even earlier and the informants were most probably born in the 19<sup>th</sup> century, and thus the language they spoke may have been very different from the language of the speakers who live in the same area now. Additionally we decided to include all open-access papers which included data on CLs in dialects and were available through the journal portal *Hrĉak*.<sup>4</sup> Furthermore, the data were supplemented with the dialectological literature (mainly journals and PhD dissertations) available at the library of Institute of Croatian language and linguistics in Zagreb, which is mostly inaccessible outside Croatia.

### 7.3.2 Data quality

In this section we would like to comment on the quality of publicly available printed excerpts of dialectological texts based on examples (1–4), which are taken from three different handbooks published in this century.

- (1) /0.00/ – Kà smo ċinili maturālnu večeru, i tō smo išli u Dūbrovaĉku Rĳĉku. Tada bila tete Jele. Ti neĉeš to znat Źe je to. /0.08/ I tamo, sve to skupa, bili, bālali. Vidimo mi: profesur ti s nŃ bālā stālno. /0.14/ A onda, ajde pjevaj – Kāte pjevā ko staglin, razumĳeš, izjutra rano, sve do zore smo ostali. /0.20/

<sup>4</sup>For more information visit <https://hrĉak.srce.hr/>.

I poslje smo se uputili, a oni dvoje skupa. Nama se odmađ štaklo. A leti se vjenčali, razumješ.

Marija Matana, Dubrovnik. Recorded by Martina Lobaš, 1989  
(Menac-Mihalić & Celinić 2012: 70)

- (2) Mladić bi se vjerō, pa bi hōdō nēkoliko vrēmēna. Pōslije bi dōšō u rōditējā, da se pītā, da se prōsī. Dohōdijo mu je ōtac, ili brāt, ōni dvā bi dōšli u prōšnu. Mladić i nēkā nēgova svōjta – ako nē bi ĩmo ōca, ĩmō bi dūnda. Ōnda bi se ugovōrilo nakon kōlikō će se vjēnča. I ōnda bi bila proglāšēna, trī pūta i bila bi svādba – kad bi se odlūčilo. Tū bi se pjēvalo, pīlo. Pjēvali su, nāzdravļali: Lijepō ime Āne, Bog jē živijo, mnōgo ljētā srētna bila, mnōgo ljētā živjela...

BHDZ, VII, 1996, 238, 241, 243  
(Lisac 2003: 113f)

- (3) Živila sam ū slami ōsam gōdīna. A pūno đecē. Dēsetero đecē sam rōdila. I skūpla ōnu đecu, sād nēko je ĩ ĩumrlo, odrānla sam i sēdmero. I u tōj slāmi sam i odrānla. Bījedno, bījedno odrānla. I mālō pōmalo i đeca, kākō kōje mālō jāče, ono pōmāže, pōmogne mi. I tākō ĩ jōnda jā prēdem ōvdekā. Tām sam bila kō šume. Pōšle, vāla bōgu, đeca dōbra, pāmetna mi đeca, dōbro mi.

Đūja Todōrović. 82. yr., Trnjaci [Semberija], Subotić 1973, 125  
(Okuka 2008: 109)

- (4) Jednostavno ne samo što gubimo materijalna dobra, mi, mi imam... mi gubimo žrtv... ne, narod uopšte, ne, ne samo svoje bližnje, nego uopšte narod...

The Tübingen corpus of spoken Bosnian language, Transcript BH

The first three examples represent *Istočnohercegovački* dialect. In (1) the stress is marked only above the words in the first sentence, while in other excerpts it is marked consistently everywhere. The name of the speaker is given in examples (1) and (3), while age is given only in the latter. Both examples provide the year in which the recording was taken. In (2) only information on the source of the excerpt is given, while finding all other data requires access to the original. Information on the year and the place of the fieldwork, like the sex, age and education of informants, is crucial and it should always be provided by the dialectologist.

In some cases informants told folk stories, which were recorded and transcribed as dialectological material. Although we believe that this kind of material is valuable for many reasons, much speaks against including it in the analysis of

a given dialect. Above all, this kind of material is often learned by heart from senior members of the community, and contains structures which are not actively in use. The interpretation of structures found in this kind of material as a distinctive features of the dialect/idiom of interest may give a distorted overall picture of that dialect/idiom.

We would like to address one more data quality problem. If we compare the examples presented in (1–3) with the one in (4), we see that the excerpt presented in (4) provides a more natural depiction of the speech flow. (4) is an example of the speech of a mobile non-rural middle aged woman from Bosnia (born in Serbia). It is hard to believe that the speech flow of rural informants can be as fluent as is presented in examples (1–3). Therefore, we assume that those who transcribed the interviews undertook certain interventions and made the texts look more like written texts and less like free, unprepared spoken language.

### 7.3.3 Examples in this chapter

Since as we demonstrated, dialectologists do not present their data uniformly, we decided to present the examples provided in this chapter in their original form, i.e. according to the transcription which was used by the author we quote. In the glossed examples we also provide the dialect or local idiom name. The names of the dialects are used according to the terminology proposed in the previous Section 7.2. Some authors write about the idioms of certain villages and they do not specify to which dialect these belong. In such cases, we estimated the dialect according to the borders of dialects in dialectological maps in Ivić et al. (2001), Lisac (2003) and Okuka (2008). Sometimes in the reviewed literature, the dialectological material was described not only with respect to the dialect, but also the subdialect or even the local idiom. Whenever these data are available, we provide them in the running text for the sake of future research. As we show below, the CL system may vary even between local idioms which belong to the same subdialect or even dialect.

Some examples presented here come from dialectological handbooks or papers and they are cited from running texts and not from transcripts. As it turns out, in many cases the quoted authors do not use full sentences, i.e. they start their examples without a capital letter and finish them without punctuation marks. We quote such examples exactly as they appear in the original running text. However, when we quote examples from transcripts and if we do not need the whole sentence to illustrate our argumentation, we use the symbol [...] to indicate that the beginning and/or the end of the sentence is missing.

## 7.4 Inventory

### 7.4.1 Pronominal clitics

In the following subsections we concentrate on CL forms which diverge from those used in BCS standard varieties. Furthermore, we discuss CL forms which caused many disputes among scholars in ex-Yugoslavia, for instance the third person feminine accusative CL *ju* and reflexive CL *si*. Note that we did not find data for pronominal CLs of the first and second person singular. Dialectologists usually find only archaic or innovative features or features which are divergent in some way to be worth noting and commenting. We assume that in such cases the forms very probably correspond to those in the respective standard variety. This would indicate that pronominal CLs of the first and second person singular do not diverge from forms in BCS standard varieties.

Every subsequent empirical study of CLs in BCS dialects has to begin with the established inventory of the units which will be examined. In that respect, the following subsections on the inventory of CLs may provide valuable information. However, readers who are not interested in this comprehensive account of the inventory can skip this part.

#### 7.4.1.1 Feminine pronominal clitics

##### 7.4.1.1.1 Feminine pronominal clitics in the accusative

In Section 6.3.1 we discussed certain differences in the third person singular feminine accusative CL in standard varieties. While standard Bosnian and Serbian use *je* as the default, an increase in the use of *ju* has been observed in standard Croatian since the end of the 20<sup>th</sup> century. If we look at Table 7.2, we see that both *ju* and *je* forms are attested in Old and Neo-Štokavian dialects.<sup>5</sup>

While in the far East of Croatia in the local idiom of Ilok (Neo-Štokavian *Šumadijsko-vojvođanski* dialect), in the idioms of Baranja (Old Štokavian *Slavonski* dialect) and in Zagreb (*Turopoljski* Kajkavian dialect) the form *ju* is in use, Croats in Southern Croatian areas such as Sinj, Bitulić and Imotski (Neo-Štokavian *Zapadni* dialect) prefer *je* (cf. Lisac 2003: 130, Sekereš 1977: 331, Hoyt 2012: 64, Ćurković 2014: 185, Šimundić 1971: 120).<sup>6</sup> In the neighbouring Neo-Štokavian

---

<sup>5</sup>For some CLs we did not find any information in the reviewed dialectological literature. We indicate these cases with “data NA” in the relevant table fields.

<sup>6</sup>This does not mean that all idioms of *Šumadijsko-vojvođanski* employ *ju* as a CL; we are aware of certain differences. For instance, Radovanović (2006: 259) claims that the form *ju* is not attested in the language of her informants from Kolubara.

Table 7.2: CL forms of the third person singular feminine pronoun

Dialect (subdialect or idiom)	her.ACC	her.DAT
<i>Šumadijsko-vojvođanski</i> (Ilok)	<i>ju</i>	data NA
<i>Istočnohercegovački</i>	<i>je</i>	<i>jon</i>
<i>Zapadni</i> (Sinj, Bitelić)	<i>je</i>	<i>jon</i>
<i>Slavonski</i>	<i>ju</i>	data NA
<i>Srednjobosanski</i> (Fojnica) (Sarajevo)	<i>je/ju je</i>	<i>joj</i>
<i>Zetsko-južnosandžački</i> (Istočnocrnogorski)	<i>ju</i>	data NA
<i>Kosovsko-resavski</i>	<i>ju</i>	<i>ju</i>
<i>Timočko-lužnički</i> (Vlasinski)	<i>gu</i> beside <i>ju</i>	<i>đu</i> beside <i>voj</i>
<i>Timočko-lužnički</i> (Lužnički)	<i>ju</i>	<i>voj</i>
<i>Svrljiško-zaplanjski</i>	<i>ju</i>	<i>voj, vu</i>
<i>Prizrensko-južnomoravski</i>	<i>gu, ga, ja, je, ju, u</i>	<i>gu, gi, i, je</i>

*Istočnohercegovački* dialect the CL *je* dominates as well (see 5), although there are some local idioms such as Grude where only *ju* is attested (cf. Halilović 1996: 174, Peco 2007a: 200, Peco 2007b: 311).<sup>7</sup> In some local idioms of *Istočnohercegovački* dialect, for instance in Banja Vrućica, and in the local idiom spoken in the Neretva river valley, the CL *ju* does not exist at all (cf. Dragičević 2007: 371, Vukša Nahod 2014: 142).

- (5) Bog *jē* živijo  
 God her.ACC live.PTCP.SG.M  
 ‘May God give her a long life’ (Istočnohercegovački; Lisac 2003: 113)

In contrast to the mixed *je/ju* distribution in Neo-Štokavian dialects, according to the dialectological data it seems that in Old and Middle Štokavian dialects the form *ju* is dominant. It is the only variant found in the *Istočnocrnogorski* idiom (*Zetsko-južnosandžački* dialect) and in the idioms of North Metohija (*Kosovsko-resavski* dialect) (cf. Peco 2007a: 200, Bukumirić 2003: 223).<sup>8</sup> Halilović (1996: 174) even assumes that this form originates from idioms spoken in Montenegro.

<sup>7</sup>According to Peco (cf. Peco 2007a: 200, Peco 2007b: 311) the Grude idiom is part of the *Istočnohercegovački* dialect, whereas dialectological map in Lisac (2003: 162) has Grude as a part of the *Zapadni* dialect. In both cases it is Neo-Štokavian.

<sup>8</sup>Peco (2007a: 200) uses the term *Zetsko-gornjopolimski* dialect instead of *Zetsko-južnosandžački* dialect.

## 7 Clitics in dialects

- (6) [...] da *ju* izvūčë iz jàmë.  
that her.ACC pull.out.3PRS from pit  
‘[...] to pull her out of the pit.’ (Zetsko-južnosandžački; Okuka 2008: 189)

However, it seems that the form *je* is prevalent in the Old Štokavian *Srednjo-bosanski* dialect (cf. Brozović 2007: 126, Halilović et al. 2009: 58). In two Middle Štokavian idioms of the *Timočko-lužnički* dialect, *Vlasinski* and *Lužnički*, like in the *Svrljiško-zaplanjski* dialect, the third person singular feminine accusative CL is *ju* (cf. Okuka 2008: 272f, 254). In contrast to the latter two Torlac dialects, the *Prizrensko-južnomoravski* dialect additionally uses various other forms such as *gu*, *ga*, *ja*, *je* and *u* besides *ju* (cf. Stevanović 1950: 111, Okuka 2008: 237, Mladenović 2010: 51). There are, however, certain differences among local idioms of that dialect. Stevanović (1950: 111) claimed that in the local idiom of Đakovica the pronominal CL *je* is never used and data from the beginning of the 21<sup>st</sup> century do not indicate any changes (cf. Mladenović 2010: 51). However, some other idioms of *Prizrensko-južnomoravski* do have the CL *je* (cf. Mladenović 2010: 51).

### 7.4.1.1.2 Feminine pronominal clitic in the dative

As may be seen in Table 7.2, the dative forms of the feminine pronominal CL vary between dialects even more than the accusative ones. The speakers of the local idiom of the Neretva valley use the CL *jon* exclusively (cf. Vukša Nahod 2014: 143), whereas among speakers of the local idiom of Bitelić only the older generation uses this variant while the language of younger speakers is slowly changing in the direction of the standard variety in this regard (cf. Ćurković 2014: 186).<sup>9</sup> The dative CL *jon* is attested in the local idiom of Dubrovnik as well: see example (7).

- (7) nema *jon* spasa  
NEG.have.3PRS her.DAT salvation  
‘she cannot be saved’ (Istočnohercegovački; Okuka 2008: 301)

All the local idioms mentioned are Neo-Štokavian. While local idioms spoken in the Neretva valley and in Dubrovnik belong to *Istočnohercegovački*, the local idiom of Bitelić belongs to the *Zapadni* dialect.

In Middle Štokavian Torlac dialects such as *Timočko-lužnički* and *Svrljiško-zaplanjski* the third person singular feminine dative CL is *voj* (cf. Okuka 2008: 272, 254).<sup>10</sup> Additionally, the form *vu* is also attested in the latter (cf. Okuka 2008: 254).

<sup>9</sup>Šimundić (1971: 120) claims that *jon* is the younger form.

<sup>10</sup>However, it is not the only possible form for the third person feminine dative CL since in the *Vlasinski* idiom speakers also use *đu* besides *voj* (cf. Okuka 2008: 273).

As in the case of the third person singular feminine accusative, various forms of the dative feminine pronoun are attested in the *Prizrensko-južnomoravski* dialect: *gu*, *gi*, *i* and *je* (cf. Okuka 2008: 237, Mladenović 2010: 51).

In contrast, in some Old Štokavian dialects, such as *Srednjobosanski*, the CL *joj*, the standard form in all varieties of BCS, is used (cf. Halilović et al. 2009: 58).

#### 7.4.1.2 Masculine pronominal clitics

In Table 7.3 we give a short overview of different CL forms for the third person singular masculine pronoun. Since there are no data on the dative CL form, we assume that it does not differ from the form used in the standard variety. The accusative forms mentioned in the dialectological literature will be discussed in detail below.

Table 7.3: CL forms of the third person singular masculine pronoun

Dialect (subdialect or idiom)	him.ACC	him.DAT
<i>Zapadni</i> (Sinj, Bitelić)	<i>nj</i>	data NA
<i>Istočnohercegovački</i> (Grude, Neretva)	<i>nj</i> , <i>jn</i> <i>nje</i>	data NA

As mentioned in Section 6.3.1 the CL *nj*, which is used exclusively with prepositions, is considered archaic in contemporary standard Serbian, but seems to be frequent in some Neo-Štokavian idioms.

Speakers of *Istočnohercegovački* dialect use the third person singular masculine accusative CL *nj* for both animate and inanimate referents (cf. Peco 2007a: 200, Peco 2007b: 311). In the local idiom of Bitelić the CL *nj* is used only after prepositions with an additional vowel [a], whereas in the local idiom spoken in the Neretva valley it is used after all prepositions (cf. Ćurković 2014: 185, Vukša Nahod 2014: 143). Besides the CL form *nj*, in the local idiom spoken in the Neretva valley the equivalent form *jn* exists as well (cf. Vukša Nahod 2014: 142f).<sup>11</sup> In addition to the two mentioned accusative forms, in the local idiom of the Neretva valley the variant *nje* is found (cf. Vukša Nahod 2014: 143). This form is employed only after prepositions, as in example (8).<sup>12</sup>

<sup>11</sup>Vukša Nahod (2014) does not specify in what context speakers of the local idiom of the Neretva valley use the CL *jn*. Moreover she does not state whether there are other differences between *nj* and *jn* besides formal ones.

<sup>12</sup>Vukša Nahod (2014) does not specify whether the CL *nje* follows only a certain type of preposition, or any preposition used with the accusative case.

## 7 Clitics in dialects

- (8) ná *nje*      *se*      *mětnē*  
 on him.ACC REFL put.3PRS  
 ‘one puts it on him/it’      (Istočnohercegovački; Vukša Nahod 2014: 143)

At the end of this subsection we would like to point out that in the dialectological literature reviewed we did not find any information on CL forms used without prepositions for the third person masculine pronoun in the accusative. Therefore, we assume that it does not differ from the form used in BCS standard varieties, where only non-clitic form may be used in this context.

### 7.4.1.3 Plural pronominal clitics

#### 7.4.1.3.1 Plural pronominal clitics in the accusative

In some Štokavian dialects accusative pronominal plural forms differ strikingly from the forms used in the contemporary standard varieties of BCS. A short overview of these forms is presented in Table 7.4.

Table 7.4: CL forms of the accusative plural pronouns

Dialect (subdialect or idiom)	we.ACC	you.ACC	they.ACC
Šumadijsko-vojvođanski (Kolubarski)	<i>ne</i>	<i>ve</i>	data NA
Istočnohercegovački (Zapadnocrnogorski) (Kula, Cuce)	<i>ne</i>	<i>ve</i>	<i>hi, hig, hin, ih,</i> <i>i, ig, ik, ji, jig, jih</i>
Kosovsko-resavski	<i>ne</i>	<i>ve</i>	<i>i, ju</i>
Slavonski	data NA	data NA	<i>je, i</i>
Zetsko-južnosandžački	<i>ne</i>	<i>ve</i>	data NA
Svrljiško-zaplanjski	<i>ni</i>	<i>vi</i>	<i>i</i>
Prizrensko-južnomoravski	<i>ne</i>	<i>ve</i>	<i>gi, ge, giv, i, i(h), ji</i>

Okuka (2008: 74) and Pešikan (1965: 152) claim that speakers of the *Zapadnocrnogorski* subdialect use the CL form *ne* instead of the first person plural accusative CL *nas*. However, following Peco (2007a) we must emphasize that *ne* is not typical of the entire territory of the *Istočnohercegovački* dialect. Specifically,



this form is not present in local idioms of Eastern Herzegovina (cf. Peco 2007a: 297). Additionally, Peco (2007a: 197) claims that he found a single example of the archaic second person plural accusative CL *ve* on the territory of Eastern Herzegovina, in the local idiom of Kula, as in example (9).

- (9) kămo ve jöš?  
 where you.ACC still  
 ‘where are you going?’ (Istočnohercegovački; Peco 2007a: 197)

Okuka (2008: 141) lists *ne* and *ve* as the plural CL accusative forms used in the Neo-Štokavian *Kolubarski* subdialect (*Šumadijsko-vojvođanski* dialect). The mentioned forms are preserved in the Middle Štokavian *Prizrensko-južnomoravski* dialect as well (cf. Stevanović 1950: 110, Mladenović 2010: 46). The CL forms *ne* and *ve* are also characteristic of the Old Štokavian *Zetsko-južnosandžački* dialect (cf. Barjaktarević 1966: 88). Moreover, they are a trait connecting the idioms of the *Zetsko-južnosandžački* dialect to the Old Štokavian idioms of the *Kosovsko-resavski* dialect and to idioms in the southeastern part of the Neo-Štokavian *Istočnohercegovački* dialect area (cf. Lisac 2003: 121). Additionally, Bukumirić (2003: 221) claims that the old CL forms *ne* and *ve* are well preserved in idioms of North Metohija (*Kosovsko-resavski* dialect).

According to Okuka (2008: 255) the first and second person plural accusative CLs *ni* and *vi* are preserved in the Middle Štokavian *Svrljiško-zaplanjski* dialect. These forms are used in the *Timočko-lužnički* dialect as well. Moreover, Ivić (1957: 201) claims that as accusative CLs, *ni* and *vi* are older than *ne* and *ve*. The usage of the CL *i* (and not *gi*) for the third person plural accusative differentiates the *Svrljiško-zaplanjski* dialect from the neighbouring Eastern and Southeastern Serbian idioms (cf. Okuka 2008: 255).<sup>13</sup> In contrast to the Torlac dialects just mentioned, which do not show such a great degree of variation with respect to the third person plural accusative CL, in the *Prizrensko-južnomoravski* dialect the following forms are attested: *gi*, *ge*, *giv*, *i*, *i(h)* and *ji* (cf. Okuka 2008: 237, Mladenović 2010: 52).

Variation affects the third person plural accusative CL in the *Istočnohercegovački* dialect as well, where scholars attest dozens of such forms. Peco (cf. Peco

<sup>13</sup>The existence of the third person plural accusative CL *gi* is a trait which connects Eastern and Southeastern Serbian idioms with Northeastern Macedonian and Western Bulgarian idioms (cf. Okuka 2008: 20).

## 7 Clitics in dialects

2007a: 202, Peco 2007b: 311) lists the following CL forms of the third person plural in the accusative and genitive: *hi*, *hig*, *hin*, *i*, *ig*, *ih*, *ik*, *ji*, *jig*, and *jih*.<sup>14,15</sup> As claimed by Peco (2007a: 202f), not all of these forms are equally widespread, and sometimes the same informant can switch from one form to another during one session. The sentences below exemplify the CL forms in the local idioms of Kula (10), Dabar (11) and Divin (12).

- (10) [...] a đeca sve viču: èto *hi*.  
and children all yell.3PRS there them.GEN  
'[...] and the children are all yelling: there they are.'  
(Istočnohercegovački; Peco 2007a: 281)
- (11) Bilo *ig* *je* prije više.  
be.PTCP.SG.N them.GEN be.3SG before more  
'There were more of them before.'  
(Istočnohercegovački; Peco 2007a: 285)
- (12) Ovi *ji* fino očekaju [...]  
these them.ACC nicely welcome.3PRS  
'They welcome them nicely.'  
(Istočnohercegovački; Peco 2007a: 287)

Besides the third person plural accusative form *i*, which appears in several Štokavian dialects, the *Slavonski* dialect preserves the old accusative plural CL form *je* (cf. Farkaš & Babić 2011: 48).

### 7.4.1.3.2 Plural pronominal clitics in the dative

In some dialects and their local idioms, dative plural CLs are the same as those which are part of standard varieties, as for example in the local idiom of Sarajevo (*Srednjobosanski* dialect) (cf. Halilović et al. 2009: 58). However, Table 7.5 reveals a great deal of variation in respect of dative plural CL forms. The data is reconstructed from the scattered information we found in dialectological literature.

In the *Istočnohercegovački* dialect the archaic CL form *ni* for the first person plural dative is rather rare; for instance, it can be found in the Montenegro village

<sup>14</sup>It seems that *hi* is a very old CL form. Halilović et al. (2009: 14) state that the CL *hi* was used in the local idiom of Sarajevo in the 18<sup>th</sup> century. They class this CL as a general Bosnian phenomenon, especially in more archaic idioms (cf. Halilović et al. 2009: 14). When speaking about the local idiom of Sarajevo in the 19<sup>th</sup> century, Halilović et al. (2009: 21) claim that there were differences between Muslim, Orthodox and Catholic speakers with respect to the CL form they tended to use. So, for instance Muslim speakers purportedly tended to use the enclitic *hi* the most often, *hin* more rarely, and *ih* the least, while Orthodox and Catholic speakers tended to use *i(h)* (Halilović et al. 2009: 21).

<sup>15</sup>Some of those forms, such as *hi*, *hin*, and *him*, are used in the *Srednjobosanski* dialect as well (Halilović 2005: 41).

Table 7.5: CL forms of the plural pronouns in the dative

Dialect (subdialect or idiom)	we.DAT	you.DAT	they.DAT
<i>Srednjobosanski</i> (Sarajevo)	<i>nam</i>	<i>vam</i>	data NA
<i>Istočnohercegovački</i> (Nikšićka Župa, Cuce)	<i>nam, ni</i> (rarely)	<i>vi, vam</i>	<i>him, im, jim, jin</i>
<i>Šumadijsko-vojvođanski</i> (Kolubarski, Šumadija)	<i>ni</i>	<i>vi</i>	data NA
<i>Zetsko-južnosandžački</i>	<i>ni</i>	<i>vi</i>	data NA
<i>Kosovsko-resavski</i>	<i>ni</i>	<i>vi</i>	<i>ju</i>
<i>Svrliško-zaplanjski</i>	data NA	data NA	<i>im</i>
<i>Prizrensko-južnomoravski</i>	<i>ni</i>	<i>vi</i>	<i>gi, gim, giv, i, im, ji, mgi</i>

Nikšićka Župa and in the Cuce tribe. In contrast, the archaic CL form *vi* for the second person plural dative is quite common and appears in everyday language, as does the form *vam*, which is in use in standard varieties (cf. Pešikan 1965: 152, Peco 2007a: 196f). The example presented in (13) is from the local idiom of Divin.

(13) [...] ali *vi* ne mògu dat òdgovōr [...].

but you.DAT NEG can.1PRS give.INF answer

‘[...] but I cannot give you the answer [...].’

(Istočnohercegovački; Peco 2007a: 287)

Okuka (2008: 63, 72) partially agrees with Peco (2007a: 196f) and underlines that the main trait of idioms in Eastern Herzegovina is the usage of the CL *vi* instead of *vam*. However, he admits that not all idioms of the *Istočnohercegovački* dialect use this feature. For instance, in the *jugozapadnosrbijanski* subdialect the CL *vi* appears only optionally, while in the *Sjevernozadnosrbijanski* subdialect only *vam* is used as a CL (cf. Okuka 2008: 78). Lisac’s (2003: 103) opinion differs slightly from Okuka’s (2008) and Peco’s (2007a); according to him, generally speaking in the *Istočnohercegovački* dialect as a whole the second person plural dative CL is *vam*, while the CL *vi* appears only in idioms spoken in Eastern Herzegovina.

However, authors (e.g. Barjaktarević 1966: 88, Lisac 2003: 121, Okuka 2008: 177) do agree that the CLs *ni* and *vi* are characteristic of the Old Štokavian *Zetsko-*

*južnosandžački* dialect.<sup>16</sup> The mentioned trait connects idioms of the *Zetsko-južnosandžački* dialect to the southeastern idioms of *Istočnohercegovački* and to idioms of *Kosovsko-resavski* (cf. Lisac 2003: 121, Okuka 2008: 205). Bukumirić (2003: 221) claims that the CL forms *ni* and *vi* are well preserved in idioms of North Metohija (*Kosovsko-resavski* dialect) and that the forms *nam* and *vam* are quite rare. These CLs are also present in the *Prizrensko-južnomoravski* Torlac dialect and in the Neo-Štokavian *Šumadijsko-vojvođanski* dialect (in idioms of central Šumadija and in the Kolubarski subdialect) (cf. Stevanović 1950: 110, Mladenović 2010: 46, Remetić 1985: 291, Okuka 2008: 141, 237).

Besides the third person plural dative CL *im*, Peco (Peco 2007a: 203, Peco 2007b: 311) mentions *jim* and *jin* as forms present in the Neo-Štokavian *Istočnohercegovački* dialect. Below are examples from the local idioms of Nevesinje (14) and Borač (15).

- (14) Mène šćèri zòvū da *jim* ìdēm.  
me.ACC daughters call.3PRS that them.DAT go.1PRS  
'My daughters are calling me to go to them.'  
(Istočnohercegovački; Peco 2007a: 282)
- (15) Švábo *him* né šće nīšta.  
German them.DAT NEG fut.3SG nothing  
'The German will not do anything to them.'  
(Istočnohercegovački; Peco 2007a: 283)

Whereas in the Middle Štokavian *Svrljiško-zaplanjski* Torlac dialect the third person plural dative CL is *im* (cf. Okuka 2008: 255, 237), speakers of the *Prizrensko-južnomoravski* Torlac dialect use several variants such as: *gi*, *gim*, *giv*, *i*, *im*, *ji*, *mgi* (cf. Okuka 2008: 237, Mladenović 2010: 34). One example with the CL *gi* from the local idiom of Prizren is presented in (16).

- (16) [...] otiša da *gi* čestíta.  
leave.PTCP.SG.M that them.DAT congratulate.3PRS  
'[...] he left to congratulate them.'  
(Prizrensko-južnomoravski; Okuka 2008: 247)

Ivić (1957: 202) claims that the CLs *ju* and *gi* as a third person plural dative form are in use only in those idioms which use *ni*, *vi* and/or *ne*, *ve* as first and second person plural dative and accusative CLs.

<sup>16</sup>In some subdialects of the *Zetsko-južnosandžački* dialect, such as in *Sjeničko-novopazarski* subdialect, dative CLs for the 1<sup>st</sup> and 2<sup>nd</sup> person are replaced by accusative ones (cf. Okuka 2008: 186).

### 7.4.2 Verbal clitics: Aoristal/conditional clitics of the verb *biti* ‘be’

In contrast to pronominal CLs it seems that verbal CLs do not vary much. In the *Istočnohercegovački*, *Zapadni*, *Šumadijsko-vojvođanski*, *Slavonski* and *Kosovsko-resavski* dialects the CL *bi* ‘would’ of the conditional auxiliary is used for all persons (cf. Peco 2007b: 331, Kurtović Budja 2009: 96, Radovanović 2006: 302, Remetić 1985: 327, Dragičević 2007: 377, Golić 1993: 106, Bukumirić 2003: 267). Lisac (2012: 42) admits that in the majority of Croatian idioms the usage of the CL form *bi* for all persons prevails, but he emphasises that in the local idiom of Dubrovnik this form and forms appearing in the standard are used equally often.<sup>17</sup> Peco (2007b: 331) believes that the CL form *bi* used for all persons is spreading as a trait from dialects into standard language; if not in its written, then certainly in its spoken registers.<sup>18</sup> Accordingly, Aladrović (2011: 165) reports this feature in the written language of elementary school students from Požega.<sup>19</sup>

Some variation with respect to conditional auxiliary CLs was detected in the local idiom spoken in the valley of the river Fojnica (*Srednjobosanski* dialect) and in the Cuce tribe (*Istočnohercegovački* dialect). Namely, the CL *biš* ‘they would’ was attested there (cf. Brozović 2007: 137, Pešikan 1965: 171). Furthermore, Pešikan (1965: 171) attested *bihu* as a third person plural auxiliary in the Bjelice and Zagarač Montenegrin tribes.

### 7.4.3 Reflexive clitic *si*

The REFL<sub>2ND</sub> CL *si* is only part of standard Croatian, while Bosnian and Serbian normativists do not include it in the inventory of standard Bosnian and standard Serbian (for more information see Section 6.3.3).<sup>20</sup> However, as we will show in this section, this form is present in Štokavian dialects spoken on Bosnian and Serbian territory.<sup>21</sup>

The REFL<sub>2ND</sub> CL *si* is quite common in Kajkavian dialects: see the example from the *Gornjolonjski* Kajkavian dialect presented in (17) (cf. Brlobaš & Lončarić 2012: 242).

<sup>17</sup>Lisac (2012: 42) adds that the speakers of Čakavian also use, among others, *bin*, *biš*, *bi*, *bimo*, *bite*, and *bi*, while in Kajkavian one form, *bi*, is usually used for all persons (for Čakavian see Menac-Mihalić 1989).

<sup>18</sup>For the results of our analysis of spoken Bosnian with respect to this matter see Section 8.7.3.

<sup>19</sup>According to the dialectological map, Požega belongs to the Old Štokavian *Slavonski* dialect, but younger generations probably speak the *Istočnohercegovački* dialect.

<sup>20</sup>Our typology of reflexives is presented in Section 2.5.4.2.

<sup>21</sup>Moreover, the mentioned form is also present in the spoken variety of Bosnian: for more details see Section 8.7.4.

- (17) mōram si mále počinōti  
 must.1PRS REFL little rest.INF  
 ‘I have to rest a little’ (Gornjolonjski; Brlobaš & Lončarić 2012: 243)

This form is also widely used in the local idiom of Zagreb (Hoyt 2012: 65), and in the local idiom of Žumberak (cf. Težak 1985: 25) – see example (18) below.<sup>22</sup>

- (18) Kúpijo sam si knjīgu.  
 buy.PTCP.SG.M be.1SG REFL book  
 ‘I bought myself a book.’ (Žumberak idiom; Težak 1985: 255)

According to the dialectological data, the CL *si* is not very typical of Štokavian and Čakavian dialects, although it is used occasionally.<sup>23</sup> The reflexive CL *si* does not exist in the following Neo-Štokavian idioms: in the local idioms of Bitelić and Imotski (*Zapadni* dialect), in the local idioms of the Neretva valley and in the local idiom of Banja Vrućica (*Istočnohercegovački* dialect), in idioms of central Bosnia (*Zapadni* dialect) and in idioms of Kolubara (*Šumadijsko-vojvođanski* dialect) (cf. Ćurković 2014: 192, Šimundić 1971: 120, Vukša Nahod 2014: 142, Dragičević 2007: 371, Peco 1990: 207, Radovanović 2006: 255).<sup>24</sup> Those idioms only have the full reflexive form *sebi* in the dative. However, the dative REFL<sub>2ND</sub> CL *si* can be found in some Neo-Štokavian idioms, e.g. in Western Herzegovina (*Zapadni* and *Istočnohercegovački* dialect), although it is claimed to be rare (cf. Peco 2007b: 311). This form is also found in some Old Štokavian idioms of Northeastern Bosnia (*Slavonski* dialect) (cf. Peco 1985: 269). In contrast, in the Middle Štokavian *Svrljiško-zaplanjski* Torlac dialect the usage of the dative CL *si* is frequent (cf. Okuka 2008: 255). Ivić (1957: 205) claims that the mentioned CL can be found in the Middle Štokavian *Prizrensko-južnomoravski* dialect as well. Mladenović (2010: 45) later corroborated this claim and found this CL in six out of nine investigated idioms of the *Prizrensko-južnomoravski* dialect.<sup>25</sup> Furthermore, it seems that Old Štokavian idioms are closer to the Middle Štokavian idioms with respect to the reflexive CL *si*. Specifically, the form in question is also found in some Old Štoka-

<sup>22</sup>In the idiom of Žumberak the features of all three dialects, Štokavian, Kajkavian and Čakavian, are present.

<sup>23</sup>Vranić (2003: 158) claims that in the Čakavian idioms of Pag island the long reflexive form is used more often than the CL one. Moreover, she provides Čakavian examples in which the reflexive in the dative is replaced with the construction: ‘preposition + reflexive in accusative’ or ‘preposition + personal pronoun in accusative’ (cf. Vranić 2003: 158). Vranić (2003: 158) claims that such substitutions are quite frequent.

<sup>24</sup>Some central Bosnian idioms belong to the Old Štokavian *Srednjobosanski* dialect.

<sup>25</sup>For a recent corpus linguistic study on the reflexive CL *si* in Torlak dialect, see Ćirković (2021).

vian idioms of Northeast Bosnia (*Slavonski* dialect) and in Novopazarsko-sjenički idioms (*Zetsko-južnosandžački* dialect) (cf. Peco 1985: 269, Barjaktarević 1966: 90).

#### 7.4.4 Stress on clitics in BCS dialects

It is a well-known fact that CLs behave differently in Kajkavian and Čakavian dialects. Therefore, it should not come as a surprise that both pronominal and verbal CLs can be stressed there, which is a consequence of the general rule of moving stress to the penultimate syllable of the stress unit. An example of this feature from the Kajkavian local idiom of Virje is presented in (19).

- (19) samo *sěm* se obudila  
 just be.1SG REFL wake.up.PTCP.SG.F  
 ‘I just woke up’ (Podravski; Maresić 2011: 463)

However, something similar is present in Štokavian dialects as well. For instance, in the Neo-Štokavian local idiom of Bitelić (*Zapadni* dialect), the CL for the third person plural accusative can be a long syllable (cf. Ćurković 2014: 186), as may be seen in example (20).

- (20) ôn *ī* je pítā  
 he them.ACC be.3SG ask.PTCP.SG.M  
 ‘he asked them’ (Zapadni; Ćurković 2014: 186)

Moreover, in the Middle Štokavian *Svrljiško-zaplanjski* Torlac dialect the stress can be placed on any syllable in a word, i.e. it can also be placed on the CL (Okuka 2008: 254). This trait differentiates the *Svrljiško-zaplanjski* dialect from its neighbouring *Timočko-lužnički* Torlac dialect (cf. Okuka 2008: 257).

## 7.5 Internal organisation of the clitic cluster

### 7.5.1 Clitic ordering within the cluster

In many dialects the order of CLs in the cluster can differ from the order in BCS standard varieties.<sup>26</sup> The most common difference concerns the order of the reflexive CL *se* and verbal CL *je*, and is attested in both Old and Neo-Štokavian dialects.<sup>27</sup> With respect to the former, Brozović (2007: 150) reports the reversed

<sup>26</sup>For CL order in the cluster in BCS standard varieties see Section 2.4.2.1.

<sup>27</sup>As we already pointed out in Section 6.4.2.2 although the CL sequence *se je* is (hypothetically) possible in BCS standard varieties, in contrast to the CL sequence *je se* which is not possible

## 7 Clitics in dialects

*je se* order for the local idiom spoken in the Fojnica valley (21) and Kolenić (1999: 46), for the local idiom of Ilača (22).<sup>28</sup>

(21) òna *je se* obúkla  
she be.3SG REFL dress.PTCP.SG.F  
'she dressed herself' (Srednjobosanski; Brozović 2007: 150)

(22) ònda *je se* glèdalo  
then be.3SG REFL watch.PTCP.SG.N  
'then it was watched' (Slavonski; Kolenić 1999: 46)

Examples of *je se* CL order are found in the Neo-Štokavian *Šumadijsko-vojvođanski* dialect, which is a neighbouring dialect of the Old Štokavian *Slavonski* dialect (cf. Nikolić 1966: 279, Okuka 2008: 136). The example in (23) is from the local idiom of Petnica.

(23) Majka *mi je se* zvala  
mother me.DAT be.3SG REFL call.PTCP.SG.F  
'My mother was called' (Šumadijsko-vojvođanski; Okuka 2008: 136)

Lisac (2003: 58), Halilović (2005: 33), and Ćurković (2014: 309) provide examples of the reversed *je se* order in the *Zapadni* dialect. Lisac (2003: 58) even claims that the verbal CL *je* consistently appears before the reflexive CL *se* in the *Zapadni* dialect – see example from the local idiom of Derventa in (24).

(24) Àntūn *je se* čúvō ùvik.  
Antun be.3SG REFL guard.PTCP.SG.M always  
'Antun always guarded himself' (Zapadni; Lisac 2003: 58)

This kind of reversed CL order can also be found in the speech of younger generations. Aladrović (2011: 165) mentions it as a dialectal feature in the language of elementary school students from Požega. Furthermore, the reversed *je se* CL cluster can be found in Štokavian idioms which are spoken in the territory of

---

there, it is discussed rather controversially by normativists. Specifically, some grammarians recommend deletion of the verbal CL *je*, i.e. haplogy of unlikes. However, as we show in the next section, haplogy is not restricted only to standard varieties. It also occurs in varieties which are not under the direct influence of language norms, i.e. in dialects. When it occurs, the process also solves the problem of reversed order.

<sup>28</sup>This is in accordance with Baotić's (1985: 371) observations on the construction in question in Northern Bosnia. It is attested by him in local idioms of the *Slavonski* and *Srednjobosanski* dialects.



other dominant languages. Ivić's example (25) comes from the idiom of Galipolje Serbs.

- (25) Ako *je se* uženila [...]
   
if be.3SG REFL marry.PTCP.SG.F
   
'If she got married [...]' (Galipolje idiom; Ivić 1957: 395)

Ivić (1957: 395) claims that this kind of reversed order is far more common in BCS dialects than the one presented below in examples (33)–(36), where the verbal CL *je* precedes pronominal CLs. In Ivić's opinion the *je se* order developed as a consequence of haplology of unlikes, which first resulted in *se je > sē* (Ivić 1957: 395). Afterwards, the verbal CL *je* was restored in front of the reflexive CL *se* by analogy with other verbal CLs (cf. Ivić 1957: 395).

However, the divergent position of the CL *se* in the CL cluster is not always connected to its position relative to the verbal CL *je*. Okuka (2008: 91) reports cases of *se li* (26) and *se ga* (27) cluster strings in the Neo-Štokavian *Lički* subdialect.

- (26) oće *se li* nēde mūci potōpiti
   
FUT.3SG REFL Q somewhere can.INF flood.INF
   
'will flooding be possible somewhere'
   
(Istočnohercegovački; Okuka 2008: 91)

- (27) jâ *se ga* sjěćâm
   
I REFL him.GEN remember.1PRS
   
'I remember him'
   
(Istočnohercegovački; Okuka 2008: 91)

The REFL<sub>LEX</sub> *se* can precede a genitive pronominal CL not only in the *Istočnohercegovački*, but also in the *Šumadijsko-vojvođanski* dialect (cf. Nikolić 1966: 280). The order presented in (27) can be found not only in Štokavian, but also in Čakavian dialects. In Lisac (2009: 41) we found a similar example from the *Buzetski* Čakavian dialect (28) with an accusative CL after REFL<sub>LEX</sub> *se*.

- (28) dok *se ga* nē pupije
   
until REFL him.GEN NEG drink.up.2PRS
   
'until it is drunk up' (Buzetski; Lisac 2009: 41)

Reversed order of the verbal CL *bi* and a pronominal CL in the cluster is attested in the Old Štokavian local idiom of Crmnica (29) and in the Neo-Štokavian local idiom of Bitelić (30) (cf. Okuka 2008: 180, Ćurković 2014: 309).<sup>29</sup>

<sup>29</sup>The original version in Okuka (2008) is: "Ne-bik-ti ja to učinijo, pa-da-mi-bi-da iljadu dinara".

7 *Clitics in dialects*

- (29) Ne bik        *ti* ja to učinijo,        pa da *mi*        *bi*  
 NEG COND.1SG you I that do.PTCP.SG.M well that me.DAT COND  
 da                iljadu        dinara.  
 give.PTCP.SG.M thousand dinars  
 ‘I wouldn’t do that, even if you gave me a thousand dinars.’  
 (Zetsko-južnosandžački; Okuka 2008: 180)
- (30) Ōnda *ti*        *bi*        dōli        dōšlo [...].  
 then you.DAT COND down come.PTCP.SG.N  
 ‘Then (they) would come down [...].’        (Zapadni; Ćurković 2014: 309)

The example from the local idiom of Bitelić in (31) indicates that in the *Zapadni* dialect CL order in this kind of cluster does not always differ from the order in standard BCS varieties. Therefore, we conclude that the reversed order is just a possibility and not a rule in these dialects.

- (31) Stārī *bi*        *ti*        ljūdi        ūvečē [...].  
 old COND you.DAT people in.the.evening  
 ‘Old people would in the evening [...].’        (Zapadni; Ćurković 2014: 309)

Pešikan (1965: 209) also reports reversed order of  $REFL_{LEX}$  and conditional auxiliary CLs in the local idiom of Rijeka (32).<sup>30</sup>

- (32) ōna *se*        *bi*        prepāla  
 she REFL COND get.scared.PTCP.SG.F  
 ‘she would get scared’        (Zetsko-južnosandžački; Pešikan 1965: 209)

The verbal CL *je* can appear not only before reflexive CLs (see examples (22)–(25) presented above), but also in front of pronominal CLs (cf. Ivić 1957: 394, Pešikan 1965: 210, Nikolić 1966: 279, Okuka 2008: 256).<sup>31</sup> Examples (33) and (34) below are from the local idioms of Pričinović and Zagrač, while (35) is from the *Svrljiško-zaplanjski* dialect.

<sup>30</sup>He suggests that from the diachronic perspective cases similar to (29) and (32) signal that the conditional auxiliary CL is younger than the pronominal and reflexive CLs. The assumption that the order of CLs in the cluster was influenced by their relative age can be found in diachronic literature (e.g. Grickat 1972: 95, Zimmerling & Kosta 2013: 189) and seems plausible. The relative order of the reflexive CL *se* and the conditional auxiliary CLs attested in (32) was already present in OCS. However, in texts written in the Croatian redaction of Church Slavonic, i.e. the younger variety, the reflexive CL *se* follows the conditional auxiliary CLs (cf. Gadžijeva et al. 2014: 318).

<sup>31</sup>This kind of reversed CL order in which pronominal CLs are preceded by the verbal CL *je* is also attested in colloquial Serbian and in {bs, hr, sr}WaC corpora. For more information see Section 6.4.1.

- (33) *komèsija je me pítala*  
 commission be.SG me.ACC ask.PTCP.SG.F  
 ‘the commission asked me’ (Šumadijsko-vojvođanski; Nikolić 1966: 279)
- (34) *òn je mu røkā*  
 he be.3SG him.DAT tell.PTCP.SG.M  
 ‘he told him’ (Zetsko-južnosandžački; Pešikan 1965: 210)
- (35) *kuj e ga volel*  
 who be.3SG him.ACC love.PTCP.SG.M  
 ‘who loved him’ (Svrlijško-zaplanjski dialect; Okuka 2008: 256)
- (36) *Kò je mu se poklonio.*  
 who be.3SG him.DAT REFL bow.PTCP.SG.M  
 ‘Who bowed to him.’ (Galipolje idiom; Ivić 1957: 394)

Stevanović (1950: 152) claims that pronominal CLs can stand in front of verbal CLs (the present tense of *biti* ‘be’) in the local idiom of Đakovica:

- (37) *Dvá mu su sína žènetà.*  
 two him.DAT be.3PL sons married  
 ‘Two of his sons are married.’  
 (Prizrensko-južnomoravski; Stevanović 1950: 152)

The CLs in example (38) from the local idiom of Vitina found in Peco (2007b: 345) are ordered according to cluster ordering rules of the standard varieties, with one exception – there are two dative CLs in one cluster.

- (38) *Vèlik ti mu je národ dohòdijo, [...].*  
 big you.DAT him.DAT be.3SG people come.PTCP.SG.M  
 ‘A great mass of people came to him, you know [...].’  
 (Istočnohercegovački; Peco 2007b: 345)

According to the cluster ordering rules in standard BCS varieties there is only one slot for dative CLs. In his theoretical work on CLs Bošković (2001: 62) mentions the possibility of two dative CLs in one cluster and claims that when both the ethical and argumental dative are present in a sentence, the former must precede the latter. Peco’s example from the Neo-Štokavian *Istočnohercegovački* dialect seems to nicely support Bošković’s theory.



- (41) dà *jē* je vīdio, zóvnijo bi nās  
 that her.ACC be.3SG see.PTCP.SG.M call.PTCP.SG.M COND us.ACC  
 ‘if he had seen her, he would have called us’  
 (Zapadni dialect; Šimundić 1971: 120)
- (42) Ōn in je je ukrāl  
 he them.DAT her.ACC be.3SG steal.PTCP.SG.M  
 ‘He stole her from them’ (Srednjočakavski; Vranić 2003: 165)

### 7.5.2.3 Haplology of unlikes

As described in Section 7.5.1, in many idioms the reflexive CL *se* and the verbal CL *je* ‘is’ appear in an order which diverges from the one found in standard BCS varieties. In contrast, some idioms such as the local idiom of Nevesinje adopt haplology of unlikes, i.e. the solution used in BCS standard varieties – see example in (43).<sup>34</sup>

- (43) Kad *se* oslobòdila kmetàrija [...].  
 when REFL free.PTCP.SG.F serfs  
 ‘When the serfs freed themselves [...].’  
 (Istočnohercegovački; Peco 2007a: 281)

However, as examples (21)–(25) reveal, haplology of unlikes does not always occur. Moreover, even idioms which belong to the same dialect sometimes differ with respect to this phenomenon: compare examples (43) and (44). In contrast to examples (21)–(25), CLs in example (44) are ordered just like they would be in BCS standard varieties. This is attested in the *Istočnobosanski* subdialect.

- (44) kad *se* je zàratilo  
 when REFL be.3SG start.war.PTCP.SG.N  
 ‘when the war started’ (Istočnohercegovački; Okuka 2008: 77)

## 7.6 Position of the clitic or the clitic cluster

### 7.6.1 Second position

CLs can definitely follow phrases with two content words in Neo-Štokavian dialects. As we pointed out in Section 6.5.4. Second position, second word and DP, this is also the case in standard Bosnian and Serbian, while the Croatian norm

<sup>34</sup>For an explanation of the haplology of unlikes, see Section 2.4.2.2.

## 7 Clitics in dialects

recommends phrase splitting or DP. While browsing dialectological literature we find examples such as those from the local idioms of Studenci (45) and Zvirici (46). However, bear in mind that 2P is not the only option in the *Istočnohercegovački* dialect: phrase splitting is also possible (see examples (59), (60) and (61) below).

- (45) Čítavo pòlje *se* mòre tòpit  
entire field REFL can.3PRS flood.INF  
'The entire field can be flooded' (Istočnohercegovački; Peco 2007b: 341)
- (46) Mòj dèdo *bi* rèkā: [...].  
my grandfather COND say.PTCP.SG.M  
'My grandfather would say: [...]' (Istočnohercegovački; Peco 2007b: 342)

The typical position of CLs in the local idiom of Sarajevo does not involve phrase splitting, as shown in example (47) (Halilović et al. 2009: 62).

- (47) Íêvā cìpela *mi je* malèhna  
left shoe me.DAT be.3SG small  
'my left shoe is small' (Srednjobosanski; Halilović et al. 2009: 62)

### 7.6.2 Delayed placement of clitics

According to Raguž (2016: 273) the Old Štokavian local idiom of the village Bogdanovci does not display the tendency common in standard Croatian to place the CL after the first stressed word. In (48) and (49) there are no barriers which would prevent the reflexive CL *se* from taking 2P. Similar examples are also found in the local idiom of Bizovac (cf. Klaić 1959: 144). Both of these idioms belong to the *Slavonski* dialect.

- (48) Ti nē mōraš *se* ùdat.  
you NEG have.to.2PRS REFL marry.INF  
'You do not have to get married.' (Slavonski; Raguž 2016: 273)
- (49) A za svàtove nīsu pògače *se* pèkle.  
and for wedding NEG.be.3PL breads REFL bake.PTCP.PL.F  
'And for the wedding, breads were not baked.'  
(Slavonski; Raguž 2016: 273)

Examples with DP can be found in other idioms of the *Slavonski* dialect, such as those spoken in South Baranja. However, if initial constituents are heavy, DP is not the only possible position for CLs in this dialect. Namely, in idioms of South

Baranja and in the local idiom of Našice CLs can also follow heavy constituents (cf. Sekereš 1977: 412f, Sekereš 1966: 264).

Aladrović (2011: 165) reports DP as a dialectal trait in the written language of elementary school students from Požega (50). We can thus conclude that DP as a dialectal feature is not present only in the language of the older population.

- (50) ja želim se vratiti  
 I want.1PRS REFL return.INF  
 ‘I want to return’ (Istočnohercegovački; Aladrović 2011: 165)

Nikolić (1966: 279) and Okuka (2008: 136) also report divergence in CL positioning in some idioms of the Neo-Štokavian *Šumadijsko-vojvođanski* dialect. Instead of moving towards the 2P, CLs move towards the end of the sentence, as in the examples from the local idioms of Kolubara (51) and Banat (52). CLs in the Banat area might be influenced by the neighbouring Romanian language. Nevertheless, the example from the Kolubara area shows that peculiarities of CL positioning cannot be ascribed exclusively to the influence of other languages, as the area is placed in the middle of the *Šumadijsko-vojvođanski* dialect area (in central Serbia).

- (51) od one kože prave opanke nam  
 from that leather make.3PRS shoes us.DAT  
 ‘from that leather, they make shoes for us’  
 (Šumadijsko-vojvođanski; Okuka 2008: 136)

- (52) Šta vi sad se oblačite?  
 what you now REFL get.dressed.2PRS  
 ‘Why are you getting dressed now?’  
 (Šumadijsko-vojvođanski; Okuka 2008: 136)

Although Nikolić (1966: 279) claims that CLs in the *Šumadijsko-vojvođanski* dialect often split semantically tightly bound phrases (see examples (64) and (67) in Section 7.6.3), examples with DP are easily found in his material, such as (53) from the local idiom of Pričinović.

- (53) a tāj domàćin narédio im  
 and that host order.PTCP.SG.M them.DAT  
 ‘and that host ordered them’ (Šumadijsko-vojvođanski; Nikolić 1966: 280)

As we can see, the phenomenon of DP occurs in various dialects. In Okuka’s (2008: 77) words: in the *Istočnobosanski* subdialect CLs shift backwards. The examples of DP presented below are from the *Istočnobosanski* subdialect (54) and from the local idiom of Borač (55).

## 7 Clitics in dialects

- (54) sva ogubala se  
entire become.leprous.PTCP.SG.F REFL  
'she became completely covered in warts'  
(Istočnohercegovački; Okuka 2008: 77)
- (55) Prije trides godina Ahmet Cik je kôsio na taj dan.  
before thirty years Ahmet Cik be.3SG scythe.PTCP.SG.F on that day  
'Thirty years ago on that day Ahmet Cik was mowing grass.'  
(Istočnohercegovački; Peco 2007a: 285)

In the idiom of Galipolje Serbs spoken in the Macedonian city Pehčevo, CLs have to follow the negative present tense form of *biti*, and therefore examples of DP such as the one in (56) below can occur.

- (56) Takò on nijè je poslušavo.  
so he NEG.be.3SG her.ACC listen.PTCP.SG.M  
'So he did not listen to her.'  
(Galipolje idiom; Ivić 1957: 395)

Many examples with DP concur with the use of *da* particle, such as those from the local idiom of Pričinović presented in (57) and (58).<sup>35</sup>

- (57) òće d' ubije me  
want.3PRS that kill.3PRS me.ACC  
'he wants to kill me'  
(Šumadijsko-vojvođanski; Nikolić 1966: 280)
- (58) nêće da donèsē ga  
NEG.want.3PRS that bring.3PRS him.ACC  
'he does not want to bring him'  
(Šumadijsko-vojvođanski; Nikolić 1966: 280)

### 7.6.3 Phrase splitting

Phrase splitting is attested in both Old and Neo-Štokavian dialects, and even in some Kajkavian dialects. In most cases, verbal CLs split attributes from their head nouns, but there are some examples in which dative pronominal CLs cause phrase splitting.<sup>36</sup>

<sup>35</sup>In this respect dialects definitely diverge from standard Bosnian and Serbian, in which CLs must follow any complementisers. Moreover, those examples speak strongly against claims that the only possible and correct position of CLs is directly after the *da* particle and are in accordance with the results of our study on CC out of *da*<sub>2</sub>-complements. For more information see Sections 6.5.3 and 13.3.

<sup>36</sup>In this respect there are not many differences between BCS standard varieties and dialects. For more information, see Section 6.5.5.



Examples in which the attribute and its noun are split can be found in the Old Štokavian *Slavonski* dialect (cf. Sekereš 1977: 412f, Golić 1993: 103) and in the Neo-Štokavian *Istočnohercegovački* dialect (cf. Peco 2007a: 283). For the latter, see the example from the local idiom of Borač presented in (59).<sup>37</sup> In this idiom we also find examples in which adverbs are split from a noun in the genitive (60).<sup>38</sup> Note that we find similar examples (37) with a split quantified phrase (*dva sina* ‘two sons’) in the Middle Štokavian *Prizrensko-južnomoravski* dialect.

- (59) *Övā je država plamírala da prāvī škôrlu.*  
 this be.3SG country plan.PTCP.SG.F that make.3PRS school.ACC  
 ‘This country planned to build a school.’  
 (Istočnohercegovački; Peco 2007a: 283)

- (60) *Víše smo hájra vídeli [...].*  
 more be.1PL benefit see.PTCP.PL.M  
 ‘We saw more benefit [...].’ (Istočnohercegovački; Peco 2007a: 283)

Furthermore, in the *Istočnohercegovački* dialect CLs can be inserted between parts of compound pronouns like *tko god* (61).<sup>39</sup>

- (61) *kò se god bòji, slàbo će próci*  
 who REFL ever fear.3PRS poorly FUT.3SG pass.INF  
 ‘whoever is afraid will fare poorly’  
 (Istočnohercegovački; Sekereš 1977: 338)

In contrast, in idioms of Baranja (*Slavonski* dialect) and of Bačka (*Zapadni* dialect) CLs normally follow compound pronouns (cf. Sekereš 1977: 340, 413).

In the local idiom spoken in the Neretva valley CLs can split prepositional phrases, like in example (62).<sup>40</sup>

- (62) *[...] slàdā je od šèčera bíla*  
 sweeter be.3SG than sugar be.PTCP.SG.F  
 ‘[...] she was sweeter than sugar’  
 (Istočnohercegovački; Vukša Nahod 2014: 188)

<sup>37</sup>This kind of split phrase is found to be acceptable in both standard Croatian and standard Serbian. For more information, see Section 6.5.5.

<sup>38</sup>Examples in which a noun in the genitive case is split from its head are controversially discussed in the theoretical syntactic literature, see Section 6.5.5.

<sup>39</sup>Although this kind of split phrase is found to be acceptable in both standard Croatian and standard Serbian, it is considered to be quite uncommon in both – see Section 6.5.5. Moreover, it is attested in the corpus Bosnian Interviews.

<sup>40</sup>This phenomenon is rather controversial in the theoretical literature on CL placement, see Section 6.5.5.

As we pointed out in Section 6.5.5 splitting a forename from a last name is not recommended in standard Serbian. However, in dialects CLs can split the tightly bound forename from the surname: see example (63) from the local idiom of Kolašin and (64) from the local idiom of Banovo Polje (cf. Okuka 2008: 67, Nikolić 1966: 279). Moreover, we would like to emphasise that the latter example is attested in a dialect spoken on Serbian territory. In the previous century Pešikan (1965: 209) reported the same for *Starocrnogorski* idioms.<sup>41</sup>

- (63) Vuk *mi je* Šćepanović reko.  
 Vuk me.DAT be.3SG Šćepanović say.PTCP.SG.M  
 ‘Vuk Šćepanović told me.’ (Istočnohercegovački; Okuka 2008: 67)
- (64) Milan *mi je* Pëćić bïo kao ùpravnik bólnice.  
 Milan me.DAT be.3SG Pecić be.PTCP.SG.M like director hospital  
 ‘Milan Pecić was like a hospital director to me.’  
 (Šumadijsko-vojvođanski; Nikolić 1966: 279)

Here we would like to point out that in both examples provided above, it is a CL cluster that splits the phrase, which theoretical syntacticians Progovac (1996) and Radanović-Kocić (1996) strongly dislike (for more information, see Section 6.5.5).

In the *Istočnohercegovački* and *Šumadijsko-vojvođanski* dialects one more rather controversial structure is attested. Namely, Okuka (2008: 67, 74) and Pešikan (1965: 209) report cases in which verbal CLs split conjoined phrases: see examples (65)–(67) presented below.<sup>42</sup>

- (65) Ja *smo i* on zajedno utekli  
 I be.1PL and he together escape.PTCP.PL.M  
 ‘He and I escaped together’ (Istočnohercegovački; Okuka 2008: 67)
- (66) Mi *ćemo i* oni to zajedno  
 we FUT.1SG and they that together  
 ‘We and they are going (to do) that together’  
 (Istočnohercegovački; Okuka 2008: 67)

According to Okuka (2008: 67), this construction is widespread in the *Istočnohercegovački* dialect. This, however, is not the only Neo-Štokavian dialect in

<sup>41</sup>According to dialectological maps some *Starocrnogorski* idioms are part of Neo-Štokavian *Istočnohercegovački*, and others of the Old Štokavian *Zetsko-južnosandžački* dialect.

<sup>42</sup>For controversial discussion on this structure from the theoretical point of view, see Section 6.5.5.

which the construction in question is attested. Example (67) below is from the local idiom of Pričinović, i.e. the *Šumadijsko-vojvođanski* dialect.

- (67) *jâ smo i žena sâmi*  
 I be.1PL and woman alone  
 ‘my wife and I are alone’ (Šumadijsko-vojvođanski; Nikolić 1966: 279)

In the Neo-Štokavian *Banatsko-pomoriški* subdialect, possessive CLs lean on attributes and split phrases (68) (cf. Okuka 2008: 148).

- (68) *pòkōjni mi muž*  
 deceased me.DAT husband  
 ‘my deceased husband’ (Šumadijsko-vojvođanski; Okuka 2008: 148)

It seems that language contact does not negatively influence phrase splitting. In the idiom of Galipolje Serbs CL clusters can be inserted between the attribute and the noun (107).<sup>43</sup> As we can see from these two examples, phrase splitting is attested even in idioms which are in direct contact with other languages which have CLs: the *Banatsko-pomoriški* subdialect is spoken on the Romanian border and the idiom of Galipolje Serbs is in direct contact with Macedonian. In addition, phrase splitting is attested even in the Old Štokavian local idiom of Vršenda, which is in direct language contact with Hungarian (cf. Gorjanac 2011: 111f): see examples (69) and (70) below.

- (69) *Jedna je krava bila za čitavu družinu.*  
 one be.3SG cow be.PTCP.SG.F for entire group  
 ‘There was one cow for the entire group.’ (Slavonski; Gorjanac 2011: 111)
- (70) *Naša je stara imala petero dice [...].*  
 our be.3SG old have.PTCP.SG.F five children  
 ‘Our mother had five children [...].’ (Slavonski; Gorjanac 2011: 112)

As we already mentioned, phrase splitting is attested even in Kajkavian dialects, in which both auxiliary CLs (71) and pronominal CLs (72) can be placed between an attribute and its noun. The former example is from the local idiom of Turropolje, while the latter is from the *Gornjolonjski* dialect.

- (71) *Mōja je mǎma govoriġla [...].*  
 my be.3SG mother speak.PTCP.SG.F  
 ‘My mother was saying [...].’ (Turopoljski; Lončarić 1994: 139)

<sup>43</sup> A few pages later, Ivić (1957: 397) even claims that in the Galipoljski idiom CLs usually follow the first member of the noun phrase.

## 7 Clitics in dialects

- (72) ova *mi* peč nigdar neće dobre goreti.  
this me.DAT stove never NEG.FUT.3SG good burn.INF  
'This stove of mine will never burn well.'  
(Gornjolonjski; Brlobaš & Lončarić 2012: 242)

### 7.6.4 Clitic first (1P)

As described in Section 6.5 CLs cannot have sentence-initial position in any of the BCS standard varieties, which are all based on the Neo-Štokavian *Istočnohercegovački* dialect. Furthermore, in BCS standard varieties CLs cannot follow the conjunctions *i* and *a* 'and'. However, in the dialectological literature we find examples from Neo-Štokavian dialects showing different behaviour:

- (73) *jâ i smo* tî ôsam dánâ da krûva nè vidi  
I and be.1PL you eight days that bread NEG see.3PRS  
'it happened that you and I did not see bread for eight days'  
(Istočnohercegovački; Okuka 2008: 91)

In example (73) recorded in the *Lički* subdialect of the Neo-Štokavian *Istočnohercegovački* dialect, the verbal CL *smo* 'are' follows the conjunction *i*.<sup>44</sup> We may speculate that this peculiar CL order could have been triggered by the neighbouring *Srednjočakavski* dialect.<sup>45</sup> Similar examples are found in the Neo-Štokavian *Šumadijsko-vojvođanski* dialect: see (74). According to Okuka (2008: 136) in this dialect enclitics can turn into proclitics, and often follow the conjunction *i*.

- (74) Došo i mi pusti *ga*, i *su ga*  
come.PTCP.SG.M and we let.INF him.ACC and be.3PL him.ACC  
strélali.  
shoot.PTCP.PL.M  
'He came and we let him go, and they shot him.'  
(Šumadijsko-vojvođanski; Okuka 2008: 136)

We can assume that the trait in question is spreading from the eastern edge of *Šumadijsko-vojvođanski* to other parts. The reason for such an assumption is

<sup>44</sup>Readers have to bear in mind that this is not an example of conjoined phrase splitting. A CL which splits a conjoined phrase follows the first element of that phrase. In this example the CL follows the unaccented coordinative conjunction *i*.

<sup>45</sup>According to Lisac (2009: 113), in the *Srednjočakavski* subdialect CLs can take 1P in a sentence, as is usual for Čakavian dialects. Furthermore, CLs can even bear stress in this dialect (cf. Lisac 2009).

Okuka's (2008: 148) report that the *Banatsko-pomoriški* subdialect shows Romanian influences, one of which is proclitisation of CLs. He provides the following example (75) with an auxiliary CL which takes absolute initial position in the sentence.

- (75) *Su*    *bíli*                    *u* *célo* *sèlo*.  
 be.3PL be.PTCP.PL.M in entire village  
 'They were in the entire village.'  
 (Šumadijsko-vojvođanski; Okuka 2008: 148)

Further in the east, in the local idiom of Rekaš (*Kosovsko-resavski* dialect) in Romania, CLs can take 1P: see example (76). However, such behaviour is not the rule, since they can appear in 2P as well (cf. Vulić 2009: 171).

- (76) *se*    *moli*            *Boga*  
 REFL pray.3PRS God  
 'he is praying to God'                    (Kosovsko-resavski; Vulić 2009: 171)

1P is reported for verbal (77), pronominal (78) and reflexive CLs (79) in the local idiom of Đakovica (*Prizrensko-južnomoravski* dialect). Stevanović (1950: 152) claims that dative pronominal CLs are more common in 1P than other pronominal CLs.

- (77) *Su*    *bíle*                    *pét* *bráča*.  
 be.3PL be.PTCP.PL.F five brothers  
 'They were five brothers.'  
 (Prizrensko-južnomoravski; Stevanović 1950: 152)

- (78) *Mu*    *víkam*    *mojému* *strícu*.  
 him.DAT yell.1PRS mine    uncle  
 'I am telling my uncle.' (Prizrensko-južnomoravski; Stevanović 1950: 152)

- (79) *Se*    *zvào*                    *táko*.  
 REFL call.PTCP.SG.M like.that  
 'He was called that.'                    (Prizrensko-južnomoravski; Stevanović 1950: 152)

Karaševó-Croats, who live in seven villages in the Romanian part of Banat and speak the *Timočko-lužnički* Torlac dialect, place CLs in the proclitic position under the influence of Romanian (cf. Lisac 2003: 147) – see the example presented in (80).

- (80) *smo*    *se*    *nazdravili*  
 be.1PL REFL toast.PTCP.PL.M  
 'we made a toast'                    (Timočko-lužnički dialect; Lisac 2003: 147)

## 7 Clitics in dialects

Language contact with Macedonian and Bulgarian is probably also the reason why speakers of the *Prizrensko-južnomoravski* and *Timočko-lužnički* dialects can place CLs in the sentence-initial position (cf. Okuka 2008: 239–267). The example in (81) below is from *Timočko-lužnički*.

- (81) *Če dojde li?*  
FUT.3SG come.3PRS Q  
'Will he come?' (Timočko-lužnički; Okuka 2008: 267)

Although Lisac (2003: 27) claims that Štokavian's interesting feature, turning enclitics into proclitics, is due to language contact, it seems that at least some of the examples with CLs in the initial position or after conjunctions such as *i* and *a* cannot be explained through the influence of other languages. Brozović (2007: 150) provides examples from the local idiom spoken in the Fojnica valley (*Srednjobosanski* dialect) in which CLs do not follow the first stressed word, such as those quoted in (82) and (83).

- (82) *se mōžēš prilādit*  
REFL can.2PRS catch cold  
'you can catch a cold' (Srednjobosanski; Brozović 2007: 150)
- (83) *i je rēkō*  
and be.3SG say.PTCP.SG.M  
'and he told them' (Srednjobosanski; Brozović 2007: 150)

The *Srednjobosanski* dialect borders neither non-Štokavian dialects nor any other languages. Therefore, we assume that its atypical positioning is language-internally caused and may have something to do with the fact that it is an Old Štokavian dialect.

### 7.6.5 Endoclititics

ENDOCLITICS (a term proposed by Radanović-Kocić 1988) are a phenomenon very similar to phrase splitting, involving the insertion of a CL in a morphological word, i.e. between affix and stem, like in examples (84)–(86) from the local idioms of Derventa, Neretva valley, and Zmijanje (cf. Lisac 2003: 58, Okuka 2008: 67, Vukša Nahod 2014: 195). In addition, the occurrence of endoclititics is reported by Nikolić (1966: 279) and Halilović (2005: 23) for the *Šumadijsko-vojvođanski* and *Srednjobosanski* dialects.

- (84) *nâj mi je drăži*  
 most me.DAT be.3SG dearer  
 ‘he is my dearest’ (Zapadni; Lisac 2003: 58)
- (85) [...] *a rîbu nâj san višê ùvatijâ jèdnu nôc [...]*.  
 and fish most be.1SG more catch.PTCP.SG.M one night  
 ‘[...] and I caught the most fish one night [...].’  
 (Istočnohercegovački; Vukša Nahod 2014: 195)
- (86) *naj bi bolje uspjevo krompijer*  
 most COND better succeed.PTCP.SG.M potato  
 ‘the potato would succeed the best’  
 (Istočnohercegovački; Okuka 2008: 67)

As we can see, in Neo-Štokavian dialects superlative forms can be split not only by a verbal CL as in (85) and (86), but also by a CL cluster as in example (84). Furthermore, in the local idiom of Retkovci a pronominal CL can be inserted into negated forms in the present tense of the verb *biti* (87).

- (87) *Nî mi je znô kâst.*  
 NEG me.DAT be.3SG know.PTCP.SG.M tell.INF  
 ‘He was not able to tell me.’ (Slavonski; Kolenić & Bilić 2004: 18)

Moreover, it seems that endoclititics exist in Istrian Čakavian dialects as well.<sup>46</sup> Kalsbeek (2003: 107) documented cases of a CL inserted between parts of a negative imperative, like in example (88).

- (88) *Ne ga muōj zvaljât!*  
 NEG him.ACC IMP.2SG dirty.INF  
 ‘Don’t dirty it!’ (Čakavian; Kalsbeek 2003: 107)

Finally, in Istrian Čakavian the CL *li* can be placed between the stem and the ending of the future auxiliary (cf. Kalsbeek 2003: 107), like in the example presented in (89).

- (89) *Ćeliš jùtre rivât tò storìt?*  
 FUT.Q.2SG tomorrow manage.INF that get.done.INF  
 ‘Will you be able to get that done tomorrow?’  
 (Čakavian; Kalsbeek 2003: 107)

<sup>46</sup>Since Kalsbeek does not provide additional information about his data, we could not determine exactly from which Čakavian dialect those examples originated.

## 7 Clitics in dialects

Similar examples with the interrogative CL *li* inserted between parts of the verb *htjeti* are found in the Neo-Štokavian local idiom of Imotski – see example (90) below.

- (90) [...] òće *li* mo lètriku.  
want Q 1SG electricity  
‘[...] do we want electricity.’ (Zapadni; Šimundić 1971: 212)

### 7.7 Clitic climbing

CC is not discussed in the dialectological literature, but as the central part (Part III) of this monograph is dedicated to this topic, we tried to find some examples of CC in the transcripts from dialectological literature.<sup>47</sup> Although most of the transcripts include mainly simple structures, we found some examples of CTPs and their infinitive complements.<sup>48</sup> In example (91) from the local idiom of Divin, the archaic pronominal dative CL *vi* climbs over the raising matrix CTP *moći*. Furthermore, we found CC of the pronominal CL *me* (92) in *Sremski* idioms and CC of the REFL<sub>LEX</sub> *se* in the *Lički* subdialect (93).

- (91) [...] ali *vi*<sub>2</sub> ne mògu<sub>1</sub> dat<sub>2</sub> òdgovōr [...].  
but you.DAT NEG can.1PRS give.INF answer  
‘[...] but I cannot give you the answer [...]’  
(Istočnohercegovački; Peco 2007a: 287)
- (92) kad *me*<sub>2</sub> stànū<sub>1</sub> grčevi vātati<sub>2</sub>  
when me.ACC start.3PRS cramps catch.INF  
‘when I start getting cramps’ (Šumadijsko-vojvođanski; Nikolić 1964: 368)
- (93) ako *se*<sub>2</sub> mīslī<sub>1</sub> žènit<sub>2</sub>  
if REFL think.3PRS marry.INF  
‘if he plans to get married’ (Istočnohercegovački; Okuka 2008: 90)

While the CTPs in examples (91) and (92) are raising predicates (modal and phasal), the CTP in example (93) is a subject control predicate.

CC is found in Old Štokavian *Slavonski* dialect as well (see (87) above). In this example, the pronominal CL *mi* climbs out of a subject-controlled infinitive and splits the negative present tense form of *biti* ‘be’. Besides examples with CC, we also find examples without CC. These come from the local idiom spoken in the

<sup>47</sup>See Section 2.4.4 and Chapter 10 for a basic explanation of the phenomenon of Clitic Climbing.

<sup>48</sup>See Section 2.5.1 for basic information on CTP types.





## 7 Clitics in dialects

- (98) nìje        ga<sub>2</sub>        ni    stìgō<sub>1</sub>        da    vīdī<sub>2</sub>  
NEG.be.3SG him.ACC NEG get.PTCP.SG.M that see.3PRS  
'he did not even get to see him'  
(Šumadijsko-vojvođanski; Nikolić 1964: 368)
- (99) kako vas<sub>2</sub>    je        jèdān tèo<sub>1</sub>        da    túčē<sub>2</sub>  
how you.ACC be.3SG one    want.PTCP.SG.M that beat.3SG  
'how one of them wanted to beat you'  
(Šumadijsko-vojvođanski; Nikolić 1964: 368)

### 7.8 Diaclisis

We find examples of diaclisis in several dialects.<sup>51</sup> Examples (100) and (101) are from the *Zapadni* dialect.

- (100) [...] i    ònda ka    se    dòbro j<sup>e</sup>        svârī [...].  
and then when REFL good her.ACC cook.3PRS  
'[...] and then when it is cooked well [...].'  
(Zapadni; Ćurković 2014: 285)
- (101) [...] kad    b<sup>i</sup>        ònī stârci    se    skūpili        ùv<sup>e</sup>če [...].  
when COND.3PL those old.men REFL gather.PTCP.PL.M in.the.evening  
'[...] when those old men would gather in the evening [...].'  
(Zapadni; Ćurković 2014: 320)

Nevertheless, we must point out that diaclisis is not the rule in the *Zapadni* dialect, since examples with CL clusters like the one presented in (102) are also attested.

- (102) [...] Òbūkl<sup>i</sup>        b        se    ù onū svàsku ròbu [...].  
dress.PTCP.PL.M COND.3PL REFL in those wedding clothes  
'[...] They would dress in those wedding clothes [...].'  
(Zapadni; Ćurković 2014: 284)

Diaclisis is also attested in the local idiom of Pričinović and in the local idiom of Babina Greda:

- (103) Pa    je        l    múzē        se?  
well be.3SG Q milk.3PRS REFL  
'Well, is it being milked?' (Šumadijsko-vojvođanski; Nikolić 1966: 279)

<sup>51</sup>For a definition of diaclisis, see Section 2.4.5.

- (104) U dva sata *smo* noću *se* dizali.  
 at two o'clock be.1PL at.night REFL get.up.PTCP.PL.M  
 'At two o'clock at night we got up.' (Slavonski; Farkaš & Babić 2011: 149)

## 7.9 Clitic doubling

CL doubling is considered to be a feature specific to the Balkan languages; it involves structures where pronominal CLs double overtly expressed direct or indirect objects.<sup>52</sup> It is common to all Slavonic and non-Slavonic South-Eastern Balkan languages (Stevanović 1950: 114). Mišeska Tomić (2006: 239, 2008: 426) claims that pronominal CLs do not double direct or indirect objects in standard varieties of BCS. The same observation applies to the Northern Serbian dialects. Conversely, pronouns are CL-doubled in all the South-eastern Serbian dialects (Mišeska Tomić 2008: 463). Whereas all indirect objects are regularly CL-doubled in the western periphery of the south-eastern Serbian dialect area, in the south-easternmost parts of the area both direct and indirect lexical objects are only optionally CL-doubled (Mišeska Tomić 2008: 463).

Stevanović (1950: 113f) provides examples of CL doubling in the *Prizrensko-južnomoravski* dialect and claims that such examples appear in the *Kosovsko-resavski* dialect too. One of his examples from the local idiom of Đakovica is presented in (105). Barjaktarević (1966: 112) finds CL doubling in the *Zetsko-južnosandžački* dialect: example (106) is from the local idiom of Trnava. Furthermore, Ivić (1957: 356) claims that CL doubling can be found in the idiom of Galipolje Serbs (107).

- (105) Dáj *mu* [njèmu].  
 give.IMP him.DAT him.DAT  
 'Give him.' (Prizrensko-južnomoravski; Stevanović 1950: 113)
- (106) [Mène] *me* zovê.  
 me.ACC me.ACC call.PRS  
 'He calls me.' (Zetsko-južnosandžački dialect; Barjaktarević 1966: 112)
- (107) Uvâ *mi* *je* divôjka poznâta [minè].  
 this me.DAT be.3SG girl familiar me.DAT  
 'This girl is familiar to me.' (Galipolje idiom; Ivić 1957: 356)

<sup>52</sup>We did not include CL doubling among other parameters in Chapter 2 because it is not attested in the standard varieties.

Nonetheless, CL doubling is not obligatory in the latter idiom, i.e. constructions without CL doubling are used as well (Ivić 1957: 357). In addition, Ivić (1957: 357) argues that CL doubling is far more common in the neighbouring *Prizrensko-južnomoravski* dialect than in the idiom of Galipolje Serbs. He assumes that CL doubling in the idiom of Galipolje Serbs is not a result of direct Macedonian influence. He finds arguments for this assumption in differing word order (cf. Ivić 1957: 357). Ivić (1957: 357) believes that CL doubling was already present in this idiom when Galipolje Serbs lived in Barjamič and that it was the result of Greek influence. CL doubling is also attested among the Croatian population in Janjevo and Letnica (108), who speak the *Prizrensko-južnomoravski* Torlac dialect, and in the *Moliški* dialect (109).<sup>53</sup>

- (108) on *me* [mene] videja  
 he me.ACC me.ACC see.PTCP.SG.M  
 ‘he saw me’ (Prizrensko-južnomoravski; Lisac 2003: 149)
- (109) [njega] su *ga* ubili  
 him.ACC be.3PL him.ACC kill.PTCP.PL.M  
 ‘they killed him’ (Moliški; Lisac 2003: 59)

## 7.10 Summary

### 7.10.1 Inventory

We notice a considerable number of forms for pronominal CLs. First, we see that in many idioms of both Old and Neo-Štokavian dialects both forms *ju* and *je* are attested for the feminine singular accusative. We come across dialects which use *je* exclusively, *ju* exclusively, and those which use both. Second, somewhat unexpectedly we find a large number of varying forms of other pronominal CLs which have not found their way into any of the three standard varieties. In this respect the *Prizrensko-južnomoravski* dialect spoken in Southern Serbia and Kosovo turns out to be the most varied as it shows the greatest number of forms (i.e. four forms for her.DAT, six forms for her.ACC, six forms for they.ACC and seven forms for they.DAT). Without going into much detail, we observe an uneven distribution of variation according to the person and number categories. Namely, dialects tend to show more variants for pronouns of the third person plural than of first and second person plural. In contrast, it seems that pronominal forms for the first

<sup>53</sup>There are several theories about the origins of Croatian speakers from Molise. However, all of the theories agree that they most probably came to Molise from Štokavian territory.

and second person singular in dialects do not differ much from forms attested in standard BCS varieties.

As mentioned in Section 6.3.3 the standard varieties differ with respect to the reflexive CL *si*. However, our dialectological overview shows that this form is found not only on Croatian, but also on Bosnian and Serbian language territory: it is attested in a scattered area comprising some idioms of Western Herzegovina, Northern Bosnia, South Eastern Serbia and Montenegro (*Zapadni, Istočnohercegovački, Svrlijsko-zaplanjski, Prizrensko-južnomoravski, Slavonski* and *Zetsko-južnosandžački* dialects). Furthermore, this form is also present in spoken Bosnian – for more details, see Section 8.7.4.

We do not find a great deal of variation as to the inventory of verbal CLs. As in the spoken varieties, in many dialects (e.g. *Istočnohercegovački, Zapadni, Šumadijsko-vojvodanski, Slavonski* and *Kosovsko-resavski*) the conditional auxiliary form *bi* is used for all persons.

### 7.10.2 Internal organisation of the clitic cluster

In many dialects we find non-standard order in the CL cluster. The most common divergent pattern concerns the order of the reflexive *se* and its position relative to the verbal CL *je*, conditional auxiliary CL *bi*, the polar question marker CL *li* and pronominal CLs. Furthermore, in some Serbian dialects (e.g. *Šumadijsko-vojvodanski, Zetsko-južnosandžački, Svrlijsko-zaplanjski* and the idiom of Galipolje Serbs) the verbal CL *je* appears in front of pronominal CLs just like all other verbal CLs. Here the CL cluster contains a single slot for all verbal CLs and is thus simpler than in the standard languages. In one idiom belonging to the *Prizrensko-južnomoravski* dialect pronominal CLs can stand in front of verbal CLs (the present tense of *biti*) and in the *Zapadni* and *Zetsko-južnosandžački* dialects pronominal CLs can stand before conditional auxiliary CLs. In addition, we would like to point out that although in many dialects some kind of non-standard CL order in the CL cluster is attested, sometimes a CL order which does not diverge from the standard is attested besides that “reversed” order.

Dialects present a varied picture of the use of the pronominal CL *je* (third person singular feminine accusative CL) and the homophone verbal CL *je* (present tense third person singular of *biti* ‘be’). Some local idioms (e.g. of Imotski, Tuholj and Pag) do not exhibit repeated morph constraint, i.e. allow the repetition of *je*, while others (e.g. of Bitelić) use haplology. The same variability is found in the case of the co-occurrence of *se je* versus haplology of unlikes.

### 7.10.3 Position of the clitic or the clitic cluster

Unlike in standard varieties of BCS, in dialects CLs can take the sentence-initial position, and they can follow the conjunctions *i* and *a* ‘and’. Such occurrences are mainly attested in dialects neighbouring with varieties which do allow 1P like Čakavian or the Romanian language, but as examples from the *Srednjobosanski* dialect indicate, not all 1P occurrences can be ascribed to language contact. Similarly, DP is a relatively widespread feature found not only in contact varieties. In this respect dialects definitely differ from the standard Bosnian and Serbian varieties since in the former, CLs do not always follow the *da*-complementiser. As those examples show, placement of CLs directly after the *da*-complementiser is not the only correct possibility, which is in accordance with the results of our study on CC out of *da*<sub>2</sub>-complements (for more information see Sections 6.5.3 and 13.3).

A further finding concerns phrase splitting, which is attested in both Old and Neo-Štokavian dialects. In most cases attributes are split from their nouns by a verbal CL, but there are some examples with pronominal CLs in the dative. Many types of phrase splitting attested in dialects (e.g. splitting of a prepositional phrase, conjoined phrase, quantificational phrase, forename and surname) are controversially discussed in the theoretical syntactic literature. In addition, in dialects we found examples of CL clusters splitting phrases, which theoretical syntacticians Progovac (1996) and Radanović-Kocić (1996) judge to be unacceptable (for more information see Section 6.5.5).<sup>54,55</sup> Moreover, we even came across one type of split not attested in the standard languages: endoclitics, i.e. CLs that split one morphological word form.

It is interesting to note that we do find single examples of diaclisis. Due to the small number of instances we cannot draw any further conclusions. Although

<sup>54</sup>This feature is detectable not only from dialectological data but also from the corpus of Spoken Bosnian *Bosnian Interviews*, see Section 8.9.5.1

<sup>55</sup>We do not claim that dialectological data is superior to the data of Progovac (1996) and Radanović-Kocić (1996) but rather that it is different. Namely, we point out the differences between the varieties described in works of a theoretical character and works on dialects. These differences indicate that the theoretical models of BCS CLs most likely do not hold for dialectological data. Nevertheless, the fact that formal theories cannot account for dialectological data does not make such data inferior to or less valid than data provided by formal linguists. Dialectological data is valid for itself, that is, for varieties classified as dialects. Moreover, the fact that dialectological data is usually collected to describe phonetic and phonological properties, with only a superficial interest in morphology and practically none in syntax, speaks for itself. In addition, the fieldwork is usually conducted by descriptive linguists. We believe that in a way, the latter two facts ensure that there is no theoretical agenda that could have distorted the dialectological data in any way.

we have only several examples of CC, we can claim that it is attested not only from infinitive complements, but in the Serbian *Šumadijsko-vojvodanski* dialect also from *da*<sub>2</sub>-complements. These findings are in accordance with the results of our corpus study presented in Chapter 13.

Finally, we can speculate that the *Prizrensko-južnomoravski* dialect has a different CL system than the majority varieties of BCS because it has not only differing ordering patterns but also the possibility of regular 1P and uses CL doubling.





# 8 Clitics in a corpus of a spoken variety (Bosnian)

## 8.1 Introduction

This chapter contains a pilot study on the usage of CLs in a spoken variety of BCS. This topic has so far remained untouched, probably due to the lack of good spoken data. On the basis of a corpus of Bosnian interviews, we study the inventory and, in particular, the internal organisation of the CL cluster and the position of the CL or cluster. First, we are interested in the inventory of CLs and the types of simple and mixed clusters found in this variety. Second, we would like to give a data-driven account of CL placement. In this regard we inspect the heaviness of the constituents preceding the CLs.

As this chapter is dedicated to spoken language, we annotate the data with respect to syntactic features typical of spoken language which complicate determination of clause boundaries, such as disfluency phenomena, right dislocation, and others. In the final step, we thus look for potential correlation between the position of the CL or CL cluster and these syntactic structures.

This chapter has the following structure: in Section 8.2 we start with a short overview of the state of the art concerning CLs in spoken BCS. In Section 8.3 we formulate our research questions. Then in Section 8.4 we describe the analysed Corpus of Bosnian Interviews (for additional information see also Chapter Overview of Corpora available for BCS). The general principles of spoken language analysis are discussed in Section 8.5 This feeds into our data preparation, including our annotation scheme, as presented in Section 8.6. The results of our data-driven study are discussed in the order that we used in Chapters 2, 6 and 7. We focus only on those parameters of variation for which we had enough data. The inventory of CLs and attested CL clusters is presented in Section 8.7. Section 8.8 is devoted to internal organisation of CL clusters, while the positioning of single CLs and CL clusters is the focus of Section 8.9. This section is followed by Section 8.10 on diacrisis. In Section 8.11 we discuss the impact of certain syntactic structures on CL positioning. The final Section 8.12 contains a summary of the findings.

## 8.2 State of the art: Clitics in spoken BCS

We are aware of the peculiarities of the syntax of spoken language, where intonation plays a crucial role and where syntactic features diverging from written language may be found. Unfortunately, we are confronted with the fact that beyond dialectology (see Section 7.3), spoken BCS is seriously understudied. What we do find in the literature, however, are some scattered conjectures based on linguists' intuition. Here it is important to underline that we do not have any theoretical basis specifically for Bosnian; the claims below were made for the Croatian spoken variety.

Silić was the first to touch upon the differences in CL placement between the spoken and written Croatian varieties. He claimed that the written variety is subject to rhythmic rules, and consequently CLs should follow the first stressed word or be inserted into the first phrase (cf. Silić 1978: 391). This means that phrase splitting is not only a norm, but also natural and expected in the written Croatian variety (cf. Silić 2006: 225).

In contrast, CL placement in the spoken variety is claimed to be freer than in written language, since the spoken variety is governed by "logical factors" (cf. Silić 1978: 391, 1984: 28).<sup>1</sup> Furthermore, Silić (2006: 225) argues that CLs after (heavy) phrases are common and natural in the spoken variety and in some registers which are similar to the spoken variety. Alexander (2009: 63) also claims that placement of CLs after (heavy) phrases is within the norm of the spoken Croatian variety. A third indication of differences in CL placement between the spoken and written Croatian varieties can be found in Kedveš & Werkmann (2013: 464). They cite observations made by teachers in Croatian high schools on allegedly incorrect CL placement as one of the most common mistakes in students' speech (Kedveš & Werkmann 2013: 464). The teachers' observations are based on their normativist expectations of CL placement (CLs should follow the first stressed word and are not to be placed after heavy constituents) on the one hand, and on the other hand on the fact that students' speech (i.e. the spoken variety) obviously does not meet these expectations.

## 8.3 Research questions

Based on the observations on CLs in standard BCS varieties and in dialects presented in Chapters 6 and 7, as well as on scattered conjectures about CLs in

---

<sup>1</sup>Silić does not define logical factors.

spoken BCS (Croatian) varieties presented in the section above, in this chapter we address the following research questions:

- RQ1: Does the inventory attested in the spoken Bosnian variety differ from the inventory in written standard Bosnian (and other standard varieties)?
- RQ2: What is the maximum size of CL clusters in the spoken Bosnian variety?
- RQ3: Does CL ordering in the spoken Bosnian variety differ from the ordering in written standard Bosnian (and other standard varieties)?
- RQ4: Do morphological processes like haplology of unlikes or suppletion take place within clusters in the spoken Bosnian variety?<sup>2</sup>
- RQ5: Which type of placement, 1P, 2P or DP, is dominant in the spoken Bosnian variety?
- RQ6: How heavy are constituents which precede CLs in the spoken Bosnian variety?
- RQ7: Does the spoken Bosnian variety allow phrase splitting and (pseudo)diacclisis?<sup>3,4</sup>
- RQ8: Are syntactic features typical of spoken language like disfluency phenomena and right dislocation correlated with CL placement in the spoken Bosnian variety?

## 8.4 Corpus of Bosnian interviews

### 8.4.1 Data quality

Like for dialectological studies, a major obstacle for the study of CLs in spoken BCS varieties is availability and quality of data resources.<sup>5</sup> Namely, as explained in Section 4.5, only very few corpora with data from spoken BCS varieties are available. We analyse the Bosnian Interviews corpus (Stevanović 1999), which contains 13 narrative interviews conducted with refugees from the territory of

---

<sup>2</sup>For more information on haplology of unlikes see Section 2.4.2.2.

<sup>3</sup>For more information on phrase splitting see Section 2.4.3.5.

<sup>4</sup>For more information on (pseudo)diacclisis see Section 2.4.5.

<sup>5</sup>For a discussion of this problem with respect to the study of CLs in BCS dialects see 7.3.

Bosnia (for the available sociolinguistic data see the next section) in 1994.<sup>6</sup> The minimal corpus annotation includes only deictics and regional features which mainly include pronunciation, but we introduce some additional layers of annotation (see Sections 8.5 and 8.6.1).

However, there are two major obstacles affecting our study. The biggest is lack of access to the audio recordings, which would make possible the disambiguation of some unclear parts of the transcript. This is particularly important as the transcription does not meet fully the standards which in recent years have been achieved in the burgeoning field of spoken language study.<sup>7</sup> Second, the main problem for our analysis of CL placement is the lack of consistent annotation of breaks which would allow the identification of the intonational units which deviate from clausal units as the main interactional unit. Instead, punctuation signs, in particular commas, are used for segmentation in an unsystematic way. The following examples show how breaks after *u stvari* ‘in fact’ were marked with a comma (1), three dots (2), hyphen (3) or were not marked at all (4).<sup>8</sup>

- (1) [...] [u stvari], sad je u zarobljeništvu [...].  
in matter now be.3SG in captivity  
‘[...] in fact, he is now in captivity [...].’ (BG1)
- (2) [...] [u stvari]... od kad su bili ovi izbori [...].  
in matter from when be.3PL be.PTCP.PL.M those elections  
‘[...] in fact ...since this election was held [...].’ (BG1)
- (3) [...] [u stvari] – obeležavali su naša vrata [...].  
in matter mark.PTCP.PL.M be.3PL our doors  
‘[...] in fact - they marked our doors [...].’ (TZ)
- (4) [...] [u stvari] dva mjeseca sam bio bez posla [...].  
in matter two month be.1SG be.PTCP.SG.M without job  
‘[...] in fact, I was without a job for two months [...].’ (BG1)

This kind of unsystematic annotation is especially clear in the case of breathing breaks, which, according to the rules of standard BCS orthography, should be represented with commas, but which are missing from the studied corpus.

<sup>6</sup>The reasons for choosing this particular corpus for the study of CLs in a spoken variety can be found in Section 4.6.2.

<sup>7</sup>For an example of a consistent annotation system for spoken German (GAT) see Selting et al. (1998) and for spoken Russian see the system proposed by Kibrik & Podlesskaya (2003, 2006) and presented concisely below.

<sup>8</sup>The code given in brackets matches the code of speaker in Bosnian Interviews. This is why we keep the round brackets. For details consult Table 8.1.

### 8.4.2 Sociolinguistic features of the corpus

In all, 16 people were interviewed. However, two of them (second speaker in transcript DJ and second speaker in transcript IL) played a secondary role, since the length of their utterances counted in words is much smaller in comparison to that of the remaining 14 participants. Table 8.1 summarises the socio-linguistic metadata from the Bosnian Interviews corpus on sex, age, profession or education of the interviewees, and nationality and religious background of their family members (in as much detail as available).

The transcripts were anonymised, so we cannot draw any conclusions about the exact place of origin and areas inhabited by the interviewees. In other words, establishing the interviewees' dialectal backgrounds is impossible.

It is important to understand the political and cultural background of Bosnia within Yugoslavia before the war broke out in 1991. Many people had ancestors and relatives from different Yugoslavian countries and of different religions. Since they all lived in Yugoslavia, many interviewees considered themselves Yugoslavs and atheists, as that was a common political orientation at the time. Nevertheless, upon the interviewer's request to specify their background in more detail, in most cases they provided the ethnic identity and the religion of their parents. Although, as it is clear from Table 8.1, interviewees have different social and ethnic backgrounds, the corpus compilers labelled the variety spoken by them as spoken Bosnian since most of them were born in Bosnia and they all lived there for years before they came to Germany.

Further, we can see that the group varies with respect to many sociolinguistic factors. The age spread is at least 33 years, and the speakers represent different layers of society, which we can conclude from their professions and education. Thus, the group is heterogeneous with respect to sociolinguistic factors, so there is no such factor which could clearly influence the linguistic results.

## 8.5 Principles of analysis of spoken language

As mentioned above, with the exception of dialectology, research on spoken BCS has developed poorly. Therefore, as will become clear in the following, the principles of analysis in this chapter are based on influential literature from other linguistic areas.

The first important issue in the analysis of spoken language concerns segmentation. As mentioned in Section 8.4.1, due to the lack of consistent annotation of breaks, segmentation is based on syntactic criteria. We follow the view of Thompson & Couper-Kuhlen (2005: 484) that "the clause is in fact the locus of

8 *Clitics in a corpus of a spoken variety*

Table 8.1: Sociolinguistic information about participants.

Code	Sex	Born	Information <sup>a</sup>	Profession or education	Transcript size <sup>b</sup>
BG1	M	1957	Croat, Catholic	architectural technician	2,271
BG2	M	1940	Croat, Catholic	driver	2,362
BH	F	NA <sup>c</sup>	atheist, husband Muslim	textile technician	3,279
BJ	F	1950	not declared Muslims	NA	1,970
BL	F	NA	Serbian, Orthodox	NA	2,892
BR	M	1948	atheist, parents Muslims	higher administrative school	3,109
DJ	M	1941	Muslim	mine worker	2,310
	NA	NA	NA	NA	196
DO	F	1962	Muslim	economic technician	2,463
IL1	F	1964	father Orthodox, mother Catholic	pedagogical academy	1,756
IL2	M	1958	parents Muslims	NA	789
KR	F	1949	Croat, Catholic	special needs educator	2,537
MO1	M	1957	atheist Yugoslav parents Croats, Catholics	PhD political sciences	2,013
MO2	F	1959	atheist Yugoslav, parents Serbians, Orthodox	lawyer	919
TZ	F	1973	parents Serbians	medical high school	2,920
VI	F	NA	Orthodox, Yugoslav	NA	6,632
					= 38,422

<sup>a</sup>Available information on nationality and religious background.

<sup>b</sup>In words spoken only by interviewees.

<sup>c</sup>Born in Serbia, lived in Bosnia for 20 years.

interaction in everyday conversation”. The authors continue: “[t]he clause, then, with its crucial predicate, appears to be a unit which facilitates the monitoring of talk for social actions” (cf. Thompson & Couper-Kuhlen 2005: 485). In our annotation scheme, we thus focus on syntactic clauses, but additionally take into consideration some other structures that do not coincide with clauses, as they are characteristic of spoken language. This is necessary to appropriately determine the position of a CL in a clause (see Section 8.6). Our annotation scheme is inspired by an approach which distinguishes types of ELEMENTARY DISCOURSE UNITS (EDUs), presented by the Russian linguists (Kibrik & Podlesskaya 2003, 2006), but has a stronger focus on purely structural features. In their approach, Kibrik & Podlesskaya (2003, 2006) combine formal, semantic and cognitive features which are often difficult to distinguish in our data. Therefore, we also draw on the work by Crible (2016, 2018), who designed a more consistent annotation system for various disfluency phenomena.

In the first step of data processing, we split the transcript into syntactic clauses. We focused on those which contain CLs, as illustrated in (5):<sup>9</sup>

- (5) U Bosni *sam* živela dvadeset godina [...].  
 in Bosnia be.1SG live.PTCP.SG.F twenty years  
 ‘I lived in Bosnia for twenty years [...].’ (BH)

The analysis of such clauses does not cause any problems. However, as Crible (2016: 38) points out, spoken language in its spontaneous forms is characterised by “the frequent occurrence of so-called disfluencies, which are generally considered to be cues of ongoing processes of language production and comprehension”. In the following typology, types 1–4 represent fluencemes as proposed by Crible (2018). In addition to EDUs defined as syntactic clauses we distinguish further types (see 5–7), based on Kibrik & Podlesskaya (2006), that are smaller than syntactic clauses and are not necessarily linked to disfluency. Further we added our own types 8–10.

1. IDENTICAL REPETITIONS “include any words formally similar to each other and directly contiguous” (Crible 2018: 73), as in 6, where the predicate *dala* ‘gave’ is repeated:

- (6) Sada je [dala, dala] treći u izbjeglištvu.  
 now be.3SG give.PTCP.SG.F give.PTCP.SG.F third in exile  
 ‘Now she finished, finished the third (grade) in exile.’ (BG1)

<sup>9</sup>The ekavian pronunciation is most probably due to the fact that this interviewee was born and lived in Serbia for 15 years (see Table 8.1).

## 8 Clitics in a corpus of a spoken variety

2. PARTIAL REPETITIONS are structures in which a word or phrase is broken and repeated in such a way that the broken part could be omitted, resulting in a well-formed unit, as in (7) where the interviewee corrected the broken part with *večera* ‘dinner’.

(7) I napravi se nekakva [veče... večera] a conto Božića [...].  
and make.3PRS REFL some din... dinner a conto Christmas  
‘And some din... dinner is made a conto Christmas [...].’ (KR)

3. FALSE STARTS “are interruptions that leave a segment syntactically and/or semantically incomplete and where no elements from the previous, abandoned context are taken up in what follows” (Crible 2018: 73). This is illustrated in example (8) in which the interviewee interrupts the segment *niku* and continues with *ovdje* ‘here’.

(8) [Nismo niku...] ovdje, dok smo ovdje eto u  
NEG.be.3PL anywh... here while be.1PL here well in  
Njemačkoj [...].  
Germany  
‘We were not anywh... here, while we are here, well, in Germany  
[...].’ (VI)

4. SUBSTITUTIONS correspond to any segment replaced by another with semantic and/or syntactic modification.<sup>10</sup> Unlike in repetition, in substitutions the repaired segment does not have to be contiguous. In (9) the interviewee substitutes the segment *ulazila sam* ‘I was coming in’ with the contiguous segment *ušla sam* ‘I came in’.

(9) [Ulazila sam, ušla sam] kad sam  
coming.in.PTCP.SG.F be.1SG come.in.PTCP.SG.F be.1SG when be.1SG  
htjela [...].  
want.PTCP.SG.F  
‘I was coming in, I came in when I wanted [...].’ (BJ)

5. RENDERED TOPIC (left or right dislocation) is a nominal (including prepositional) phrase that precedes or follows the clause and is not syntactically

<sup>10</sup>Crible (2018: 74) distinguishes between morphological and propositional substitutions, which we do not consider necessary.



integrated.<sup>11</sup> A reliable feature which helps to qualify a noun phrase as dislocation, and not as an actant or an adjunct in the clause, is the presence of an anaphoric pronoun of the third person which is co-referent with the topic (cf. Kibrik & Podlesskaya 2006: 7). In example (10) below the anaphoric pronoun *ona* ‘she’ signals that the nominal phrase *komunistička policija* ‘communist police’ is a rendered topic.

- (10) Jer [komunistička policija], ona je progonila  
 because communist police she be.3SG pursue.PTCP.SG.F  
 svakoga [...].  
 everyone  
 ‘Because the communist police, it pursued everyone [...].’ (BG1)

6. RETROSPECTIVE EDUS (afterthoughts) are units smaller than syntactic clauses which follow the clause or phrase they refer to. In contrast to Kibrik & Podlesskaya (2006) we do not distinguish any subtypes depending on different cognitive processes (cf. Kibrik & Podlesskaya 2006: 10). Unlike in dislocation, in retrospective EDUs there is no anaphoric pronoun which is co-referent with the element in the afterthought. In example (11) the afterthought noun phrase *praznik rada* ‘Labour Day’ follows *prvi maj* ‘first of May’.

- (11) Prvi maj, [praznik rada], se proslavljao, na primjer [...].  
 first May holiday work REFL celebrate.PTCP.SG.M on example  
 ‘May First, Labour Day, was celebrated, for example [...].’ (BR)

7. DISCOURSE STRUCTURING ELEMENTS (DSEs) can consist of single words or phrases.<sup>12</sup> They differ functionally from other types of small EDUs, as they do not carry any propositional information. Therefore, they can be eliminated from the clause due to irrelevance to the propositional content; furthermore, they cannot be subject to questions and cannot be negated (see Birzer 2015: 85f for a consistent account). Example (12) contains three DSEs: *ali* ‘but’, *recimo* ‘let’s say’, *evo* ‘here’:

<sup>11</sup>Kibrik & Podlesskaya (2006: 7) do not mention the possibility of a nominal phrase following the clause in their definition, but from our examples it is obvious that such instances can occur.

<sup>12</sup>The term “regulatory EDU” for this type of small EDU originated with Chafe (1994: 63ff) and was adapted by Kibrik & Podlesskaya (2006: 12f). However, we prefer the term “discourse structuring element” (DSE) proposed by Birzer (2015).

- (12) [Ali] [recimo], [evo] starija kćerka [...].  
 but say.IMP.1PL here older daughter  
 ‘But, let’s say, here, the older daughter [...].’ (BL)

8. OMISSIONS can be divided into ellipsis and aposiopesis. The term ELLIPSIS refers to cases where the missing element can be restituted semantically. APOSIOPESES is used in a purely syntactic and not rhetorical sense: it represents syntactic units where constituents are missing and are not recoverable from the context (cf. Karlík 2017a). In (13) we can reconstruct a predicate related to ‘studying’. In contrast, in example (14) the missing element cannot be reconstructed. We annotated such instances as aposiopesis.

- (13) Ova druga kćerka [ ]<sub>ellipsis</sub> na fakultetu u Z. a sin završava  
 this other daughter on faculty in Z. and son finish.3PRS  
 treću godinu zanata u L.  
 third year craft in L.  
 ‘This other daughter [studies] at the faculty in Z., and the son is finishing the third year of craft (school) in L.’ (BG1)

- (14) [...] samo da prestane da puca, da se može  
 just that stop.3PRS that shoot.3PRS that REFL can.3PRS  
 [ ]<sub>aposiopesis</sub> i žao mi [...].  
 and sorry me.DAT  
 ‘[...] only to stop shooting, to be able to [?...]... and I am sorry [...].’ (DO)

9. ANACOLUTHA appear in utterances where the speaker switches to a different construction within one clause (or in a broader syntactic context), which leads to a grammatically and/or semantically ill-formed structure (cf. Karlík 2017b). In example (15) the interviewee first doubles the REFL<sub>LEX</sub> CL *se* which belongs to the verb *izraziti se* ‘express’ and then omits the REFL<sub>LEX</sub> CL *se* which belongs to the verb *sjetiti se* ‘remember’.

- (15) [...] kako *se*<sub>2</sub> sada da *se*<sub>1</sub> izrazim<sub>1</sub> ne mogu<sub>2</sub> da  
 how REFL now that REFL express.1PRS NEG can.1PRS that  
 [ ]<sub>anacoluthon</sub> sjetim<sub>3</sub> [...].  
 remember.1PRS  
 ‘[...] how myself now to express myself, I can’t remember [...].’ (BL)

10. **INSERTED CLAUSE** Like Crible (2018: 75) we distinguish insertions in the sense of what she calls parenthetical insertions – “propositional segments functioning as a ‘parenthetical aside’ [...] – located in the sequence of fluencemes to which it adds some background information without directly modifying the content of the utterance”. In our data insertions are mainly relative clauses or parentheticals, such as *gdje smo živjeli* ‘where we lived’ in (16) and *zna se dobro* ‘it is well known’ in (17).

(16) [...] u okruženju, [gdje smo živjeli], smo  
 in environment where be.1PL live.PTCP.PL.M be.1PL  
 bili ređi.  
 be.PTCP.PL.M rarer  
 ‘[...] in the environment where we lived, we were rarer.’ (KR)

(17) U prošlom ratu, [zna se dobro], nije [...] ne  
 in previous war know.3PRS REFL well NEG.be.3SG NEG  
 bi bilo ništa novo [...].  
 COND.3SG be.PTCP.SG.N nothing new  
 ‘During the last war – this is well known – there would not have  
 been anything new [...].’ (BG1)

## 8.6 Data preparation

### 8.6.1 Annotation scheme

Our objective for the annotation was economy, transparency, and adequacy of analysis. We distinguished three main steps in the annotation process: segmentation into clauses, annotation of categories related to CLs (inventory and position-related phenomena mentioned in Chapter 2), and annotation of syntactic structures described in Section 8.5. The full coding scheme is given in Tables 8.2–8.4.

### 8.6.2 Inventory-related categories

The first topic investigated is inventory-related categories which include distribution of CL types, the types of clusters and the ordering of CLs in clusters in spoken Bosnian in comparison to standard written Bosnian (and other standard varieties). We include this in Table 8.2.

In the CL Type category, we annotated not only single CLs but also all occurrences of two and more CLs in a clause. The distinction between clusters and

Table 8.2: The coding scheme of inventory-related categories

Category	Values
CL	VERB = any verbal CL except <i>je</i> VERB_je = verbal CL <i>je</i> PRON_acc PRON_dat PRON_gen REFL = reflexive CL REFL-X = reflexive CL used in haplogy of unlikes QUEST = the polar question marker <i>li</i> any combination of the above CLs
cluster	0 = no cluster 1 = cluster
(pseudo)diaclisis	0 = no diaclisis 1 = diaclisis 2 = pseudodiaclisis

(pseudo)diaclisis was annotated separately. This allowed us to obtain information on the types of CLs which clusterise, and compute the maximal size of a cluster.

As to morphonological processes, we are particularly interested in the interaction of the verbal CL *je* and the reflexive CL *se*. In written language their co-occurrence usually leads to haplogy of unlikes.<sup>13</sup> Therefore, the reflexive CL *se* appearing without the verbal CL *je*, which has been haplogised, was annotated separately as REFL-X.<sup>14</sup>

Some CL forms, in particular the REFL<sub>2ND</sub> *si* and the pronominal *ju*, were not included in the annotation scheme presented in Table 8.2 due to their infrequency. However, they were observed in the data and we comment on them in Section 8.7.

<sup>13</sup>Haplogy of unlikes and pseudodiaclisis are the only phenomena related to CC which are included in the coding scheme. Skipping the annotation of other phenomena related to CC is motivated by the overall small size of the corpus. We devote Part III to CC and base the discussion on empirical material retrieved from large web corpora.

<sup>14</sup>Note that this element of annotation relates to the mere surface structure of haplogy. It does not refer to the typology of reflexives proposed in Section 2.5.4.

### 8.6.3 Position-related categories

Secondly, we are interested in CL position in the clause. Categories which had to be annotated for the study of this topic are shown in Table 8.3. One of the phenomena considered under positioning is phrase splitting, where the CL is placed within the hosting phrase. This, however, can take place only if a hosting phrase can be split. Therefore, in the coding scheme we distinguished clauses which do not satisfy the conditions for phrase splitting (coded as 0), clauses where phrase splitting takes place (coded as 1), and clauses where phrase splitting could theoretically be possible (phrases appeared in a position before the CL, but phrase splitting did not take place; coded as 2).

We distinguished three types of positions of the CL or cluster: 1P (where there is no initial host constituent in the clause or alternatively the CLs directly follow an insertion), 2P and DP. In order to establish the placement type, we took into account the special syntactic structures discussed above in Section 8.5 and listed in Table 8.4. In most cases we were simply interested in whether these phenomena, including DSEs or inserted clauses, appeared before the CLs in the

Table 8.3: The coding scheme of position-related categories

Category	Values
position of CL	1P = initial position 2P = second position DP = delayed position
position of CL in relation to the number of words	number of stressed words (prosodic units) counted from the beginning of the clause or from the beginning of the insertion if a CL occurs after an insertion
length of host (only for DP)	number of graphemes of the host constituent in DP
length of the initial constituent	number of graphemes of the initial constituent (for 2P the length of the initial constituent coincides with the length of the host constituent)
phrase splitting	0 = phrase splitting impossible 1 = phrase splitting 2 = no phrase splitting, although possible

Table 8.4: The coding scheme of syntactic structures

Category	Values
A.	position of a syntactic structure in relation to CL(s)
	identical repetition
	0 = no
	1 = yes, before CL
	2 = yes, after CL
	3 = yes, including CL
	partial repetition
	0 = no
	1 = yes, before CL
	2 = yes, after CL
	3 = yes, including CL
	false start
	0 = no
	1 = yes, before CL
	2 = yes, after CL
	3 = yes, including CL
	substitution
	0 = no
	1 = yes, before CL
	2 = yes, after CL
	3 = yes, including CL
B.	presence of a syntactic structure before CL(s)
	rendered topic
	0 = no
	1 = yes
	Retrospective
	0 = no
	1 = yes
	DSE
	0 = no
	1 = yes
	ellipsis
	0 = no
	1 = yes
	aposiopesis
	0 = no
	1 = yes
	anacoluthon
	0 = no
	1 = yes
	inserted clause
	0 = no
	1 = yes



## 8.7 Inventory

### 8.7.1 Distribution of clitics in the corpus of spoken Bosnian

The annotated clauses contain 4727 CLs. It is interesting to note that CLs are very rarely attested in repetitions (both identical and partial), false starts or substitutions. Speakers repeated or replaced (when self-correcting) a total of 132 CLs in 117 clauses. This number accounts for only 3% of all CL usages in the corpus. In the majority of cases the replacement in partial repetitions and substitutions contains a verbal CL (100 of 117 clauses). We did not count CLs appearing in repetitions and substitutions in the overall distribution shown in Figure 8.1.

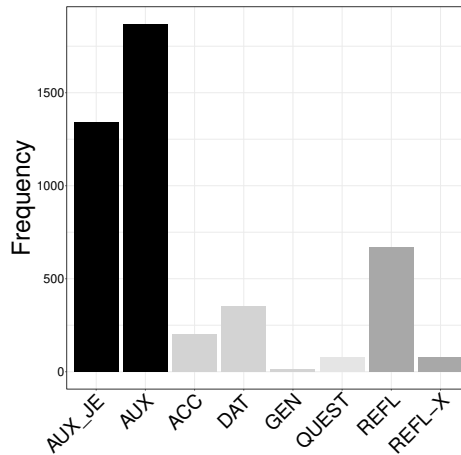


Figure 8.1: Distribution of single CLs in the corpus

To start with, in general the CL inventory coincides with the inventory attested in written standard Bosnian (and other standard varieties). Nevertheless, the reflexive CL *si* was identified as an additional CL element in the inventory of the spoken Bosnian variety that is not present in the inventory of either standard Bosnian or standard Serbian. We shortly comment on this CL in Section 8.7.4. Verbal CLs are the most frequent CL and they make up 69% of the sample ( $N = 3210$ ; almost half of that ( $N = 1341$ ) are occurrences of the CL *je* 'is'). The second most frequent type is reflexive CLs ( $N = 750$ ), followed by pronominal CLs ( $N = 565$ ). The question marker *li* ( $N = 79$ ) is mostly used in its reduced form, transcribed as *l'*, like in (19).



- (19) E, sad da l' će te biti mržnje, da l' neće, ne  
 eh now that Q FUT.3SG this be.INF hatred that Q NEG.FUT.3SG NEG  
 znam [...]  
 know.1PRS  
 'Eh, now, will there be hatred, or not, I don't know...' (VI)

### 8.7.2 Pronominal clitics

The frequencies of pronominal CLs are as follows: dative CLs are the most frequent ( $N = 350$ ), followed by accusative ( $N = 198$ ), and genitive ( $N = 15$ ). The high frequency of dative CLs is related to the non-argumental, i.e., possessive (20) and ethical (21) dative.

- (20) [...] muž mi je tu išo u školu [...].  
 husband me.DAT be.3SG here go.PTCP.SG.M to school  
 '[...] my husband went to school here [...]' (DO)
- (21) I ne znam ti koje već, šta već [...].  
 and NEG know.1PRS you.DAT which already what already  
 'And I honestly do not know whichever, whatever [...]' (BH)

As mentioned in Section 6.3.1 written standard BCS varieties differ with respect to their usage of the accusative pronominal CLs *ju* and *je* 'her'. Therefore, we examined the corpus to determine the distribution of these forms in the spoken Bosnian variety. We found 13 clauses with pronominal CLs in the accusative third person singular feminine form. However, all of them contained the CL form *je*, like in (22). We thus found no empirical evidence for the CL *ju* being part of the CL inventory in the language of the recorded speakers.

- (22) [...] ne vidim je nikako.  
 NEG see.1PRS her.ACC nohow  
 '[...] I don't see it at all.' (BG1)

Nevertheless, regardless of the absence of the CL form *ju* from the corpus of spoken Bosnian, we should refrain from generalisations concerning the usage of this CL in spoken Bosnian as such.<sup>17</sup>

<sup>17</sup>Caution is necessary here. Only 16 informants contributed to the analysed corpus of spoken Bosnian. Our statement does not mean that there is no CL *ju* at all in any Bosnian varieties, nor that the CL *ju* cannot generally be attested in the spoken Bosnian variety. The reason for this caution is the size of the corpus of spoken Bosnian on the one hand and data from bsWaC on the other. If we compare the distribution of accusative CLs *ju* [tag="Pp3fSa" & word="ju"] and

Further, we notice that the CL for third person plural accusative is reduced by some speakers to a form transcribed as *i'* instead of *ih* 'them': see example (23) below. This is in line with forms found in most dialects (see Section 7.4.1.3).

- (23) Od koga *i'* brani – od komšija.  
 from whom them.ACC protect.3PRS from neighbours  
 'Whom is he protecting them from – from the neighbours.' (VI)

### 8.7.3 Verbal clitics

As pointed out above, verbal CLs are quantitatively the most frequent CL type in the whole corpus of spoken Bosnian, with the CL *je* 'is' as the most frequent CL form ( $N = 1341$ , 29% of all CL occurrences in the corpus).

We already indicated in Section 7.4.2 that in some Štokavian dialects there is only one form of the conditional auxiliary for all persons and that this form is spreading from dialects into spoken BCS varieties. This syncretism is also attested in the corpus data: the interviewees use the CL form *bi* 'would' for all persons in the conditional. This is nicely illustrated in (24), where the interviewee uses the CL form *bi* and not the CL form *bih* which is prescribed in standard BCS varieties.

- (24) [...] kad *bi* ja stvarno mogla još više dati [...].  
 when COND.1SG I really can.PTCP.SG.F still more give  
 '[...] if I could really give even more [...]' (VI)

Inflected forms of the conditional occur 5 times, only as forms of the first person singular and plural *bih* (25) and *bismo* (26), whereas the uninflected form *bi* for those and the second person plural is used 14 times.

- (25) [...] dodao *bih* još možda malo tačniji odgovor.  
 add.PTCP.SG.M COND.1SG still perhaps little more.precise answer  
 '[...] I would add maybe a slightly more accurate answer.' (MO1)
- (26) [...] ono što *bismo* mi željeli [...].  
 That what COND.1.PL we wish.PTCP.PL.M  
 '[...] the thing we would want...' (MO1)

---

*je* [tag="Pp3f5a" & word="je"] in bsWaC, we get the following results: 27,433 occurrences of *ju* (95.6 per million) and 33,305 occurrences of *je* (116.1 per million). As we can see, the difference in the distribution of the competing forms is not that extensive at all. Moreover, the CL *ju* is attested in dialectological data from the language territory of Bosnia and Herzegovina presented in Section 7.4.1.1.

Regardless of the small number of observations, we may assume that they are probably a case of diastatic variation. Namely, the inflected forms prescribed in standard BCS are used by spouses who obtained a higher education than the rest of the interviewees. The male speaker (25) had a PhD degree and the female (26) was a lawyer.

#### 8.7.4 Reflexive clitics

As stated in Section 6.3.3 there is diatopic variation in the inventory of reflexive CLs between the BCS standard varieties. The REFL<sub>2ND</sub> *si* is only recognised by authors describing the Croatian standard. It does occur in the analysed corpus, but with a very low frequency. Namely, in 750 occurrences of reflexive CLs we find two instances of REFL<sub>2ND</sub> *si*, which is 0.3% of all occurrences. Both of the following utterances were produced by the same speaker.

(27) [...] mogli smo si dozvoliti pristojne uvjete života [...].  
 can.PTCP.PL.M be.1PL REFL allow.INF decent conditions life  
 ‘We were able to afford decent living conditions [...].’ (KR)

(28) [...] i mogao si je priuštiti.  
 and can.PTCP.SG.M REFL be.3SG afford.INF  
 ‘[a]nd he could allow himself.’ (KR)

As mentioned in Section 7.4.3, the occurrence of the REFL<sub>2ND</sub> CL *si* is also reported for idioms of Western Herzegovina and Northern Bosnia. Considering the frequency distribution in the corpus, we have to admit that when compared with the reflexive CL *se*, the reflexive CL *si* is not frequent in Croatian either. In hrWaC, the occurrences of the CL *si* make up only 1.19% of all reflexive CL occurrences.<sup>18</sup> The difference between the standard BCS varieties is that in contrast to Bosnian and Serbian, the Croatian standard recognises the REFL<sub>2ND</sub> *si* as part of the codified system. However, the form as such occurs not only in spoken Croatian but also in spoken Bosnian (pace Ridjanović 2012: 440).

#### 8.7.5 Clusters

Clusters are attested in 461 clauses. In total, 454 clusters consisted of 2 CLs and only 7, of 3 CLs, which sheds new empirical light on the size of the CL cluster. Note that Piper & Klajn (2014: 451f) claim that a CL cluster usually consists of

<sup>18</sup>95,016 out of 7,969,617 occurrences of reflexive CLs. The low frequency of the reflexive CL *si* can probably partially be attributed to its homonymy with the verbal CL *si* and tagging inaccuracy.

two or three elements. Our corpus data show that at least in the spoken Bosnian variety CL clusters with three CLs are much rarer than CL clusters with two CLs.

The frequency distribution of the most frequent types is shown below in Figure 8.2. Note that, as already mentioned in Table 8.2, we annotated the verbal CL *je* separately from the other verbal CLs (V\_je vs V).

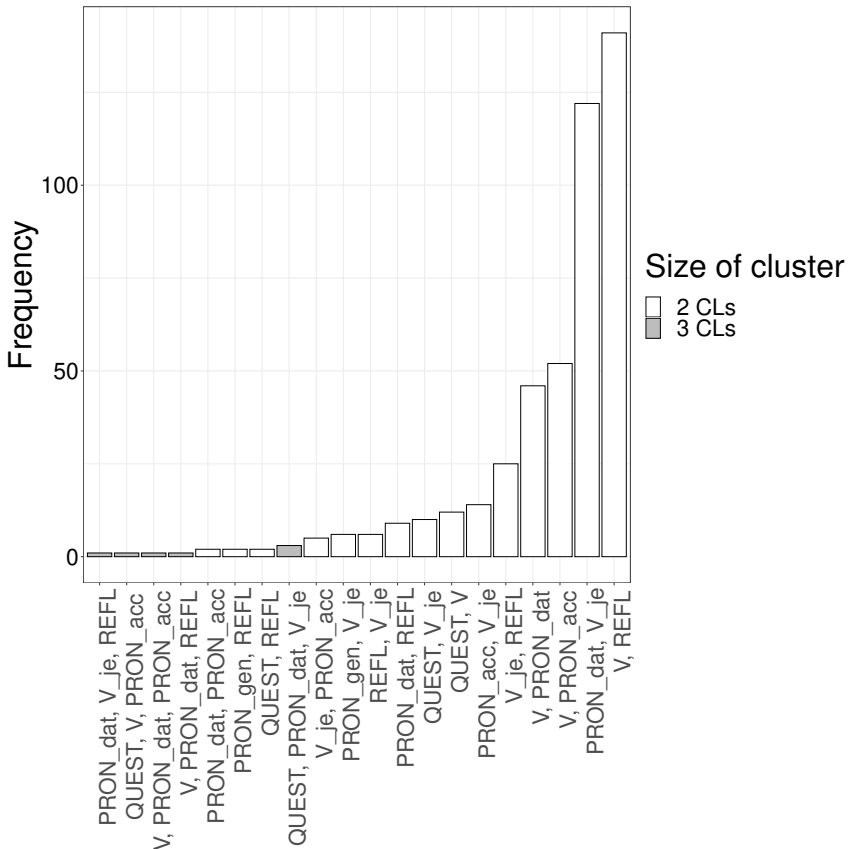


Figure 8.2: Distribution of cluster types in the corpus of spoken Bosnian

Fifteen different combinations for 2-CL clusters and 5 different combinations for 3-CL clusters are attested. The most frequent types are V(erbale) + REFL ( $N = 141$ ) and PRON\_dat + V(erbale)\_je ( $N = 122$ ) – note that this combination with a dative CL is much more frequent than the combination PRON\_acc + V(erbale)\_je ( $N = 19$ ).<sup>19</sup> In contrast, the combinations V(erbale) + PRON\_dat ( $N = 46$ ) and

<sup>19</sup>These 19 occurrences also include CL clusters with the verbal CL *je* and a pronominal CL in the accusative in an order which diverges from the CL order attested in the standard BCS varieties. For more information see below.

V(erbal) + PRON\_acc ( $N = 51$ ) are similarly frequent. As already observed in Section 8.7.2 above, the high frequency of dative CLs is due to occurrences of possessive and ethical dative. These numbers, however, indicate that possessive dative might be more frequent than ethical dative, since possessive dative is typically used with the verbal CL *je* as in example (21) provided above.<sup>20</sup>

The combination PRON\_dat + REFL appears 9 times; as expected there are no combinations of reflexive CLs with the accusative.<sup>21</sup>

## 8.8 Internal organisation of the clitic cluster

### 8.8.1 Clitic ordering within the cluster

In Section 6.4.1 we pointed out that in BCS standard varieties the ordering sequence of CLs in clusters does not differ. In our data, however, we do find two types of CL order in the cluster which diverge from the order attested in standard BCS varieties.

The first and by far the most common CL order diverging from the sequence given by Franks & King (2000: 29) and presented in Section 2.4.2.1 involves the verbal CL *je* and the reflexive CL *se*. The expected CL order *se je* allowed in standard Croatian and Bosnian appears only six times: one of those utterances is presented in (29).

- (29) [...] *onda se je Irma u mene pre* [...] (BR)  
 then REFL be.3SG Irma in me.GEN get.scare  
 ‘[...] then my Irma got scare...’

In contrast, the reversed CL order *je se* is found 25 times in clusters with 2 CLs (30), making this the fifth most frequent type of cluster, and once in a cluster with 3 CLs (31).

- (30) *Gore je se oženio* [...] (DJ)  
 up.there be.3SG REFL marry.PTCP.SG.M  
 ‘He got married up there [...]’
- (31) [...] *nego muž mi je se prep’o* [...] (B)  
 but husband me.DAT be.3SG REFL get.scared.PTCP.SG.M  
 ‘[...] but my husband got scared [...]’

<sup>20</sup>Bear in mind that in our annotation scheme we did not distinguish between possessive and ethical dative CLs.

<sup>21</sup>This is not very surprising since REFL<sub>LEX</sub> have genitive and not accusative complements.

Still, both patterns are much less frequent than the haploglised structure (which omits the verbal CL *je*), discussed in the next section.

The second CL ordering sequence which diverges from the order attested in standard BCS varieties also involves the verbal CL *je*. Although the established CL order in standard BCS has the CL *je* appearing in the final position of the ordering sequence, we found 5 clusters in which the verbal CL *je* precedes a pronominal accusative CL, like in example (32).

- (32) [...] koji je me dobro poznavao i cijenio [...].  
 which be.3SG me.ACC good know.PTCP.SG.M and respect.PTCP.SG.M  
 ‘[...] who knew me very well and respected me [...]’ (BG1)

Although these two types of CL ordering in a cluster attested in spoken Bosnian diverge from the standard BCS varieties, they do not come as a surprise. As already mentioned in Section 7.5.1, they are also attested in dialects spoken on BSC territory.

### 8.8.2 Morphological processes within the cluster

In order to identify possible microvariation, we analysed reflexive pronouns with regard to the following categories:

1. haplogy of unlikes (*se* appearing alone where a *se je* cluster is expected),
2. co-occurrence with *je* (*se je* appearing in a cluster),
3. *se* and *je* appearing in (pseudo)diacclisis.

The data contain 122 clauses with possible co-occurrence of the reflexive CL *se* and the verbal CL *je*, with the distribution in Figure 8.3.

In Figure 8.3 we see that in 68.8% of cases the reflexive CL *se* appears without *je* and in 25.4% of cases, with *je*. This speaks for the preference of haplogy in spoken Bosnian. All *je se* clusters are simple clusters. Similarly, the haplogical forms are generated by one verb. All analysed instances are in a past-tense context. Our data thus show that haplogy of unlikes described by many grammarians of the standard BCS varieties (e.g. Težak & Babić 1996: 246, Barić et al. 1997: 596, Jahić et al. 2000: 471, Ridjanović 2012: 302, 333, Piper & Klajn 2014: 450) is not the rule in spoken Bosnian.<sup>22</sup> Although *je* is often omitted as an auxiliary and in simple clusters, the non-haplogical forms are only twice less frequent.

---

<sup>22</sup>See also Section 6.4.2.2.

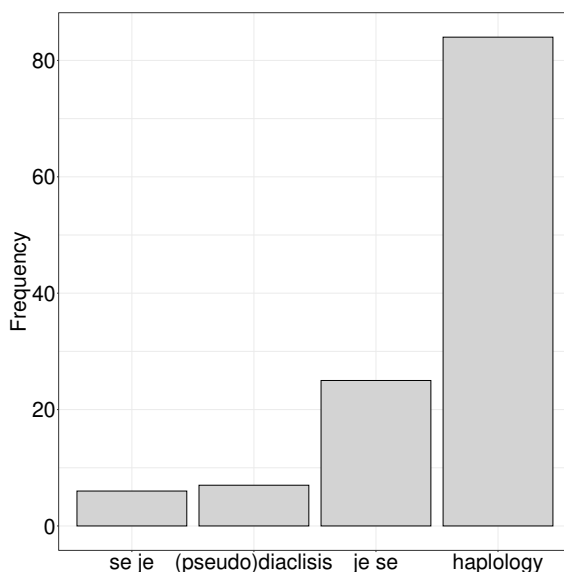


Figure 8.3: Distribution of different constructions with the reflexive CL *se* and the verbal CL *je* in clauses

Our data do not allow for any conclusions about mixed clusters or usage as a copula.

As already mentioned in the previous section, the combination of the CLs *je* and *se* mostly appears in an order (33) which is reversed in comparison to the one attested in standard Croatian and Bosnian varieties. However, even this reversed order is less frequent than haplogy of unlikes (34).

(33) [...] znalo            *je*    *se* [...].  
    know.PTCP.SG.N be.3SG REFL  
    ‘[...] it was known [...].’ (BR)

(34) [...] znalo            *se*    uvijek [...].  
    know.PTCP.SG.N REFL always  
    ‘[...] it was always known [...].’ (BL)

Other types of variation with respect to morphonological processes within the cluster are not evident in the data. The morphonological process of suppletion, in which the feminine accusative pronominal CL *je* is replaced with its counterpart *ju* when followed by its homonym, the verbal CL *je*, is not attested at all. Note that in revised dialectological data we find no evidence for suppletion either. We may assume that the Bosnian linguist Ridjanović (2012: 434) could be right in his

claim that suppletion is a feature of deliberate speech, but more robust data is definitely needed on this matter.

## 8.9 Position of the clitic or the clitic cluster

### 8.9.1 General distribution

The following discussion on CL positioning is based on a subsample of 3829 clauses. This choice is motivated by the fact that on the basis of anonymised transcripts only, some clauses were impossible to interpret and a recording would have been needed. The 22 cases of (pseudo)diacclisis are not included in the sample and are discussed separately in Section 8.10. For the time being, we also excluded all cases where CLs were substituted by other CLs as a type of disfluency phenomenon.

We analysed the positions of single CLs and clusters separately and compared them to establish whether any differences could be observed. In total, we analysed 3,399 single CLs (26 in 1P and 164 in DP) and 430 clusters (3 in 1P, 17 in DP). The logarithmic frequencies of the position of single CLs and CL clusters are as given in Figure 8.4.

Pearson's chi-squared test does not show any significant difference between single CLs and clusters with respect to placement ( $p = 0.1991$ ). Regardless of whether the CLs appear in clusters or as single CLs, a retrograde fall, i.e., a reduction in occurrences from 2P (94% for single CLs, 95% for clusters), through 3P (4% for both single CLs and clusters) to 1P (< 1% for both single CLs and clusters), is noticeable for all CLs.

The category labelled 1P was mostly recorded for the position after an insertion; we discuss it in Section 8.11. Among single CLs in 1P ( $N = 26$ ) we identify only verbal ( $N = 19$ ) and reflexive ( $N = 7$ ) CLs, in particular the verbal CL *je*, which appears 15 times.

### 8.9.2 Placement of single CLs

We now turn to differences in the positioning of individual CLs compared to cluster types. We found occurrences of delayed placement for all types of single CLs except the polar question CL *li*. In Chapter 2 we pointed out that the polar question marker *li* differs significantly from other CLs because it does not have a non-clitic equivalent. In our data we annotated 79 appearances of the polar question marker *li* in total. The only significant fact we would like to raise is that the CL *li* takes the second position in 100% of cases. Moreover, in 100% of



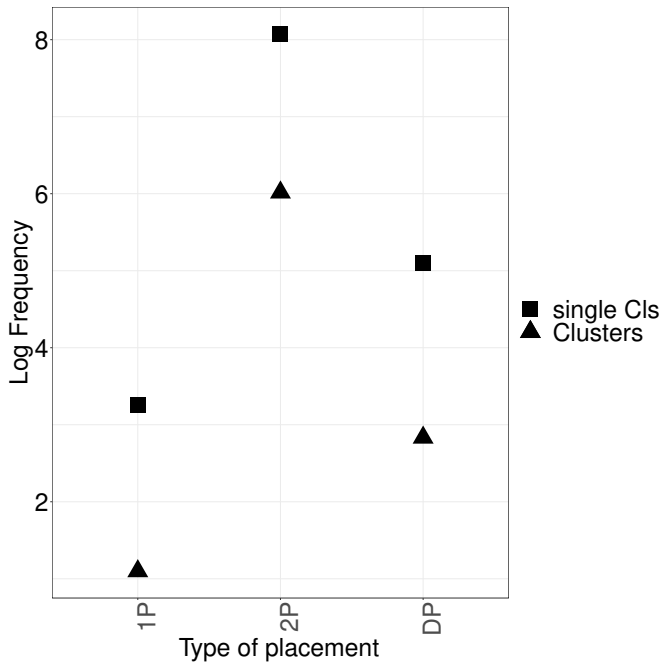


Figure 8.4: Frequencies of placement types for single CLs and CL clusters (normalised to natural logarithm for conciseness)

cases *li* follows one very short word (only 2 to 4 graphemes long). This is in line with the observation of Siewierska & Uhlířová (1998: 119) who call this CL “inflexible”.

Little variation is observed for pronominal CLs. While accusative and dative pronominal CLs were attested not only in 2P but also in DP, genitive CLs were attested only in 2P. The generally more frequent verbal and reflexive CLs differ somewhat from pronominal CLs and the polar question marker *li*. As well as in delayed placement, in rare cases they also appear in the first position (see more below). Nonetheless, delayed placement seems equally rare for all the three CL types: verbal, reflexive and pronominal.

### 8.9.3 Placement of clusters

The distribution of cluster types across positions is shown in Figure 8.6. In all, only 17 of 455 analysed clusters occupy DP. The two most frequent clusters V(erbal), REFL ( $N = 5$ ) and PRON\_dat, V(erbal)\_je ( $N = 6$ ), also have the highest frequency in DP. For six clusters only single occurrences in DP are observed.

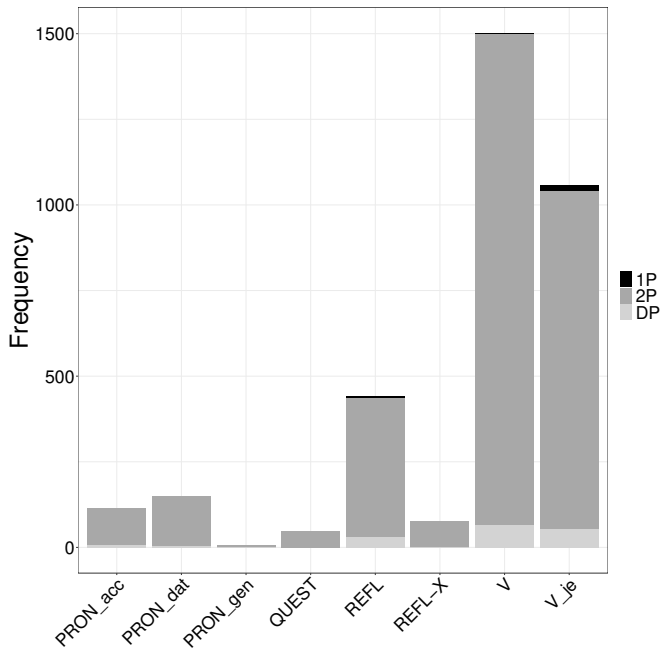


Figure 8.5: Placement of single CLs

Although reflexive CLs can take not only 2P but also DP and 1P, clusters starting with a reflexive CL show no variation in this respect. Namely, they always appear in 2P in the data. However, other CL clusters containing the reflexive CL *se* do show variation. For instance, the cluster *je se* was attested not only in 2P, but also in DP and 1P.

Clusters containing the polar question marker *li* are, similarly to *li* occurring as a single CL, observed in the data only in 2P.

#### 8.9.4 Relationship between the length of preceding phrases and clitic placement

The previous sections showed no substantial difference in terms of placement in clauses between CLs and CL clusters. Second position is by far the dominant position in spoken Bosnian. Delayed placement represents about 4% of CL occurrences.

In the following lines, we focus on the differences between 2P and DP. As already discussed in Sections 6.5.2 and Section 6.5.4, delayed placement is usually associated with breathing breaks and long initial or heavy phrases. We measure

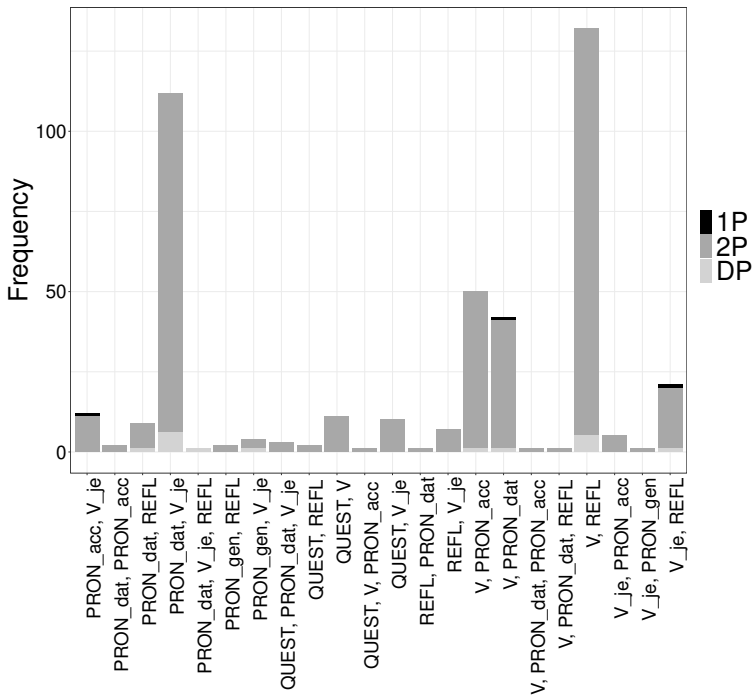


Figure 8.6: Placement of CL clusters

the length of constituents according to the number of intonational words and the number of graphemes as suggested in Kosek et al. (2018) and thoroughly explained in Section 8.6.3. This approach is rather new in studies of South Slavic languages, where most authors argue in favour of a mixed prosodic and syntactic approach.<sup>23</sup> Figure 8.7 shows the differences in placement related to the number of preceding words counted from the beginning of a clause or the end of an insertion.

One preceding word is by definition necessary for 2P. The studied data contain 2,982 such observations. Thus,  $2P = 2W$  placement holds for 77% of the data.<sup>24</sup>

According to Radanović-Kocić (1988: 108ff, 1996: 435) CLs do not usually follow an initial phrase longer than two words, unless that phrase is a subject. Only in the latter case can a phrase longer than two intonational words be a potential

<sup>23</sup>However, none of these authors offer solutions for how exactly to empirically distinguish (heavy) phrases which cannot host CLs from phrases which can host CLs. For more information on approaches to 2P effects see Sections 2.4.3.1 and 2.4.3.2.

<sup>24</sup>For more information on 2W see Section 6.5.4.1

8 *Clitics in a corpus of a spoken variety*

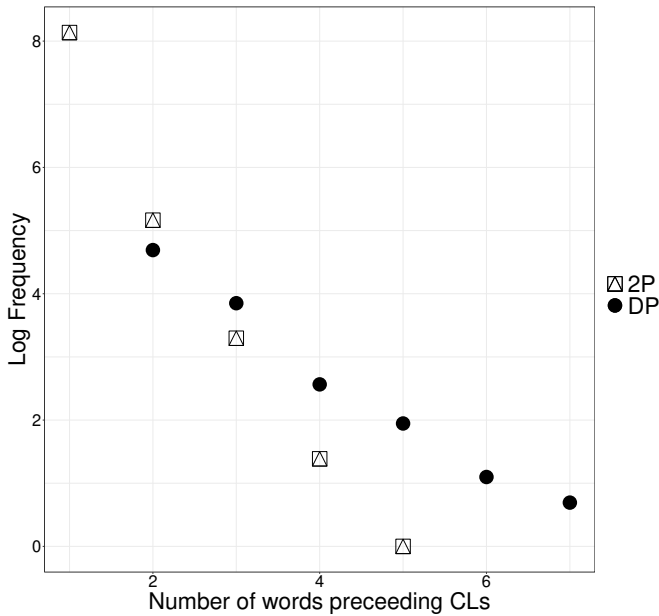


Figure 8.7: Frequencies of CLs in DP and 2P (normalised to natural logarithm for conciseness. The abscissa shows the number of words preceding the CL).

CL host. We found 175 observations with CL placement in the second position after two stressed words, and 31 observations including initial host constituents which are 3–5 words long like *do aprila devedeset i druge* ‘until April 1992’ in (35):

- (35) [Do aprila devedeset i druge] je podnošljivo bilo.  
 until April ninety and two be.3SG bearable be.PTCP.SG.N  
 ‘Until April 1992 it was bearable.’ (KR)

This example proves that also longer, non-subject-initial constituents may host CLs in spoken Bosnian Radanović-Kocić (pace Radanović-Kocić 1988: 108ff, 1996: 435).

Delayed placement appears in the transcripts of all speakers, with the exception of the very short transcript of the second speaker in interview DJ. One hundred eighty one clauses contain a CL or cluster in DP. The part preceding the delayed CL is 2 to 7 words long. Figure 8.7 suggests that when the CL is placed after the third or further words, the probability is in favour of DP; that is, in such cases at least two constituents are usually involved. Nonetheless, the constituents preceding 2P CLs may be relatively long when counted in graphemes.

This is shown in Figure 8.8.

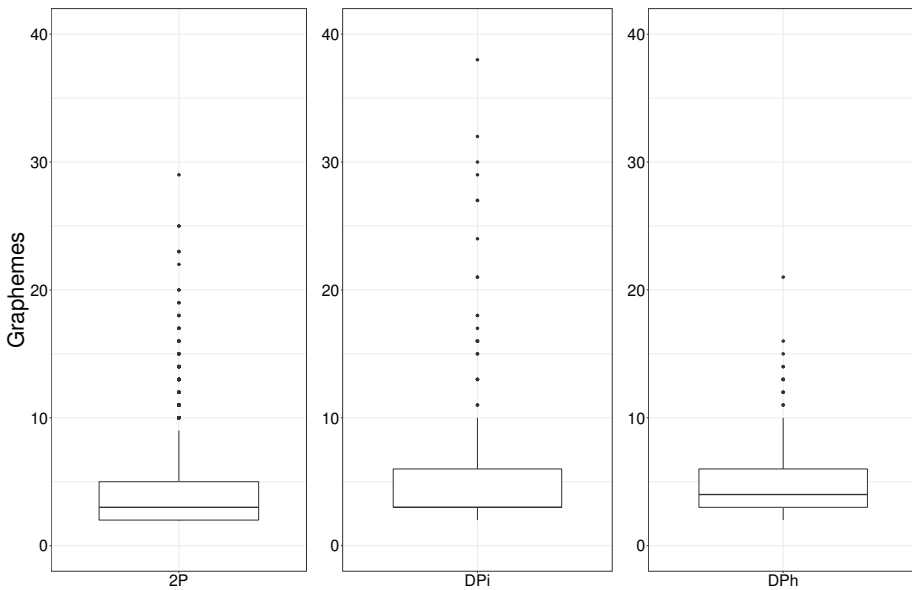


Figure 8.8: Box plot representing the frequency distribution for the length of constituents preceding CLs measured in graphemes (ordinate). 2P – initial constituents for 2P,  $DP_i$  – initial constituent in DP,  $DP_h$  – host constituent in DP.

In Figure 8.8 we present box plots where the lower whisker represents the minimum value. The edges of the box show the upper and the lower quartile (25<sup>th</sup> and 75<sup>th</sup> percentile), while the thick line represents the median. The upper whisker represents the trimmed estimator based on interquartile range, allowing the outlier values to be seen.<sup>25</sup>

The minimum, first quartile and mode of 2P are all equal to 2 (see also Table 8.5). The most frequent length of the initial constituent preceding CLs in 2P is two graphemes, as seen in 30% of observations. Twenty per cent of observations in that group are three graphemes long, which is also the median. Another 25% of observations are 4–5 graphemes long. Only 28 observations (< 1%) are longer than 9 graphemes. As given in Table 8.5, the mean is 4.12 and standard deviation (SD) 2.60.

As mentioned in Section 8.6.3, we follow Kosek et al. (2018) in describing DP. We computed two parameters: the length of the initial constituent ( $DP_i$ ) and the

<sup>25</sup>Interquartile range is the difference between the upper and the lower quartile multiplied by 1.5.

length of the actual host ( $DP_h$ ). The two types of constituents have some common distributional properties: the minimum value (2), the lower quartile (3) and the upper quartile (6). Nonetheless, they differ as to mode and median, which are equal to each other for both types of constituents. The most frequent initial constituent length is 3 graphemes (18% of observations with DP). In the case of the host it is 4 (23% of observations with DP). In both cases, only single observations exceed the value of 10. However, the outliers in host constituents are shorter (the maximum value is 29 graphemes) than the outliers for initial constituents, which reach a length of up to 38 graphemes. Thus, we observe that both types of constituents appearing in DP are, in general, longer than the initial constituents that host CLs in 2P.

Although the most frequent host in DP is longer than the initial constituent in DP, its values are more “compact”, which is visible when the standard deviation (SD) in Table 8.5 is compared. It takes the highest value for  $DP_i$ , the middle for  $DP_h$ , and the lowest for 2P. Importantly, SD of  $DP_h$  is much closer to SD of 2P than of  $DP_i$ . This result suggests that the length of the host counted in graphemes is limited in some way.

Table 8.5: Descriptive statistics for the length of the constituent preceding a CL in a clause

	Measurement in graphemes			
	Mean	SD	Median	Mode
2P	4.12	2.60	3	2
$DP_i$	6.03	6.25	3	3
$DP_h$	5.38	3.15	4	4

We tested the results for significance. None of the three distributions come from the normal distribution which can be tested with the Shapiro-Wilk normality test (2P:  $W = 0.74724$ ,  $p < 2.2e-16$ ;  $DP_i$ :  $W = 0.63624$ ,  $p < 2.2e-16$ ;  $DP_h$ :  $W = 0.84325$ ,  $p = 1.13e-12$ ). Therefore, we investigated the differences in lengths of particular constituents using non-parametric tests. The difference between the distribution of a DP initial constituent and a 2P initial constituent is significant according to the Kolmogorov-Smirnov test ( $D = 0.24803$ ,  $p = 1.384e-05$ ). We made sure that the difference is not a result of location shift. To this end we performed the Wilcoxon Rank-Sum Test, which confirmed that the true location shift is not equal to 0 ( $W = 18794$ ,  $p = 0.01412$ ).

The same holds for the difference between the DP host and the 2P initial constituent (Kolmogorov-Smirnov test:  $D = 0.23772$ ,  $p = 6.914e-09$ ; Wilcoxon Rank-Sum Test:  $W = 421830$ ,  $p = 2.214e-11$ ). Thus the length of an initially positioned host for 2P CLs and the length of the two constituents distinguished for DP differ significantly.

We now examine the relationship between the DP constituents. The observations should be treated pairwise, as this is the way these constituents occur. We first show them in Figure 8.9. The 181 observations are sorted according to the length of all constituents preceding the delayed CL. Circles and triangles represent the actual constituent lengths, while the black and grey lines depict the main trend in the data.

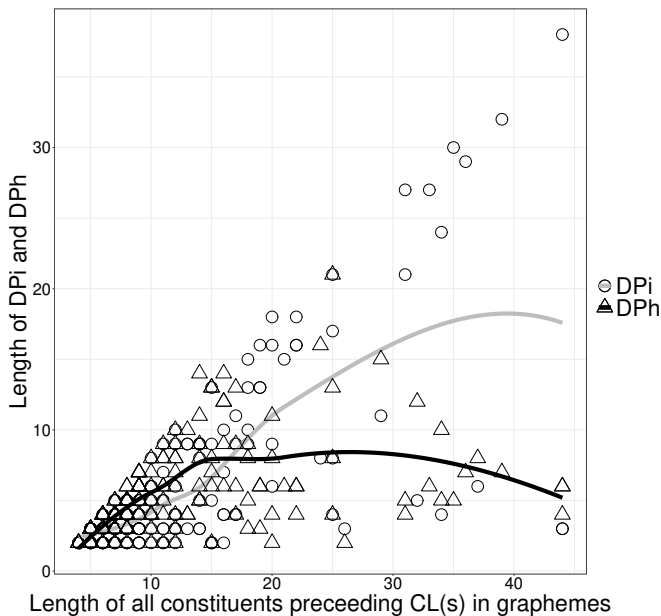


Figure 8.9: Pairwise length of initial and host constituents in DP

The longer the constituents, the bigger the difference between the initial constituent and the host. In very short constituents (up to six graphemes long) the host is often longer than the initial constituent. However, when constituents become longer, host length remains at the same level, while the initial constituent may still lengthen. Because the deviations in the data are obviously caused by very long initial constituents, we tested for the significance of the difference between the host and the initial constituent leaving out the nine (less than 5%) of

the longest initial constituents, that is, the constituents with over 20 graphemes. We used the Wilcoxon signed-rank test for paired vectors with the alternative hypothesis that the median for the initial constituent is lower than the median for the host. The result of the test ( $V = 4687.5$ ,  $p = 0.01018$ ) allows us to reject the null-hypothesis.

Therefore, we conclude that the phenomenon of DP in spoken Bosnian is a result of significantly long initial constituents which block 2P placement. Surprisingly, the actual hosts are, in most cases, even longer than initial constituents. According to the trend visible in Figure 8.9, this regularity applies to clauses where both the host and the initial constituent are about six graphemes, like in example (36) where the initial constituent is four graphemes long, whereas the actual host is eight graphemes long.

- (36) Ne da mislim, [nego]<sub>phrase1</sub> [sto posto]<sub>host</sub> sam  
 NEG that think.1PRS but hundred percent be.1SG  
 ubijeden [...].  
 convinced.PTCP  
 ‘Not that I assume, but I am one hundred percent convinced [...].’ (BG1)

In the subset of initial constituents longer than 6 graphemes, the host is shorter, and it stays at a length of 3–6 graphemes (37). Hence, the length of the host remains at the same level.

- (37) [...] [drugarica najbolja]<sub>phrase1</sub> [bila]<sub>host</sub> mi je Srkinja [...].  
 friend best be.PTCP.SG.F me.DAT be.3SG Serbian  
 ‘[...] my best (girl)friend was a Serbian [...].’ (TZ)

This phenomenon cannot be explained by the syntactic properties of constituents. It is important to observe that hosts prefer even numbers of graphemes, since two and four are the modes, while the mode of the initial constituent is an odd number, 3. Since the numbers represent word length, it is clear that when DP occurs, the host type changes. The two-grapheme words are usually grammatical words such as pronouns or determiners.

- (38) [...] tako da [ja]<sub>host</sub> sam tu [...].  
 so that I be.1SG here  
 ‘[...] so that I am here [...].’ (MO1)
- (39) [...] do temelja [to]<sub>host</sub> je uništeno [...].  
 until foundation that be.3SG destroyed.PASS.PTCP  
 ‘[...] that is completely destroyed [...].’ (BJ)



The four-grapheme words are lexical words, for example verbs and particles like in (40) and (41).

- (40) [...] da to [vodi]<sub>host</sub> nas u propast [...].  
 that that lead.3PRS US.ACC in failure  
 ‘[...] that this leads us to failure [...].’ (BG1)
- (41) Moj muž [isto]<sub>host</sub> je ranjen [...].  
 my husband also be.3SG injured.PASS.PTCP  
 ‘My husband is also injured [...].’ (BJ)

Since no acoustic data are available, we could speculate that CL placement in the case of very short, one-word constituents might be highly phonologically motivated. As suggested by Diesing et al. (2009: 71f), it may be related to intonational contour.

## 8.9.5 Phrase splitting

### 8.9.5.1 Inventory of clitics participating in phrase splitting

We now proceed to the analysis of phrase splitting in the spoken Bosnian variety.<sup>26</sup> Out of 4106 annotated clauses with CLs, 260 contained a compound phrase as a potential CL host. In other words, 260 clauses in the corpus are potential contexts for phrase splitting. However, only every fifth clause ( $N = 53$ ) was actually split by a CL.

Figure 8.10 shows which CL types are used in the corpus as elements inserted into a phrase. Phrase splitting is possible mainly by verbal CLs ( $N = 31$ ). The most common type is phrase splitting with verbal CLs *su* ( $N = 11$ ) and *je* ( $N = 15$ ), as previously observed among others by Peti-Stantić (2005: 174f). We observe no difference between the behaviour of the CL *je* and other verbal CLs with respect to phrase splitting.

Further, in the case of pronominal CLs, phrase splitting is attested only with dative CLs ( $N = 5$ ). We find two occurrences of genitive CLs in the context of multiword phrases, but without phrase splitting. Our data confirm that reflexive CLs may also split a phrase ( $N = 5$ ). The differences between verbal and other CLs in the context of phrase splitting which can be seen in Figure 8.10 are presumably motivated by frequency. Namely, verbal CLs are generally more frequent. We

<sup>26</sup>More information on phrase splitting in written standard BCS varieties can be found in Section 6.5.5, whereas more information on phrase splitting in BCS dialects can be found in Section 7.6.3.

## 8 Clitics in a corpus of a spoken variety

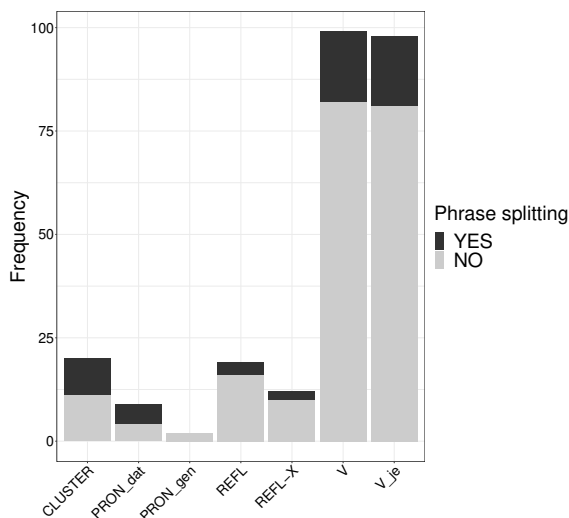


Figure 8.10: Inventory of CLs appearing in contexts where a phrase could be split

have no evidence for structural restrictions. For instance, accusative CLs were not attested in our data either as CLs which split phrases or as CLs which occur in the context of phrases which could be split but were not. Nonetheless, both variants can be easily retrieved from bsWaC.<sup>27</sup>

According to some authors (e.g. Progovac 1996, Radanović-Kocić 1988, 1996), clusters are not used as splitting elements. However, the corpus of spoken Bosnian contains seven occurrences of the cluster *mi je* (PRON\_dat + V\_je) inserted into a phrase, as in (42).

- (42) [Jedan *mi je* sin] bio otišao [...].  
 one me.DAT be.3SG son be.PTCP.SG.M leave.PTCP.SG.M  
 ‘One of my sons had left [...]’ (DJ1)

The possibility of phrase splitting with diacsis, as in (43), is not mentioned in the literature at all.

<sup>27</sup> An example of phrase splitting with the accusative CL *me* is given in (i):

- (i) [Moja *me* porodica] nije čula na dan muzičkog nastupa.  
 my me.ACC family NEG.be.3SG hear.PTCP.SG.F on day music performance  
 ‘My family did not hear me on the day of the musical performance.’ [bsWaC 1.2]

- (43) [...] [moja je mater] se udala [...].  
 my be.3SG mother REFL marry.PTCP.SG.F  
 '[...] my mother got married [...].' (VI)

### 8.9.5.2 Split phrases

Typical split phrases are subject noun phrases, consisting of a possessive attribute and a noun as in (44). This kind of phrase splitting is considered uncontroversial not only in standard Croatian, but also in standard Bosnian and Serbian.

- (44) [...] [moja je majka] mene rodila u bolnici [...].  
 my be.3SG mother me.ACC birth.PTCP.SG.F in hospital  
 '[...] my mother gave birth to me in the hospital [...].' (BR)

Furthermore, in the data we find split adverb phrases, similar to (45). This kind of phrase splitting is also found in standard BCS varieties.

- (45) [...] [vrlo su rijetko] išli u džamiju [...].  
 very be.3PL rarely go.PTCP.PL.M in mosque  
 '[...] they very rarely went to the mosque [...].' (IL)

However, splitting is not restricted to NP subject and adverb phrases only. In (46) the verbal CL *su* 'are' splits a modifier in the prepositional phrase *na istim* 'on same' from its noun *linijama* 'lines'.

- (46) Svaki puta kad zovem [na istim su linijama] [...].  
 every time when call.1PRS on same be.3PRS lines  
 'Every time I call, they are on the same lines (of front) [...].' (DO)

The example above clearly contradicts Radanović-Kocić (1996: 436), who claims that a sentence is ungrammatical when a CL is placed between a noun and its modifier in a prepositional phrase.

### 8.9.5.3 Clitic position and phrase splitting

The most frequent CL position within a split phrase is after the first stressed word ( $N = 46$ ). However, in the corpus of spoken Bosnian there are two cases where phrases which consist of more than two stressed words are split. In those utterances the verbal CL *je* is placed after the second stressed word of the phrase, as in (47).

- (47) [...] i [još mnogo je toga] izgorjelo [...].  
           and more much be.3SG that burn.down.PTCP.SG.N  
           ‘[...] and much more of that burned down [...].’ (KR)

When the CL is placed after the first word in the initial phrase, its position is naturally 2P = 2W. However, split phrases (even prepositional phrases) which are not initial are also possible, as shown in (48).

- (48) [...] [svako]<sub>phrase1</sub> [na svom je koritu]<sub>phrase2</sub> jači [...].  
           everybody on own be.3SG trough stronger  
           ‘[...] everyone is stronger on his own trough (on his own territory) [...].’ (VI)

Since phrase splitting is possible in DP, it is not necessarily motivated by 2P. However, this phenomenon is very rare, as in the whole corpus we have only six cases of non-initial phrase splitting.

## 8.10 Diaclisis

In this section we discuss the attested cases of diaclisis. Twenty-two such utterances are attested, three of which contain a matrix verb and its complement (pseudodiaclisis). As these numbers are small we restrict ourselves to some general observations without analysing frequencies. We identified the following combinations of CLs which do not form a cluster:

1. QUEST, V(erbal), REFL
2. V(erbal), PRON\_dat, REFL
3. V(erbal)\_je, REFL, PRON\_dat
4. QUEST, V(erbal)
5. V(erbal)\_je, PRON\_acc
6. V(erbal), PRON\_dat
7. V(erbal)\_je, REFL
8. V(erbal), REFL

We see that diaclisis and pseudodiaclisis always involve an interaction between a verbal CL and another CL type, most frequently a reflexive ( $N = 17$ ). Further, we observe that two (49) or three CLs (50) can appear in diaclisis. In the latter case two of them clusterise.

- (49) [...] *po gradovima su predsednici opština se odjednom*  
 in cities be.3PL presidents counties REFL suddenly  
*opredjeljivali [...]*  
 determine.PTCP.PL.M  
 ‘[...] in the cities, the county presidents were suddenly determining [...].’  
 (BH)
- (50) *Da li Bosna i Hercegovina će se osamostaliti [...].*  
 That Q Bosnia and Herzegovina FUT.3SG REFL become.independent.INF  
 ‘Will Bosnia and Herzegovina become independent [...].’ (DJ)

As mentioned in Section 2.4.5 we use the term pseudodiaclisis for matrix-embedding structures in which CC does not occur and a CL is present in the matrix. We found only three clear cases of pseudodiaclisis, including:

- (51) [...] *pa sam<sub>1</sub> uspio<sub>1</sub> se<sub>2</sub> izvuć’<sub>2</sub> kroz bašče [...].*  
 so be.1SG manage.PTCP.SG.M REFL extract.INF through gardens  
 ‘[...] so I managed to get myself out through the gardens [...].’ (BG1)

Note that we did not annotate pseudodiaclisis with *da*-complements. Interestingly, pseudodiaclisis was also attested in a CC utterance. In (52) the dative pronominal CL climbs from its infinitive complement, as it is placed before the particle *i*. However, it does not form a cluster with the verbal CL in the matrix. The possibility of such cases has not been reported before.

- (52) [...] *evo ja sam<sub>1</sub> trebao<sub>1</sub> vam<sub>2</sub> i donjet<sub>2</sub> baš*  
 here I be.1SG need.PTCP.SG.M you.DAT bring.INF really letter  
*pismo [...].*  
 ‘[...] here I should really have brought you the letter [...].’ (BR)

Finally, it is important to note that (pseudo)diaclisis does not seem to be linked to any fluencemes, i.e. specific structures typical of spoken language, which we discuss in the following section.

## 8.11 Impact of certain syntactic structures on clitic placement

### 8.11.1 Impact of structures occurring before clitics

As the first to address the impact of disfluency and structures typical of spoken language on CL positioning, we are able to present a few observations. Table 8.6 summarises the frequencies of individual types of occurrences.

Table 8.6: Special syntactic structures occurring before CL placement in the data in non-anonymised utterances. Values in brackets are frequencies relative to the frequency of a particular placement type.

Type of structure	1P <sup>a</sup>	2P <sup>b</sup>		DP <sup>c</sup>	
		<i>n</i>	%	<i>n</i>	%
repetition identical	1	32	1.9	2	1.1
repetition partial	0	20	1.2	3	1.6
aposiopesis	0	72	1.9	4	2.2
anacoluthon	0	74	2.0	6	3.3
retrospective EDU	7	213	5.8	18	9.9
inserted clause	5	120	3.3	0	0.0
rendered topic	0	56	1.5	2	1.1
DSE	8	390	10.7	18	9.9
substitution	0	30	0.8	1	0.5
false start	0	42	1.1	3	1.6
ellipsis	1	83	2.2	3	1.6
total	22	1132	31.2	42	23.2

<sup>a</sup>*N* = 29

<sup>b</sup>*N* = 3619

<sup>c</sup>*N* = 181

We annotated inserted clauses, two types of repetition, false starts, substitutions, rendered topics, retrospective EDUs, DSEs, omissions, and anacolutha. This is a necessary step for determining the position of the CLs in the clause correctly.

In the next lines we refer to the same sample as in Section 8.9, so that we can address the length of constituents. Most types of annotated structures are quite

### 8.11 Impact of certain syntactic structures on clitic placement

rare in the data, mostly accounting for around 1% of observations. Only retrospective EDUs and DSEs cross the threshold of 5%. All of the special syntactic structures appear in the context of DP or 2P. However, no difference in the relative frequency of the two types of placement can be inferred from the data. Only retrospective EDUs occur twice as often in DP clauses than in 2P clauses, and could therefore be a potential topic for further study.

- (53) Ali u školi tako, [u gimnaziji,] mnogo je bilo  
 but in school so in secondary.school many be.3SG be.PTCP.SG.N  
 Muslimana [...].  
 Muslims  
 ‘But in school like, in secondary school, there were many Muslims [...].’  
 (TZ)

With respect to 1P, we observe that CLs take it when they are preceded by inserted clauses ( $N = 5$ ; example (54)), DSEs ( $N = 8$ ) or retrospectives ( $N = 7$ ; example (55)).

- (54) [...] niko [ko god dobije tu vojnu obavezu] se ne  
 nobody who ever get.3PRS that military obligation REFL NEG  
 treba javljat [...].  
 need.3PRS apply.INF  
 ‘[...] nobody who gets invited to military service ever has to apply [...].’  
 (VI)

- (55) Jedan drug, [Musliman,] me je zvao [...].  
 one friend muslim me.ACC be.3SG call.PTCP.SG.M  
 ‘One friend, a Muslim, called me [...].’  
 (TZ)

We do not observe instances of absolute 1P defined as a true sentence-initial position. From this result, we can conclude that insertions, DSEs and retrospectives are the main triggers for 1P. This means that 1P is restricted to syntactic structures typical of spoken language.

When it comes to the usage of the question marker *li*, it is worth mentioning that speakers tend to avoid potential delays in its placement caused e.g. by long phrases, insertions and special syntactic structures. In fact, not a single insertion, special syntactic structure or any other element or category which could endanger the 2P of *li* is found. Only DSEs are used twice as in (19), which has already been discussed in Section 8.7.1:

## 8 Clitics in a corpus of a spoken variety

- (19) E, sad da l' će te biti mržnje, da l' neće, ne  
eh now that Q FUT.3SG this be.INF hatred that Q NEG.FUT.3SG NEG  
znam [...]  
know.1PRS  
'Eh, now, will there be hatred, or not, I don't know...' (VI)

### 8.11.2 Positioning of clitics within fluencemes

As mentioned in Section 8.7.1, 132 CLs in 117 clauses are repeated or substituted with other CLs. Disfluency involving CLs does not seem to have an impact on CL positioning. In repetitions involving CLs ( $N = 5$  for partial,  $N = 19$  for identical), the CL is never delayed. In 44 cases of substitutions we find only one CL in DP, which is caused by a very long initial constituent.

## 8.12 Summary

We can now answer our research questions presented in Section 8.3:

- A1: Our data from the corpus of spoken Bosnian do not reveal any major peculiarities in the inventory of CLs. Apart from cases of phonologically reduced forms, which can also be found in dialects, the inventory coincides with the inventory of standard Bosnian and Serbian varieties. We identified only one additional item, the  $REFL_{2ND}$  CL *si*, recognised only by the Croatian standard. This CL is very rare, which is in line with the claims made in the literature.
- A2: We have seen that in spoken Bosnian the most representative clusters consist of 2 elements (99%) and those with 3 components are the exception (1%). Combinations of 4 or more CLs, whether in clusters or in (pseudo)diacclisis, are not found at all. Further, in the corpus of spoken Bosnian language 18 different cluster combinations of 2 CLs and 6 different combinations of 3 CLs are attested. We have shown that the most frequent cluster combinations of two and three CLs include a pronominal CL in the dative or the verbal CL *je*.
- A3: Generally, CL ordering in clusters coincides with the ordering found in standard BCS varieties. There are, however, two divergent types of CL cluster ordering. The first involves the verbal CL *je* and the reflexive CL



*se*. Interestingly, when these CLs co-occur in a cluster, they are 4 times as likely to be attested in the non-standard CL order, that is, with *je* preceding *se*. The second case of diaphasic variation is the order of the verbal CL *je* and pronominal accusative CLs. Namely, we found examples of the verbal CL *je* preceding pronominal accusative CLs *me* and *ga*, the reverse of the order established in written standard Bosnian (and other standard varieties).

- A4: Verbal and reflexive CLs are by far the most frequent CLs in the whole corpus. However, the combination of *je* and *se* in a CL cluster is quite rare. We have shown that the Bosnian spoken variety prefers haplogologised structures, i.e. where only the reflexive CL *se* occurs while the verbal CL *je* is omitted. Furthermore, haplogy of unlikes is attested in every interview of the corpus. In contrast, there is not a single example of suppletion. Moreover, we did not find any occurrences of the pronominal accusative CL *ju*.
- A5: In spoken Bosnian 2P is the dominant CL position. Interestingly, we found no statistically significant difference in the placement of single CLs and CL clusters. They both show a strong tendency towards 2P. Moreover, the findings concerning differences in the placement of individual CL types are of major theoretical interest. First, the polar question marker *li* differs from all the other CL types as it is placed in 2P after one short word (2 to 4 graphemes) in 100% of cases. Its positioning is thus much more unified than the positioning of verbal, pronominal and reflexive CLs. Second, in our corpus 1P is restricted to verbal and reflexive CLs, as neither *li* nor pronominal CLs seem to allow it. Finally, we would like to emphasise that 1P is connected to contexts of insertions, DSEs and retrospectives; i.e., there are no utterances with absolute sentential 1P.
- A6: As to the nature of 2P in spoken Bosnian, we saw a strong tendency towards 2W: the CL occupies the position after the first word in 77% of all observations (single CLs and clusters). Our data provide interesting insights into the heaviness of a constituent in the spoken variety. The typical CL position in the clause is after the first word, which is most frequently two graphemes long. The most frequent initial constituent in DP is three graphemes long, but in general its length is not limited, while the most frequent host in DP is four graphemes long, making it longer than the initial constituent. However, even in the case of DP host length is more limited than the length of the initial constituent. The current study provides only statistical tendencies. In our view, two further types of investigations

should be undertaken in the future. First, studies based on acoustic data are necessary to allow examination of the role of phonological contour. Second, a study of written language should investigate whether similar statistical regularities can be obtained.<sup>28</sup>

A7: Our analysis of phrase splitting brings new insights, as we see that in the spoken Bosnian variety, like in the written Croatian standard, splitting is possible not only for 2P but also for DP. Further, in spoken Bosnian splitting occurs not only in subject, but also in prepositional phrases, which is in line with the findings of Diesing & Zec (2017) for written Serbian.<sup>29</sup> Contrary to the observations of some authors, we find that clusters can take part in phrase splitting. The examples attested involve possessive dative. We did not find a single example of an accusative or genitive pronominal CL taking part in splitting. There are also few examples of diaclysis with phrase splitting and even fewer examples of pseudodiaclysis.

A8: We have shown that the analysis of CL positioning in spoken language has to take into consideration disfluency phenomena and some discourse-organising structures because they affect the surface structure of the sentences. They have, however, no direct impact on CL positioning beyond 1P after insertions, DSEs and retrospectives. Neither do they trigger (pseudo)-diaclysis.

The restricted empirical base notwithstanding, we would argue that our small pilot study could serve as the point of departure for future studies on CL positioning in spoken languages, not only in BCS but also beyond. We have prepared a scheme for the annotation of disfluency and other phenomena typical of spoken language, which is a *conditio sine qua non* for the analysis of CL positioning in spoken language.

---

<sup>28</sup> Although Reinkowski (2001) analyses the positioning of CLs in newspapers and magazines, her results cannot be directly compared with ours. In her study, three CL positions are distinguished: initial (after the first word), middle (any position before the predicate, but not immediately after the first word) and final (behind the predicate), which does not coincide with our coding scheme. Additionally, many types of initial constituents allowed in our study fall outside the scope of Reinkowski's study.

<sup>29</sup> The reader should, however, bear in mind that Diesing & Zec (2017: 9f) differentiate between predicate and argument initial prepositional phrases, which a CL can split. The latter was accepted by more than 67% of participants in the acceptability judgment experiment, but it had very low scores in the production experiment (Diesing & Zec 2017: 9f). Therefore, Diesing & Zec (2017: 11f) ascribe the ungrammatical status to split prepositional arguments in Serbian. In contrast, we do not believe that such structures have an ungrammatical status in the spoken Bosnian variety.

# 9 Parameters of variation: Conclusions

## 9.1 Introduction

In this chapter we summarise the findings presented in the previous three chapters on microvariation in BCS grammaticography and dialects, and in the corpus of spoken Bosnian. We focus on the relationship between language use in spoken languages and standardisation processes which include selection and prescription of features. By taking a bird's eye view, we detect global patterns of microvariation which we discuss in the following sections. Section 9.2.1 provides an overview of the variation in the inventory. Section 9.2.2 presents variation within CL clusterisation and morphological changes which occur in CL clusters, followed by conclusions on variation in position of CLs or CL clusters. Section 9.2.3 discusses the absolute first position, Section 9.2.4, second position, delayed placement and phrase splitting, and Section 9.2.5, the heaviness of constituents. As already mentioned, clitic climbing will be dealt with separately in Part III because the data from the grammar handbooks, dialectological sources and our corpus of spoken language are too limited to allow for any sound conclusions. As we are aware that it might be hard to follow our discussion on microvariation detected between codified standard varieties on the one hand and BCS Štokavian dialects and the spoken variety of Bosnian on the other hand, we provide Figure 9.1. which shows the occurrence of the discussed features in the individual dialects or dialect groups. At the same time, readers who are not familiar with the geographical background can clearly see the borders of the ex-Yugoslavian countries. On the map we indicate only features from dialects. First, we use them to discuss which features present on the territory of certain countries were or were not selected to be features of the respective codified standard variety. Secondly, we compare features attested in dialects spoken on Bosnian territory with features attested in the spoken Bosnian variety. However, variation which was attested in spoken Bosnian is not indicated on Figure 9.1. Since data on the origins of interviewees was anonymised, we could not trace possible dialectal influences in their speech.

9 Parameters of variation: Conclusions

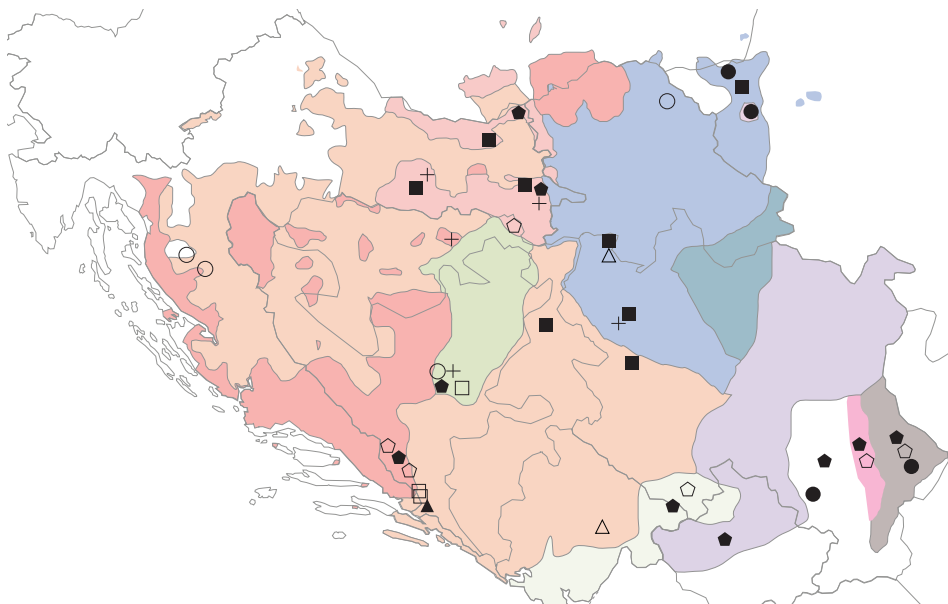






















Figure 9.1: Map showing the dialects and the distribution of selected features. Author: Branimir Brgles.

Table 9.1: Legend to Figure 9.1

	ISTOČNOHERCEGOVAČKI		CL <i>ju</i> (her. ACC)
	SLAVONSKI		CL <i>si</i> (refl. DAT)
	ZAPADNI		<i>je se</i> sequence
	SREDNJOBOSANSKI		absolute 1p
	ZETSKO-JUŽNOSANĐAČKI		conjunctions <i>a</i> and <i>i</i> followed by clitics
	ŠUMADIJSKO-VOJVODANSKI		delayed placement
	SMEDEREVSKO-VRŠAČKI		phrases with two content words followed by clitics
	SVRLJIŠKO-ZAPLANJSKI		splitting of PPs
	TIMOČKO-LUŽNIČKI		insertion of more CLs in a phrase
	KOSOVSKO-RESAVSKI		
	PRIZRENSKO-JUŽNOMORAVSKI		

## 9.2 Parameters of microvariation: Global patterns

### 9.2.1 Clitic inventory

Our findings on variation in the CL inventory can be summed up as follows. Our analysis of the standard grammar books shows differences in the inventory of the standard varieties. Only Croatian grammarians accept the reflexive CL *si* as standard. In this respect Ridjanović (2012: 440) even claims that the REFL<sub>2ND</sub> CL *si*, which is widely used in Croatian, can hardly be found elsewhere in BCS territory. The analysis of the dialectological literature, however, yields a more varied picture. Figure 9.1 clearly shows that the *si* form (transparent pentagon) is found not only on Croatian, but also on Bosnian and Serbian language territory. Namely, it is attested in scattered areas comprising some idioms of Montenegro, South Eastern Serbia, Western Herzegovina and Northern Bosnia. The presence of the REFL<sub>2ND</sub> CL *si* in idioms of Western Herzegovina and Northern Bosnia explains why this CL is also attested in the corpus of spoken Bosnian. Nevertheless, we would like to emphasise that our data suggest that this form is very rare in the spoken Bosnian variety.

Further, Croatian and Serbian authors differ in their recommendations for the usage of the third person singular feminine accusative CL *ju*. According to some Croatian authors, *ju* can be treated as a separate unit of the inventory, which is not restricted only to realisations of the CL cluster sequence with the third person singular feminine accusative and the third person singular present tense of the verb *biti* ‘be’. In contrast, in standard Serbian the third person feminine accusative pronoun *ju* can be used only in the case of suppletion. If we want to compare its situation in standard BCS varieties with the situation attested in dialects spoken in BCS language territory in Figure 9.1, we clearly see that the *ju* form (black pentagon) is attested in many idioms of Old, Middle and Neo-Štokavian dialects. The spatial distribution of the CL *ju* is not limited only to Croatian language territory. Moreover, it stretches from the West (*Zapadni, Sred-nobosanski*) to the Southeast (*Timočko-lužnički, Kosovsko-resavski* and *Prizrensko-južnomoravski*) and covers also Bosnian, Montenegrin and Serbian language territory. However, unlike in dialects spoken on Bosnian language territory, the CL *ju* is not attested at all in the data from the corpus of spoken Bosnian.

No variation as to the inventory of verbal CLs is found in standard BCS varieties. For the few varying forms that appear only in dialects, see Chapter 7. Notably, in many dialects the conditional auxiliary form *bi* is used for all persons (*Istočnohercegovački, Zapadni, Šumadijsko-vojvođanski, Slavonski* and *Kosovsko-resavski*). Moreover, our data from the corpus of spoken Bosnian corroborate

Peco's (2007b: 331) claim that the CL form *bi* used for all persons is spreading as a trait from dialects into spoken varieties.

At the end of this subsection we conclude that the selection and prescription of certain features related to CLs in standard BCS varieties do not correlate with their distribution on BCS language territory. Namely, although the CL *si* is present on Bosnian and Serbian language territory, it has not found its way into their standard varieties. Similarly the CL *ju* which, while attested in Bosnian and Serbian language territory, is restricted only to suppletion contexts in the relevant standard varieties. All three standard varieties are equally strict in their treatment of the conditional CL form *bi* used for all persons as a feature limited exclusively to non-standard varieties.

### 9.2.2 Clitic cluster and morphological processes

The maximum size of CL clusters in standard BCS varieties has been discussed by the Serbian authors Piper & Klajn (2014: 451f) and the Bosnian author Ridjanović (2012: 558). They claim that the CL cluster usually consists of two or three elements and that groups of five or more CLs are quite infrequent. Whereas there is still no solid empirical data on the maximum size of CL clusters in standard BCS varieties, in Chapter 8 we provide empirical data from the corpus of spoken Bosnian. Here, clusters consisting of only two elements are by far the most representative (99%). In contrast to the claims made for standard Serbian and Bosnian varieties, our empirical data show that in spoken Bosnian CL clusters with three components are an exception (1%). Moreover, strings of four or more CLs in a cluster are not attested at all.

Let us discuss the variation with respect to CL ordering in the cluster. The ordering of the reflexive CL *se* and the verbal CL *je* is a further clear case of microvariation. Whereas in standard Bosnian and Croatian both haplogy of unlikes and CL clusters with the sequence *je se* are allowed, the Serbian authors of a normative grammar book Piper & Klajn (2014: 452) acknowledge only the former as a feature of standard Serbian. Unlike in standard BCS varieties, both in dialects and in the spoken Bosnian variety we find ample evidence for the reversed CL order. The situation in BCS dialects with respect to the *je se* sequence is clearly visible on Figure 9.1: it is attested in central BCS territory of *Šumadijsko-vojvođanski*, *Zapadni*, *Slavonski*, *Srednjobosanski* and *Istočnohercegovački*. Thus, it is attested on Bosnian language territory. This is in accordance with the data from the corpus of spoken Bosnian where the *je se* cluster sequence is four times more frequent than *se je* prescribed in the standard Bosnian and Croatian varieties.

A second case of variation in CL ordering in the cluster concerns both diaphasic and diatopic variation. Namely both in BCS dialects (*Šumadijsko-vojvođanski* and *Svrljiško-zaplanjski*) and in the corpus of spoken Bosnian we find sentences in which the verbal CL *je* precedes pronominal accusative CLs. Since this order has been attested in the spoken Bosnian variety, we assume that it is very probably present in dialects spoken in Bosnian language territory too. Moreover, not only sentences in which the verbal CL *je* precedes accusative pronominal CLs, but also those with pronominal CLs in other cases are attested in various dialects: for more information and examples see Section 7.5.1.

Regarding the morphonological process of haplogy of unlikes in the context of co-occurrence of the reflexive CL *se* and the verbal CL *je*, we would like to put forward our empirical data from the corpus of spoken Bosnian. Although both haplogy of unlikes and the *se je* sequence are allowed in the standard Bosnian variety, the data from the corpus of spoken Bosnian show that haplogy of unlikes is far more common than the co-occurrence of these two CLs. Namely, we find haplogy (with only the CL *se* occurring) in 68.8% of cases, in 25.4% of cases a CL cluster (the CLs co-occur, the sequence *je se* is more frequent than *se je*) with *je*, and in 5.8% of cases the reflexive CL *se* and the verbal CL *je* appear in diaclisis.

### 9.2.3 Absolute first position and clitics after the conjunctions *a* and *i*

We start with our findings concerning absolute 1P, i.e., CLs which are placed at the beginning of the clause. According to the prescribed language norms of all three standard varieties, this CL position is ungrammatical. However, the dialectal map in 9.1 shows that absolute 1P (black circle) is attested in idioms of the *Šumadijsko-vojvođanski*, *Kosovsko-resavski*, *Prizrensko-južnomoravski* and *Timočko-lužnički* dialects. It is important to note that the former two are in language contact with Romanian, while the latter two are in language contact with Macedonian. We did not find CLs in the absolute 1P in the corpus of spoken Bosnian. We only came across sentences in which CLs follow insertions, DSEs and retro-spectives. These findings strongly suggest that absolute 1P is likely to occur in Štokavian contact varieties.

Our dialectological data indicate that at least some Štokavian idioms, including even idioms of *Istočnohercegovački*, allow CL positioning directly after the coordinative conjunctions *a* and *i*, unlike standard BCS varieties. Figure 9.1 shows that this feature, represented by a transparent circle, is also attested in *Šumadijsko-vojvođanski*, spoken mainly in Serbia.<sup>1</sup>

<sup>1</sup>As mentioned in Chapter 6, *Istočnohercegovački* served as a dialectal base for all three standard BCS varieties.

#### 9.2.4 Second position, delayed placement and phrase splitting

In this subsection we would like to highlight the following facts. While it seems that in standard Croatian and standard Bosnian the second position rule is understood as 2W, in the literature on standard Serbian it is emphasised that 2P is normally understood as the position posterior to the first phrase. Moreover, in the latter variety splitting the initial phrase is less preferred than placing CLs after initial compound phrases of two content words.

In contrast to standard Serbian, in which phrase splitting is uncontroversial only in very few cases, Croatian and Bosnian standards allow the insertion of CLs in far more contexts. Our dialectological data show that splitting of forename and family name, of conjoined NPs and of quantificational phrases is not only widespread in Bosnian and Croatian territory, but can also be found in Serbian language territory. Similarly, the corpus of spoken Bosnian contains ample evidence for phrase splitting. Moreover, we would like to emphasise that in the spoken Bosnian variety not only subject phrases, but also prepositional phrases can be split. As can be seen in Figure 9.1 (black triangle), the latter is also attested in the *Istočnohercegovački* dialect spoken on Bosnian language territory.

Furthermore, we would like to comment on the disagreement among theoretical syntacticians with respect to the number of CLs taking part in phrase splitting. Dialectal data from *Šumadijsko-vojvođanski* and *Istočnohercegovački* show that two CLs can be inserted into a phrase. Since this feature is attested in a dialect spoken on Bosnian language territory (see transparent triangle on Figure 9.1) it should come as no surprise that the corpus of spoken Bosnian also contains such instances. To conclude, both dialectal data and the data on the spoken Bosnian variety clearly contradict Progovac (1996) and Radanović-Kocić (1988, 1996), who claim that clusters are not used as splitting elements.

Moreover, BCS normativists disagree in their evaluations of DP. While Bosnian and Croatian authors recommend delaying the placement of CLs as a better alternative to placing CLs after compound phrases, Serbian authors propose quite the opposite. Delayed placement is widespread in dialects (see black square on Figure 9.1). We find such cases in the *Slavonski*, *Istočnohercegovački* and *Šumadijsko-vojvođanski* dialects spoken not only in Croatia and Bosnia, but also in Serbia.

#### 9.2.5 Heaviness of the initial constituent

Several authors mention the heaviness of the initial constituent as a factor responsible for DP. However, exact information on how to distinguish initial constituents which are heavy and cause DP from those which allow 2P can be found



neither in grammar books nor in the dialectological literature. Therefore, we conducted an empirical study based on the measurement of heaviness proposed by the Czech linguists Kosek et al. (2018). The chapter on spoken Bosnian provides some first hints on the heaviness of a constituent in the spoken variety. As to the nature of 2P in spoken Bosnian, we saw a strong tendency towards 2W; in 77% of all observations (single CLs and clusters) the CL occupies a position after the first word. The typical CL position in the sentence is after the first word, which is most frequently two graphemes long. The most frequent initial constituent in DP is three graphemes long, but in general its length is not limited, while the most frequent host in DP is four graphemes long, and thus longer than the initial constituent.



## Part III

# Clitic climbing



# 10 Approaches to clitic climbing

## 10.1 Introduction

Most works on CLs in Bosnian, Croatian, and Serbian address the nature of 2P effect mainly within formal theoretical frameworks (primacy of syntactic vs prosodic processes; for an overview see Chapter 2). One of the controversial issues in the literature concerns CC. An example of CC out of an infinitive complement is given in (1).

- (1) I mi *se*<sub>2</sub> planiramo<sub>1</sub> baciti<sub>2</sub> na posao.  
and we REFL plan.1PRS throw.INF at work  
'We also plan to throw ourselves into work.' [hrWaC v2.2]
- (2) Bojim<sub>1</sub> se<sub>1</sub> testirati<sub>2</sub> ih<sub>2</sub>.  
be.afraid.1PRS REFL test.INF them.ACC  
'I am afraid to test them.' [hrWaC v2.2]

CC occurs in constructions containing two or more verbal elements. In example (1), the reflexive CL *se* which belongs to the infinitive verb form *baciti* 'throw' is realised in the second position of the matrix clause. This is quite puzzling because the CL seems to have "climbed" from the infinitive complement into the matrix clause. Example (2), where the pronominal CL *ih* 'them' stays in the infinitive complement, shows that CC does not always occur. As we discuss in more detail below, CC is indeed a major source of variation in the usage of pronominal and reflexive CLs in BCS.

Part III of the book is dedicated to CC and its constraints in BCS, a hitherto underresearched topic. To our knowledge there are only four studies dealing specifically with CC in BCS (Caink 2004, Stjepanović 2004, Aljović 2004, 2005), and only the latter three address the question of constraints on CC. Besides these studies, some scattered information can be found in various works (e.g. Čamdžić & Hudson 2002, Todorović 2012), as we show later. In comparison, for Czech the syntactic conditions for CC are much better described by: Rezac (1999, 2005), Junghanns (2002), Dotlačil (2004), Rosen (2001, 2014), Hana (2007), Lenertová (2004), who discuss a whole series of constraints on CC in this West Slavonic

language.<sup>1</sup> Of them, Junghanns (2002) undoubtedly offers the most comprehensive account. In Section 2.4.3.2 we show that, unlike in Czech, next to strictly syntactic approaches (e.g. Progovac 1996, Franks 1997), also phonological (e.g. Radanović-Kocić 1988, 1996) and mixed approaches (e.g. Schütze 1994, Bošković 2000, 2001) to CL placement exist. However, as we discuss in Section 6.5.5, even the strict phonological approaches need to use syntax to explain the variation in CL placement. Moreover, some scholars argue against a purely phonological approach to the 2P phenomenon. Čavar & Wilder (1994: 441), for instance, question the assumption of phonological rules which have the power to move material around in phonological representations in order to capture marginal cases like phrase splitting. In a similar vein, we thus conclude that syntactic constraints on CL placement are relevant in BCS too. As we assume the CL systems of both languages to show many common features, we use the Czech constraints as a test ground for BCS (see Chapter 11).

We preliminarily define constraints on CC as structural features or combinations of features blocking the realisation in the matrix of a CL belonging to the embedding. Constraints *sensu stricto* can only be detected by testing minimal pairs, providing negative evidence where one sentence is evaluated as acceptable and the other as unacceptable. Nearly all the scholars who have worked on constraints on CC in BCS discussed below rely exclusively on their own linguistic intuition as native speakers, which may entail certain problems (see a more detailed discussion of those problems in Sections 3.1 and 3.3.3.5). This does not hold for the work for Czech of Junghanns (2002), which is based on examples found in corpora.

In part III of the book we first give a brief presentation of the main theoretical accounts of CC. Afterwards, we zoom in on the linguistic data. At first glance, the distribution of CC shows a confusingly high degree of variability. We follow the research scheme presented in Section 3.3.1: intuition/theory – observation – experiment. Part III of the book, dedicated to CC, has the following structure: On the basis of the existing research literature we present constraints in Czech and compare these data to BCS (intuition/theory) – see Chapter 11. We then present two empirical corpus studies on CC (observation). The studies are based on a common methodology, which we explain in Chapter 12.

First we present a corpus-based study of CC out of *da*<sub>2</sub>-complements which are characterised by the presence of an element sometimes interpreted as a complementiser and of an inflected verb (Chapter 13).<sup>2</sup> This is an interesting topic

<sup>1</sup>Cf. also Franks & King (2000: 247) on CC in Slovene. We will not take into account works on CC in Romance languages.

<sup>2</sup>For more information on *da*<sub>2</sub>-complements see Section 2.5.3.

because CC out of other complements with inflected verbs is a rare phenomenon. Here, we focus exclusively on Serbian because *da*<sub>2</sub>-complements are much more frequently used in Serbian than in Croatian, especially in the context of raising and subject control verbs.<sup>3</sup>

Next we present an empirical in-depth study on diaphasic variation with respect to the raising–control dichotomy and its impact on CC out of infinitive complements (Chapter 14).<sup>4</sup> This study focuses on corpora which contain texts with standard Croatian on the one hand and colloquial Croatian language features on the other. Two reasons motivated us to choose Croatian as our target language. First, in Croatian infinitive complements are used not only with raising, but also with subject and object control CTPs (which is not the case in Serbian and Bosnian). Second, only for Croatian are there electronically stored and publicly available big corpora compiled not only for colloquial, but also for standard language.<sup>5</sup>

On the basis of the observation chapters (with corpus studies) we proceed to Chapter 15. In that chapter we conduct a full-fledged psycholinguistic experiment consisting of acceptability judgment tasks. With this study we want to contribute new, experimentally collected data to CC research. Namely, our aim is to broaden the set of structures considered in accounts of CC in BCS. The general question which lies behind the psycholinguistic study is whether on the one hand any particular contexts can be recognised as triggers for obligatory CC, and on the other hand whether there are any features which can be detected as constraints on CC. Just like in the corpus study presented in Chapter 14, sentences in this test contain CTPs with infinitive complements only. As already mentioned, in contrast to Serbian and Bosnian where many CTPs favour *da*<sub>2</sub>-complements, in Croatian infinitive complements are used not only with raising, but also with control CTPs. Therefore, to match our corpus-linguistic data on CC out of infinitive complements, we conducted the psycholinguistic study exclusively for Croatian.

Next we discuss our data on haplology, clusters, and pseudodiaclisis. We bring together the findings from corpus studies and the experiment. We discuss what determines CC in BCS in terms of complexity in Chapter 16. The idea of system complexity offers a unified explanation for different types of constraints on CC in BCS without the necessity of assuming particular syntactic mechanisms like clause union or restructuring (see below), which by some authors is considered

---

<sup>3</sup>More information on the raising–control dichotomy can be found in Section 2.5.2.

<sup>4</sup>For more information on diaphasic variation, see Section 2.3.

<sup>5</sup>For an overview of corpora available for BCS see Chapter 4.

a sufficient, and by others only a necessary condition for CC. Although we do not negate the relevance of restructuring for CC, theory of complexity is in our view more adequate for the empirical data. It allows for explaining the considerable amount of variation we identify also in restructuring environments. This is achieved by incorporating non-systemic factors (diaphasic variation) into the model.

## 10.2 Theoretical approaches to clitic climbing

### 10.2.1 Definitions of clitic climbing

CC has been analysed only within formal frameworks. Surprisingly, there are no functional or cognitive accounts. In the theory-neutral survey of CL systems in different languages by Spencer & Luís (2012: 162), CC is linked 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 (for instance, the matrix verb which selects that subordinate clause), even though it may have no obvious semantic or syntactic connection to that verb”. We refrain from giving an overview of the research literature, confining ourselves to a small selection of definitions of CC by various authors:<sup>6</sup>

- Hana (2007: 122) on Czech: “In a clause, clitics governed by the highest non-clitic governor (usually a non-auxiliary finite verb, [...]) obligatorily occur in Wackernagel position – in the main clitic cluster. However, there can be other clitic clusters in the domain of more embedded phrases. Clitics governed by those words can, or even tend, under certain circumstances to occur in the clitic clusters of less embedded governors, possibly in the main one. Within a finite clause, clitics governed by infinitives [...], adjectives [...], adverbs, and numerals [...] can climb up into a higher clitic cluster.”
- Junghanns (2002: 58) on Czech: “In komplexen syntaktischen Ausdrücken bewegt sich ein klitisches Pronomen aus der Einbettung in die Matrix. Eindeutiges Indiz für die Bewegung ist die Tatsache, dass das der Einbettung

---

<sup>6</sup>For a literature overview, we refer readers to the abovementioned textbook Spencer & Luís (2012), and to the concise article Dotlačil (2017), which sums up the findings of works related to generative grammar and minimalism.



## 10.2 Theoretical approaches to clitic climbing

entstammende Pronomen in der Satzoberfläche links vom Matrixverb erscheint. Diese Bewegung von Klitika ist “Clitic Climbing” (CC) genannt worden [...].<sup>7</sup>

- Rezac (2005: 7) on Czech: “Clitic climbing refers to a phenomenon whereby the clitic argument of an infinitive shows up within the clause of a c-commanding verb.”
- Dotlačil (2004: 70) on Czech: “Clitic climbing (realization of a clitic in a clause higher than the one in which the clitic originates) [ ...].”
- Aljović (2004: 169) on BCS: “Clitic climbing is the phenomenon whereby clitics climb out of the clause containing the verb they are arguments of, and attach to a higher predicate.”
- Słodowicz (2008: 152) on Polish: “a process found in sentences in which the object of the complement predicate is realized by means of a pronominal clitic appearing in the matrix clause.”

If we abstract from the theoretical embedding of individual works, this small selection of definitions shows that most authors agree that CC is associated with matrix complement structures and with the positioning of a CL in the matrix clause and not in the complement in which it originates. If we discuss these definitions, we find the following pitfalls or even shortcomings:

- Hana (2007) and Rezac (2005) focus on infinitive complements because in Czech CC is attested exclusively out of them. Serbian, however, is different because it allows CC out of *da*<sub>2</sub>-complements containing a verbal form inflected for person and number.
- Spencer & Luís (2012) and Aljović (2004) assume that the complement forms a clause on its own. As the following section on restructuring shows, this is a contentious issue.
- Some authors assume a specific operation responsible for moving the element from one position to another (Hana 2007, Aljović 2004, Junghanns 2002) whereas others do not (Rezac 2005, Dotlačil 2004, Słodowicz 2008).

---

<sup>7</sup>In complex syntactic expressions, a clitic pronoun moves from the embedding into the matrix. A clear indication of the movement is that the pronoun originating in the embedding appears in the sentence surface left of the matrix verb. This movement of clitics is called “Clitic Climbing” (CC).

- There is no consensus as to the relationship between the CL and the verb of the complement: government (Hana 2007), CL as argument (Rezac 2005) or object (Słodowicz 2008). Other authors remain agnostic on this point. This is also our position because not only argumental pronominal CLs undergo CC; so do lexical reflexives like *se* in *smijati se* ‘laugh’, which can hardly be interpreted as objects.

Summing up, we propose the following definition: CLITIC CLIMBING (CC) refers to a phenomenon whereby a clitic is not realised in a position contiguous to elements of the embedding to which it belongs, but in a position contiguous to elements of the matrix.

### 10.2.2 Clitic climbing and optionality

It is well-known that formal syntactic models rely on the axiom of parsimony. Therefore, it comes as no surprise that many authors working in a formal framework try to treat CC as an ordinary case of CL placement. Namely, they attempt to explain away the peculiarities of CC in order to formulate a unified theory of cliticisation; e.g. Čamdžić & Hudson (2002: 350) see CC and cliticisation in general “as a very simple and natural extension of ordinary grammar.” In some approaches CC is associated with the syntactic process of “restructuring”, in others with “clause union”.<sup>8</sup>

Scholars differ when it comes to the relation between restructuring and CC. Some claim that restructuring is a necessary but insufficient condition for CC, while others are convinced that CC is contingent upon restructuring. Thus we can clearly see that there are two major streams in research on CC.

On the one hand, there are authors who claim that CC is always optional, which means that if the conditions for restructuring are fulfilled, CC can, but does not have to occur (for BCS see Progovac 1993b, Progovac 1996, Čavar & Wilder 1994, Stjepanović 2004, for Czech see Rezac 2005).<sup>9</sup> In probably the best-known paper dealing specifically with CC in BCS, Stjepanović starts out from

---

<sup>8</sup>The term “restructuring” is used in the analysis of certain infinitive complements which lack clausal properties when they appear as complements of restructuring verbs (cf. Aljović 2005: 2). For a more detailed discussion of restructuring and clause union see next section.

<sup>9</sup>These scholars are convinced that the existence of cases where other processes such as licensing of negative polarity items, object preposing and *wh*-movement out of complement signal restructuring are present while at the same time CLs do not climb demonstrates that CC is optional. For more on restructuring tests see Progovac (1994: 50–53) and Stjepanović (2004: 178–179).

the hypothesis that CC is not obligatory either out of infinitive or out of *da*<sub>2</sub>-complements (cf. Stjepanović 2004: 181, 186, 205). For more on technical details of this account see section below.

Aljović (2005) gives a comprehensive account of the discussion on the connection between CC and restructuring and on the most controversial question of whether CC is optional or obligatory. Aljović (2004, 2005) convincingly points out the weak spots of an approach that considers CC to be optional within the restructuring context. If the lack of restructuring is not the reason for the lack of CC, some special mechanisms of CL placement would have to exist, and various ad-hoc explanations would be called for (Aljović 2005: 3). A further consequence of this approach is the lack of a (unified) theory of cliticisation. Moreover, it fails to predict, or needs special solutions to explain cases of obligatory CC which is observed in some languages (cf. Aljović 2005: 3).

Aljović (2005: 6) addresses three questions: “Why is clitic climbing sometimes unavailable (blocking effects)? Why is sometimes clitic climbing obligatory? Why does clitic climbing sometimes appear to be optional?” She identifies “the size of the clausal complement” as the deciding factor in CC. “Clitics climb from domains that are functionally poor”, i.e., do not contain elements such as sentence negation and interrogatives (for more details see below; cf. Aljović 2005).<sup>10</sup> She suggests that there is no optionality for CC. CLs climb obligatorily in contexts of restructuring infinitives because these restructuring complements lack the functional structure necessary to keep CLs in their original phrasal domain.

### 10.2.3 Clitic climbing, restructuring (or clause union) and movement

In the following, we will give some technical theory-internal details showing how the authors implement the conceptual issues delineated in Section 10.2.2. in their models of grammar. Due to lack of space we cannot provide a full presentation of individual theories. This chapter is addressed to readers with some basic knowledge of minimalism or related frameworks.

There are many accounts of restructuring. The main idea is that there exist types of predicates which differ as to the complement they select. Restructuring predicates are found among modal verbs, motion verbs, aspectual verbs, causative verbs and some propositional attitude verbs. They vary not only cross-linguistically but also among speakers of one language (cf. Aljović 2005: 2).

---

<sup>10</sup>In this respect, she does not differ for instance from Rezac (2005) who claims that in Czech, CC is possible only from VP (verbal phrase) complements, while CPs (complementiser phrases) and TPs (tense phrases) do not allow climbing.

Progovac (1993b, 1996), Stjepanović (2004), Aljović (2004, 2005), and Todorović (2011, 2012, 2015) are, as far as we can see, the first to extend the notion of restructuring from infinitives to complements introduced by the element *da* containing a verbal form inflected for number and person, but not for tense: so-called *da*<sub>2</sub>-complements.<sup>11</sup>

Stjepanović (2004: 175–179) provides further data for the distinction of subjunctive-selecting (S-) and indicative-selecting (I-) verbs.<sup>12</sup> Long object preposing from passivised embeddings is possible with S-verbs, but not with I-verbs, as with the latter the passive reading is lost. With S-verbs, multiple wh-fronting with one wh-phrase originating in the matrix clause and another being licenced in the embedding is possible with any order of the wh-phrases, provided that the embedded subject is not overtly realised. With I-verbs, the structurally lower wh-phrase has to follow the matrix one. Stjepanović (2004: 179) concludes that S-verbs are restructuring verbs and that domain extension is restructuring.

As Stjepanović (2004: 186–198) notes, there are two major lines of thinking about restructuring: The first considers restructuring to be a transformational process, whereby two clauses are rearranged into one. According to the second, restructuring constructions are generated as a single clause from the beginning. Stjepanović argues for the latter based on different referential features of a singular subject and an embedded collective verb, compare (3a) and (3b).

- (3) a. Petar *je* odlučio da *se* okupe u parku  
 Petar be.3SG decide.PTCP.SG.M that REFL gather.3PRS in park  
 ‘Petar decided to gather in the park.’
- b. \* Petar *je* pokušao da *se* okupe u parku  
 Petar be.3SG try.PTCP.SG.M that REFL gather.3PRS in park  
 Intended: ‘Petar tried to gather in the park.’

(BCS; Stjepanović 2004: 193)

In (3a) the singular subject *Petar* is not strictly referential with the embedded subject of *se okupe* ‘(they) gather’, which has a collective meaning and can therefore be used with the verbal plural form *se okupe*. However, (3a) is grammatical in contrast to (3b), the latter showing a mismatch between the singularity of *Petar* and the collective semantics of *se okupe*. Referring to Wurmbrand (1999), Stjepanović takes such cases for determining what restructuring is. The reasoning behind this may be summed up as “one clause per subject argument”: The

<sup>11</sup>Some of them use other terminology like S- and I-verbs or subjunctive and indicative complements, but the idea behind the different terms is the same.

<sup>12</sup>In our terminology S-verbs are equivalent to verbs with *da*<sub>2</sub>-complements and I-verbs, to verbs with *da*<sub>1</sub>-complements. For more information on those differences see Section 2.5.3.

*da*-complement in (3a) has a phonetically empty subject, big PRO, which agrees with *okupe*. As subjects are licensed by clause, the *da*-complement builds a CP on its own. Hence, non-restructuring instances like (3a) are bi-clausal. Conversely, restructuring constructions like (3b) have only one subject, as indicated by their ungrammaticality, i.e. there is no PRO, as they are mono-clausal. From this it follows that CC is just an instance of regular CL movement. As CC is optional in restructuring contexts with both infinitive and *da*<sub>2</sub>-complements, Stjepanović (2004: 206) concludes that restructuring is not the driving force for CC. Furthermore, Stjepanović (2004: 198–204) observes that restructuring verbs behave like raising verbs. She also argues that under S-verbs *da* is not a real complementiser, but that it belongs to the verbal domain, similarly to the English infinitive marker *to* and German *zu* (Stjepanović 2004: 205f).

In her PhD thesis Todorović (2012) further elaborates on indicative and subjunctive complements and claims that CC out of *da*-complements is restricted to pronominal CLs which are hosted by *da* [–veridical] and is impossible in the case of *da* [+veridical]. Following Progovac (2005), Todorović (2012) eliminates CPs from the clausal structure of Serbian, thus giving rise to a mono-clausal analysis of *da*-complementation. Therefore, *da* is not viewed as a complementiser at all. Todorović assumes two different structural positions for the indicative and subjunctive *da* in the syntactic tree. In indicative *da*-complements, CLs successively move with the verb through all functional heads on its way from VP to TSP. Thus, auxiliary and pronominal CLs cluster together, while the lower copy of the verb is pronounced. In subjunctive complements, CLs climb on their own, since [–veridical] verbs lack tense and hence do not move to TSP for checking purposes. At this point, CLs attach to the matrix verb. According to Todorović (2012), there is no CC at all, because the notion of climbing is based on the assumption of CLs moving from an embedded to a matrix clause. However, the elimination of clausal boundaries between matrix verb and non-veridical *da*-complement allows CC to be interpreted in the more general terms of CL positioning within the clause.

Progovac (1993b, 1996) proposes that BCS CLs are right-adjoined to the head of the CP. Thus, 2P CLs always appear in the second position when being hosted by material that appears either in the specifier position of a CP or in the C-head. CC is considered to be the effect of domain extension by S-verbs, which is technically the deletion of the embedded clause's CP or inflectional phrase (IP). As a result, CLs right-adjoin to the head of the matrix CP. By contrast, I-verbs do not extend their domain and the embedded CP is preserved, so CLs cannot climb and thus remain in situ.

Franks & King (2000: 245) consider CC to be associated with restructuring, in that the matrix and the embedded verbs' domains are combined into one. CLs are then positioned with respect to this single domain. Restructuring is considered to be a lexical matter in principle and may be optional, obligatory, or impossible, with respective implications for CC.

Rezac (2005) discusses the syntactic structure which is necessary for CC in Czech, demonstrating that CC is possible only when the constituent out of which a CL climbs is a bare VP complement and showing that CPs, TPs and "small vPs" (verb shells) do not allow CC. He argues that CC depends on restructuring contexts of raising and control verbs (both subject and object), with CC taking place only with restructuring infinitives but not with non-restructuring infinitives, where CLs remain in situ. With restructuring infinitives CLs climb as arguments in order to check case and  $\phi$ -features such as person or number. Furthermore, he argues that these conditions affect both CLs and full NPs equally (cf. Rezac 2005).

Rosen (2014) who analyses CC of reflexives and their haplology within HPSG, uses clause union as an independently motivated mechanism for explaining mainly word-order phenomena. According to this solution, CLs may climb due to optional raising of arguments. Argument raising is a lexically specified option.

A different approach to CC is offered by Junghanns (2002), who is the only author who systematically studies the environments enabling or blocking CC. His approach does not rely on restructuring or any other highly abstract notion, but on a mechanism of CL movement which is susceptible to information structure. According to him, CC is possible in Czech in raising, subject control matrix clauses, and in ECM environments, where the climbing CL is generated in an infinitive embedding. Junghanns (2002: 66) notes that CLs may climb from complements as well as from adjuncts and argues that there is therefore no clause union in instances of CC. As heads, CLs may only climb if there is a free verbal head above the embedded infinitive that they can use as a landing site (cf. Junghanns 2002: 85f). The lack of a free verbal head blocks CC, so CL movement from subordinations under NPs, APs, and PPs is blocked. Junghanns (2002) treats instances of CC out of infinitive complements under noun/determiner and predicative adjective phrases as special cases of incorporation into the verbal head. However, Junghanns (2002: 82f) admits that syntax proper does not explain the seeming optionality of CC in Czech. He proposes using information structure in order to explain CC and in situ realisation of CLs. Namely, he suggests that CC takes place if the CL belongs to the background of the sentence. In contrast, CLs remain in the infinitive phrase (in situ) if they are part of the topic or focus.

### 10.2.4 Outlook

Summing up, the presented works on CC revolve around the (non-)obligatoriness of CC and the attempts to reconcile CC with a unified account of CL positioning. CC involves complex structures containing two verbal elements. One popular idea is to assume a process which unites two clauses into one. If there is no boundary between the matrix and the embedded structure, CLs do not climb, but appear in their usual, second position. Many authors, even those who do not refer to clause union or restructuring in their argumentation, assume that the embedding has a functionally poor structure. Moreover, these authors point out that CC is possible only with specific types of matrix predicates like restructuring verbs or S-verbs. The extension of this class of predicates, however, remains unclear. The distinction between raising and control seems to be relevant. Further, there is no consensus as to the structure of the embedding. An empirically more adequate approach explains CC in relation to the syntactic environments blocking landing sites for the movement of CLs.

Generally speaking, no theory of CC as such is presented. As CLs show up in positions where according to the formal models they should not, the authors discuss how this aberrant behaviour can be reconciled with the axioms of their respective model. Factors beyond sentence structure in formal models are not taken into account.

It is not an exaggeration to say that all theoretical accounts of CC in BCS are based on heavy data reduction in the sense that they are based on a small selection of syntactic environments where CC occurs. We see that most authors focus mainly on instances of climbing by single CLs and do not address the potential interaction between different types of CLs. Neither do they ask whether there are differences between, for example, pronominal and reflexive CLs. A further open question concerns different types of matrix predicates. It is Junghanns (2002) who strives for a complete account of environments enabling or blocking CC in Czech. As his approach refrains from assuming a single highly abstract mechanism, it covers a much wider range of data than the other models. Therefore, it will serve as the basis for the following chapter on constraints on CC in Czech and BCS.





# 11 Constraints on clitic climbing in Czech compared to BCS: Theory and observations

## 11.1 Introduction

Since the environments where CC occurs in BCS match with CC constructions in better-studied cCL languages (Ćavar & Wilder 1994: 448), we start out with claims concerning constraints on CC in Czech and the very few constraints which can be found in the literature about CC in BCS. We approach the topic in the following steps. First, taking the state of the art predominantly concerning Czech as a point of departure, we try to apply the putative constraints to BCS by looking for similar structures or counterexamples in our database and by querying {bs,hr,sr}WaC directly.<sup>1,2</sup> We focus on structural contexts found both in Czech and in BCS, excluding the usage of CLs in Czech structures which are not attested in BCS.<sup>3</sup> The findings from the literature and our first tentative corpus data are used to formulate further hypotheses regarding possible constraints on CC. These qualitative data are our first source of observation. Second, in order to gain some sort of negative evidence, we examine some of the examples extracted from corpora which were permuted, and then tested by at least five native speak-

---

<sup>1</sup>For a detailed description of the database, see Chapter 12.

<sup>2</sup>For more information on the corpora selected and our argumentation for choosing those and not other corpora, see Section 4.6.3. For the queries used see Section 12.2 and for an exhaustive discussion of our methodological approach see Section 3.3.

<sup>3</sup>Furthermore, we do not elaborate on the difference in the structure of control complements proposed by Rezac (2005). These are not constraints in a narrow sense, since they do not prevent CC per se, but lead to certain semantic and temporal differences when CC does occur. In other words, there are cases in which Czech sentences with and without CC have different semantic interpretations. However, we believe that studying semantic and temporal differences between sentences with and without CC in BCS should be a separate study, which actually cannot be conducted before the syntactic conditions for CC have been described well. That is why we neither report Rezac (2005) observations for Czech nor compare them with data from BCS. Readers who are interested in the subject can consult Rezac (2005).

ers (informal acceptability judgments).<sup>4,5</sup> Although at this point we do not address the question of diatopic variation, we look for corresponding data in all three varieties: Bosnian, Croatian, and Serbian. For the sake of brevity, however, we present only one or two examples for one structure. If not stated otherwise, we found examples of corresponding structures in all three languages.

As stated in Section 1.3 our aim is to give a maximally adequate descriptive account of the variation in CC that we are able to detect in natural data. Following this empirical orientation, we refrain from offering our own theoretical account of the sentence structure and confine ourselves to a list of putative constraints.

For the time being we propose six basic types of constraints. Their presentation is structured as follows. In Section 11.2 we present island constraints, while Section 11.3 introduces constraints which are connected to the raising and control dichotomy of CTPs. Next, in Section 11.4 we discuss constraints related to the inner structure of the mixed CL cluster. Constraints connected to the way CLs climb are taken into consideration in Section 11.5 Furthermore, one constraint linked to sentential negation is presented in Section 11.6 Finally, some constraints related to information structure can be found in Section 11.7.

We discuss how the range of constraints on CC we found can be accounted for and what the relation between selected constraints is, i.e. whether one constraint can be deduced from another, in Chapter 16.

## 11.2 Island constraints

### 11.2.1 Infinitives in comparative sentences with *než/nego*

Several authors point out that certain types of phrases seem to defy CC. Franks & King (2000: 245) call phrases showing some sort of locality constraint “islands for clitic climbing”. The term goes back to Ross (1967), who is known to have coined a major number of syntactic terms.

Junghanns (2002: 76) observes for Czech that there is no CC from infinitive comparative complements and idiomatic phrases with *než* ‘than’. In example (1a) the pronominal CL *ho* ‘him’ governed by the infinitive *držet* ‘hold’ cannot climb into the matrix clause: compare it with its unacceptable permutation (1b).

---

<sup>4</sup>For more information on informal acceptability judgments see Section 3.1.

<sup>5</sup>The most prospective hypotheses based on both sets of data (from corpora and informal acceptability judgments) underwent further rigorous tests that used psycholinguistic methodology: see Section 15.2.

- (1) a. Je<sub>1</sub> pro tu chvíli snadnější<sub>1</sub> vyměnit<sub>2</sub> děcku plenu, než  
 be.3SG for this moment easier change.INF baby diaper than  
 ho<sub>3</sub> držet<sub>3</sub> na hrnečku a mluvit na ně.  
 him.ACC hold.INF on potty and talk to him.DAT
- b. \*Vyměnit děcku plenu ho<sub>3</sub> je<sub>1</sub> snadnější<sub>1</sub> než držet<sub>3</sub>  
 change.INF baby diaper him.ACC be.3SG easier than hold.INF  
 na hrnečku a mluvit na ně.  
 on potty and talk to him.DAT
- ‘For the moment it is easier to change the baby’s diaper than to hold him on the potty and talk to him.’ (Cz; Junghanns 2002: 76)

Examples similar to the one in (1a) can easily be found in {bs,hr,sr}WaC. Their permutations with CC (2b)–(4b) are, as expected, unacceptable.

- (2) a. Nisam<sub>1</sub> imao<sub>1</sub> izbora<sub>1</sub> nego prodati<sub>2</sub> ga<sub>2</sub>.  
 NEG.be.1SG have.PTCP.SG.M choice than sell.INF him.ACC
- b. \*Nisam<sub>1</sub> ga<sub>2</sub> imao<sub>1</sub> izbora<sub>1</sub> nego prodati<sub>2</sub>.  
 NEG.be.1SG him.ACC have.PTCP.SG.M choice than sell.INF
- ‘I had no choice but to sell him.’ [bsWaC v1.2]
- (3) a. Što sam<sub>1</sub> drugo mogao<sub>1</sub>, nego poslušati<sub>2</sub> ga<sub>2</sub> [...].  
 what be.1SG else can.PTCP.SG.M than listen.INF him.ACC
- b. \*Što sam<sub>1</sub> ga<sub>2</sub> drugo mogao<sub>1</sub>, nego poslušati<sub>2</sub> [...].  
 what be.1SG him.ACC else can.PTCP.SG.M than listen.INF
- ‘What else could I do but listen to him [...].’ [hrWaC v2.2]
- (4) a. Nema<sub>1</sub> većeg<sub>1</sub> smora<sub>1</sub> i uzaludnijeg<sub>1</sub> procesa<sub>1</sub> nego  
 NEG.have.3SG bigger annoyance and more.useless process than  
 skupljati<sub>2</sub> ih<sub>2</sub> u jedno, [...].  
 collect.INF them.ACC in one
- b. \*Nema<sub>1</sub> ih<sub>2</sub> većeg<sub>1</sub> smora<sub>1</sub> i uzaludnijeg<sub>1</sub>  
 NEG.have.3SG them.ACC bigger annoyance and more.useless  
 procesa<sub>1</sub> nego skupljati<sub>2</sub> u jedno, [...].  
 process than collect.INF in one
- ‘There is no bigger annoyance and more useless process than collecting them into one [...].’ [srWaC v1.2]

These first data suggest that clauses introduced by the comparative particle *než/ nego* are islands for CC in both Czech and BCS.



## 11.2.3 Phrases with gerunds

Junghanns (2002: 70f) shows that in Czech there is no CC out of phrases with gerunds (transgressives).<sup>7</sup> He provides the following Czech example (8a), and its permutation (8b) in which the reflexive CL *se* cannot climb out of the phrase with the gerund *opírajíce* ‘leaned’.

- (8) a. Později oba usnuli<sub>1</sub>, opírajíce<sub>2</sub> se<sub>2</sub> o sebe  
 later both fall.asleep.PTCP.PL.M lean.PTCP.PRS.PL REFL ON REFL  
 hlavami.  
 head
- b. \* Později se<sub>2</sub> oba usnuli<sub>1</sub>, opírajíce<sub>2</sub> o sebe  
 later REFL both fall.asleep.PTCP.PL.M lean.PTCP.PRS.PL ON REFL  
 hlavami.  
 head

‘Later both of them fell asleep with their heads leaning on each other.’

(Cz; Junghanns 2002: 70)

In BCS gerunds (adverbial participles) are stylistically restricted. Below we adduce examples with the present (9a) and the past (10a) adverbial participle. Examples such as the one presented in (10a) would suggest that a CL governed by an adverbial participle can move away from it, since the CL, in this case the reflexive *se*, is placed to the left of its governor, here the present adverbial participle *žalečí* ‘complaining’. However, according to informal acceptability judgments of permuted examples (9b), (10b), CLs cannot climb from the adverbial participles into the main clause.<sup>8</sup>

- (9) a. [...] kako sa zanimanjem razgledavaju<sub>1</sub> eksponate nimalo se<sub>2</sub>  
 how with interest look.at.3PRS exhibits not.at.all REFL  
 ne žalečí<sub>2</sub> na [...].  
 NEG complaining.PTCP.ADV.PRS ON
- b. \* [...] kako se<sub>2</sub> sa zanimanjem razgledavaju<sub>1</sub> eksponate nimalo  
 how REFL with interest look.at.3PRS exhibits not.at.all

<sup>7</sup>Due to lack of space we cannot discuss the terms gerund and adverbial participle. Suffice to point out that the Czech forms (transgressives) show agreement whereas the BCS equivalents do not.

<sup>8</sup>Example (9a) shows that within the adverbial participle phrase the CL does not necessarily follow the verb as claimed by Čavar & Wilder (1994: 446f).

## 11 Constraints on clitic climbing in Czech compared to BCS

- ne žalečí<sub>2</sub> na [...].  
 NEG complaining.PTCP.ADV.PRS on  
 ‘[...] how they look at the exhibits with interest, not complaining at all about [...].’ [hrWaC v2.2]
- (10) a. Slično *su*<sub>1</sub> *mu*<sub>1</sub> ponovili<sub>1</sub> moj sin Senad i učenik  
 similar be.3PL him.DAT repeat.PTCP.PL.M my son Senad and student  
 M.R. zamolivši<sub>2</sub> *ga*<sub>2</sub> da [...].  
 M.R. ask.PTCP.ADV.PST him.ACC that
- b. \*Slično *su*<sub>1</sub> *mu*<sub>1</sub> *ga*<sub>2</sub> ponovili<sub>1</sub> moj sin Senad  
 similar be.3PL him.DAT him.ACC repeat.PTCP.PL.M my son Senad  
 i učenik M.R. zamolivši<sub>2</sub> da [...].  
 and student M.R. ask.PTCP.ADV.PST that  
 ‘My son Senad and the student M.R. repeated something similar, asking him to [...].’ [srWaC v1.2]

As we can see from the examples above, there is no doubt that the constraint noticed by Junghanns (2002) for Czech is relevant in the case of BCS as well (cf. Čavar & Wilder 1994: 447). Adverbial participles prevent both reflexive and pronominal CLs from climbing.

### 11.2.4 Adjective phrases

Junghanns (2002: 71) points out that adjective phrases lack the feature of finiteness. If an adjective has a CL as a complement, this CL will not be able to climb out of the adjective phrase in which it was generated. This holds at least for adjective phrases in attributive position preceding a noun phrase. Below is Junghanns’ (2002) Czech example (11a) and its permutation (11b), in which the reflexive CL *si* cannot climb out of the adjective phrase *neznámý člověk* ‘unknown man’.<sup>9</sup>

<sup>9</sup>Alexandr Rosen (p.c.) warned us that example (11a), from the Czech writer Ludvík Vaculík, sounds very odd and that Vaculík often uses his native Moravian dialect of Czech. According to Rosen, a much better version of the same sentence in standard Czech would be [...] *vyšel jsem z telefonní budky jako sobě neznámý člověk*. However, Rosen does not dispute Junghanns’ observation that CLs cannot climb out of adjective phrases and offers a better example for the same constraint:

- (i) a. Ze dvora bylo<sub>1</sub> slyšet<sub>2</sub> křik hrajících<sub>3</sub> *si*<sub>3</sub> děti.  
 from courtyard be.PTCP.SG.N hear.INF scream playing REFL children  
 b. \*Ze dvora *si*<sub>3</sub> bylo<sub>1</sub> slyšet<sub>2</sub> křik hrajících<sub>3</sub> děti.  
 from courtyard REFL be.PTCP.SG.N hear.INF scream playing children

‘You could hear the shouts of children playing in the courtyard.’

- (11) a. [...] vyšel<sub>1</sub> jsem<sub>1</sub> z telefonní budky jako neznámý<sub>2</sub>  
 go.out.PTCP.SG.M be.1SG out phone booth as unknown  
 si<sub>2</sub> člověk.  
 REFL man
- b. \* Vyšel<sub>1</sub> jsem<sub>1</sub> si<sub>2</sub> z telefonní budky jako neznámý<sub>2</sub>  
 go.out.PTCP.SG.M be.1SG REFL out phone booth as unknown  
 člověk.  
 man
- ‘[...] I came out from the phone booth as a man unknown to myself.’  
 (Cz; Junghanns 2002: 71)

Our corpus data suggest that the same constraint is found in BCS: see example (12a) and its unacceptable permutation (12b) below.

- (12) a. [...] radim<sub>1</sub> ritmom ponuđenog<sub>2</sub> mi<sub>2</sub> rada<sub>2</sub> za opstanak.  
 work.1PRS rhythm offered me.DAT work for survival
- b. \* [...] radim<sub>1</sub> mi<sub>2</sub> ritmom ponuđenog<sub>2</sub> rada<sub>2</sub> za opstanak.  
 work.1PRS me.DAT rhythm offered work for survival
- ‘[...] I work according to the rhythm of the job offered to me for survival.’  
 [bsWaC v1.2]

However, as Junghanns (2002: 72) points out, adjective phrases in predicate position do allow for the extraction of CLs, like in Czech example (13) in which the dative CL *mu* ‘him’ is placed to the left of its governor *vděčný* ‘grateful’. Nevertheless, we would like to point out that this is not a case of CC *sensu stricto* because we are dealing with a clearly mono-clausal structure with a single predicative element.

- (13) Libor *mu*<sub>1</sub> byl<sub>1</sub> za dotaz v duchu vděčný<sub>1</sub>.  
 Libor him.DAT be.PTCP.SG.M for question in spirit grateful  
 ‘In his mind Libor was grateful to him for the question.’  
 (Cz; Junghanns 2002: 72)

Our preliminary data from {bs,hr,sr}WaC suggest that the same holds for BCS; see examples (14)–(16).

- (14) I još ću<sub>1</sub> mi<sub>1</sub> biti<sub>1</sub> zahvalan<sub>1</sub>.  
 and also FUT.1SG him.DAT be.INF grateful  
 ‘And I will also be grateful to him.’  
 [bsWaC v1.2]

## 11 Constraints on clitic climbing in Czech compared to BCS

- (15) Sutra *ćeš<sub>1</sub>* *mi<sub>1</sub>* *biti<sub>1</sub>* *zahvalna<sub>1</sub>*.  
tomorrow FUT.2SG me.DAT be.INF grateful  
'Tomorrow you will be grateful to me' [hrWaC v2.2]
- (16) Toliko *sam<sub>1</sub>* *im<sub>1</sub>* *bila<sub>1</sub>* *dužna<sub>1</sub>*.  
that.much be.1SG them.DAT be.PTCP.SG.F in.debt  
'I was so much in their debt.' [srWaC v1.2]

As is apparent from the previous examples, in Czech and in BCS adjective phrases in predicate position allow CLs to climb.

### 11.2.5 Depth and kind of embeddedness of infinitive phrases

In this subsection we will discuss several constraints which depend on the depth and kind of embeddedness.

#### 11.2.5.1 Infinitives as complements of nouns

Infinitives complementing nouns are still a somewhat unclear case.<sup>10</sup> Junghanns (2002: 72) argues that normally CC does not occur out of infinitives embedded in a determiner phrase: compare Czech example (17a) and its permutation (17b).<sup>11</sup> However, at the same time he admits that counterexamples can still be found, such as (18) (cf. Junghanns 2002: 73).

- (17) a. Nemám<sub>1</sub> právo<sub>1</sub> ti<sub>2</sub> bránit<sub>2</sub>.  
NEG.have.1SG right you.DAT restrain.INF

<sup>10</sup>Junghanns (2002: 73) uses the German term "Funktionsverbgefüge" (light verb construction).

<sup>11</sup>Alexandr Rosen (p.c.) disagrees with Junghanns' observation regarding this example: such structures have multiple attestations in the Czech National Corpus, and additionally he as a native speaker finds them acceptable.

- (i) Policisté *mi* dali neoprávněně botičku, pokutu, neměli<sub>1</sub>  
Policemen me.DAT give.PTCP.PL.M unjustified ticket fine NEG.have.PTCP.PL.M  
*mě<sub>2</sub>* právo<sub>1</sub> zastavit<sub>2</sub> [...].  
me.ACC right stop.INF  
'Policemen gave me an unjustified ticket, a fine, they had no right to stop me [...].'  
[Czech National Corpus]
- (ii) Neměli<sub>1</sub> *ho<sub>2</sub>* právo<sub>1</sub> dát<sub>2</sub> [...].  
NEG.have.PTCP.PL.M it.ACC right give.INF  
'They did not have any right to give it [...].'  
[Czech National Corpus]



- b. \*Nemám<sub>1</sub> ti<sub>2</sub> právo<sub>1</sub> bránit<sub>2</sub>.  
 NEG.have.1SG you.DAT right restrain.INF  
 ‘I do not have any right to restrain you.’ (Cz; Junghanns 2002: 72)
- (18) Já jsem<sub>1</sub> mu<sub>2</sub> to neměla<sub>1</sub> čas<sub>1</sub> vysvětlit<sub>2</sub>.  
 I be.1SG him.DAT that NEG.have.PTCP.SG.F time explain.INF  
 ‘I did not have time to explain it to him.’ (Cz; Junghanns 2002: 73)

He offers a possible explanation for the discrepancy in the acceptability of examples presented in (17b) and (18). Namely, he suggests that CC is possible only in the context of CTPs in which the verbal part has almost no descriptive content while the nominal part contains substantial descriptive content (18). If, however, both the nominal and the verbal part of the construction contain descriptive content, CC is claimed to be blocked (17b).

Here, we must emphasise that infinitives which are an adjunct or complement to a noun were recognized as general islands for CC in Croatian by Čavar & Wilder (1994: 448f) well before Junghanns (2002). However, as corpus data show, it seems that BCS does allow CC not only with light verb constructions like *imati/nemati pravo* ‘be right/wrong’, i.e., cases in which only the noun has descriptive content (19), but also with constructions like *pasti na um/pamet* ‘cross one’s mind’ in which both the noun and the verb have descriptive content (20). Compare also sentences with light verb constructions from {bs,sr}WaC in (21a)–(22a) and their acceptable permutations (21b)–(22b).

- (19) [...] a ti ga<sub>3</sub> imaš<sub>1</sub> pravo<sub>1</sub> odbiti<sub>2</sub> dati<sub>3</sub>.  
 and you it.ACC have.2PRS right refuse.INF give.INF  
 ‘[...] and you have the right to refuse to give it.’ [hrWaC v2.2]
- (20) [...] i ne pada<sub>1</sub> mi<sub>1</sub> je<sub>3</sub> na<sub>1</sub> pamet<sub>1</sub> ići<sub>2</sub> buditi<sub>3</sub>.  
 and NEG fall.3PRS me.DAT her.ACC on mind go.INF wake.INF  
 ‘[...] and it does not cross my mind to go and wake her up.’ [hrWaC v2.2]
- (21) a. Neki su<sub>1</sub> imali<sub>1</sub> potrebu<sub>1</sub> braniti<sub>2</sub> ga<sub>2</sub> od [...].  
 some be.1PL have.PTCP.PL.M need defend.INF him.ACC from  
 b. Neki su<sub>1</sub> ga<sub>2</sub> imali<sub>1</sub> potrebu<sub>1</sub> braniti<sub>2</sub> od [...].  
 some be.1PL him.ACC have.PTCP.PL.M need defend.INF from  
 ‘Some had the need to defend him from [...]’ [bsWaC v1.2]
- (22) a. Naime, mozak nije<sub>1</sub> u<sub>1</sub> stanju<sub>1</sub> prebaciti<sub>2</sub> ih<sub>2</sub> iz  
 namely brain NEG.be.3SG in state switch.INF them.ACC from  
 kratkoročnog u dugoročno pamćenje [...].  
 short in long memory

## 11 Constraints on clitic climbing in Czech compared to BCS

- b. Naime, mozak  $ih_2$  nije $_1$  u $_1$  stanju $_1$  prebaciti $_2$  iz  
 namely brain them.ACC NEG.be.3SG in state switch.INF from  
 kratkoročnog u dugoročno pamćenje [...].  
 short in long memory  
 ‘Namely, the brain is unable to move them from short term  
 memory [...].’ [srWaC v1.2]

As to BCS, our small selection of examples and their permutations seems to contradict Junghanns’ explanation. However, we would like to point out that neither for Czech nor for BCS is it known exactly which light verb constructions, i.e. infinitives as complements of a noun, allow and which block CC. This indicates that CC in the context of infinitives complementing nouns still needs to be investigated both in Czech and in BCS.

### 11.2.5.2 Infinitives as complements of nouns in prepositional phrases

A case related to but slightly different from the one mentioned in the previous subsection concerns infinitives which are complements of a noun in a prepositional phrase: see Czech example (23a). In this example the infinitive *přimět* ‘bring’ is a complement of the noun in the prepositional phrase *se snahou* ‘with aim’. According to Junghanns (2002: 75), CC is blocked in such cases.

- (23) a. [...] zeptal $_1$  se [se snahou] $_2$  přimět $_3$  ho $_3$  k  
 ask.PTCP.SG.M REFL with aim bring.INF him.ACC to  
 odpovědi.  
 answer  
 b. \* [...] zeptal $_1$  se $_1$  ho $_3$  [se snahou] $_2$  přimět $_3$  k  
 ask.PTCP.SG.M REFL him.ACC with aim bring.INF to  
 odpovědi.  
 answer  
 ‘[...] he asked, with the aim of getting him to answer.’  
 (Cz; Junghanns 2002: 75)

Below are similar examples from Bosnian (24) and Croatian (25) web corpora and their permutations, which were not accepted by our informants.

- (24) a. [...] i došao $_1$  [u situaciju] $_2$  vratiti $_3$  se $_3$  u meč.  
 and come.PTCP.SG.M in situation return.INF REFL in match

- b. \* [...] i došao<sub>1</sub> se<sub>3</sub> [u situaci]u<sub>2</sub> vratiti<sub>3</sub> u meč.  
 and come.PTCP.SG.M REFL in situation return.INF in match  
 ‘[...] and he was in a position to come back into the match.’ [bsWaC v1.2]
- (25) a. [...] i pružila<sub>1</sub> ruku [u namjeri]u<sub>2</sub> pomilovati<sub>3</sub> me<sub>3</sub>  
 and extend.PTCP.SG.F hand in intention caress.INF me.ACC  
 po obrazu [...].  
 on cheek
- b. \* [...] i pružila<sub>1</sub> me<sub>3</sub> ruku [u namjeri]u<sub>2</sub> pomilovati<sub>3</sub>  
 and extend.PTCP.SG.F me.ACC hand in intention caress.INF  
 po obrazu [...].  
 on cheek  
 ‘[...] and reached out an arm intending to caress my cheek [...].’  
 [hrWaC v2.2]

As the example above suggests, it seems that in BCS, just like in Czech, a CL cannot not climb out of an infinitive phrase which is a complement of a noun in a prepositional phrase. It is important to note that although these constructions share some features with the light verb constructions described in Section 11.2.5.1, only the former seem to function as a constraint in BCS.

### 11.2.5.3 Infinitives as complements of agreeing predicative adjectives

Junghanns (2002: 75) argues that CLs do not climb out of infinitives embedded in a predicative adjective phrase. In his example presented in (26a) the reflexive CL *se* stays in the embedding of its governor, the infinitive *vyjádřit* ‘express’ which is a complement of the agreeing predicative adjective *schopni* ‘able’. He emphasises that such cases should be strictly distinguished from CL positioning with predicative adjectives like in example (13) given above.

- (26) a. [...] jsme<sub>1</sub> schopni<sub>1</sub> se<sub>2</sub> i k této věci společně vyjádřit<sub>2</sub>?  
 be.1PL able REFL and to this matter together express.INF
- b. \* Jsm<sub>1</sub> se<sub>2</sub> schopni<sub>1</sub> i k této věci společně vyjádřit<sub>2</sub>?  
 be.1PL REFL able and to this matter together express.INF  
 ‘Can we express ourselves together regarding this matter?’  
 (Cz; Junghanns 2002: 75)

However, Junghanns (2002: 76) admits that there are counterexamples to the constraint in question. In the following example (27), the pronominal dative CL *mu* ‘him’ climbs out of the embedded infinitive *řít* ‘say’ in spite of the fact that the

latter is a complement of the agreeing predicative adjective *schopen* ‘able’. We would like to point out that in both examples, (26b) and (27), the infinitives are complements of the same agreeing predicative adjective, i.e. *schopen*.

- (27) Já *jsem*<sub>1</sub> *mu*<sub>2</sub> ted’ však *nebyla*<sub>1</sub> *schopna*<sub>1</sub> nic říct<sub>2</sub>.  
 I be.1SG him.DAT now but NEG.be.PTCP.SG.F able nothing say.INF  
 ‘But I was unable to tell him anything.’ (Cz; Junghanns 2002: 76)

Junghanns assumes that in this and similar examples, the adjective moves to the verb, where it becomes incorporated. The CL can then be extracted over the V+A head (cf. Junghanns 2002: 76). He upholds the claim that in some cases incorporation is not possible, which he supports with the unacceptable example in (26b). However, he admits that the exact conditions of CC in such structures are yet to be clarified. We would like to point out that CL type might be responsible for the difference in the acceptability of examples (26b) and (27). Namely, reflexives might be blocked from climbing out of an infinitive phrase which is a complement of an agreeing predicative adjective (26b), in contrast to pronominal CLs which might not be blocked (27). This would be plausible since Dotlačil (2004: 82) shows that in the case of CC out of an infinitive as a complement of a non-agreeing predicative a similar constraint applies only to reflexive CLs (see next section).

Let us have a look at BCS. Examples (28)–(30) extracted from {bs,hr,sr}WaC suggest that in BCS pronominal CLs can be extracted out of an infinitive which complements an agreeing predicative.

- (28) [...] *i dužan*<sub>1</sub> *ih*<sub>2</sub> *je*<sub>1</sub> *naručiti*<sub>2</sub> *prilikom prijave*  
 and obligated them.ACC be.3SG order.INF when application  
*putovanja.*  
 travel  
 ‘[...] and he is obligated to order them when applying to travel.’  
 [bsWaC v1.2]
- (29) *Spremni*<sub>1</sub> *smo*<sub>1</sub> *ti*<sub>2</sub> *pomoći*<sub>2</sub> *u svakoj situaciji [...].*  
 ready be.1PL you.DAT help.INF in every situation  
 ‘We are ready to help you in every situation [...].’ [hrWaC v2.2]
- (30) *Pored slijeđenja dužni*<sub>1</sub> *smo*<sub>1</sub> *mu*<sub>2</sub> *pružiti*<sub>2</sub> *i svoju*  
 besides following obligated be.1PL him.DAT offer.INF and own  
*ljubav [...].*  
 love  
 ‘Besides allegiance, we are obligated to offer him our love, too [...].’  
 [srWaC v1.2]

Furthermore, as our corpus-based examples (31)–(33) show, it seems that in contrast to Czech, in BCS the abovementioned restriction does not apply to reflexives.

- (31) [...] *i dužni<sub>1</sub> su<sub>1</sub> ga<sub>2</sub> se<sub>2</sub> pridržavati<sub>2</sub>.*  
 and obligated be.3PL him.GEN REFL abide.by.INF  
 ‘[...] and they are obligated to abide by it.’ [bsWaC v1.2]
- (32) *Spremna<sub>1</sub> sam<sub>1</sub> mu<sub>2,3</sub> se<sub>2</sub> vratiti<sub>2</sub> i oprostiti<sub>3</sub> [...].*  
 ready be.1SG him.DAT REFL return.INF and forgive.INF  
 ‘I am ready to return to him and forgive him [...].’ [hrWaC v2.2]
- (33) *Jesam<sub>1</sub> li se<sub>2</sub> spreman<sub>1</sub> preseliti<sub>2</sub>?*  
 be.1SG Q REFL ready move.INF  
 ‘Am I ready to move?’ [srWaC v1.2]

This constraint seems to be another difference between CC in BCS and Czech.

#### 11.2.5.4 Infinitives as complements of non-agreeing predicatives

Junghanns (2002: 77) notes that in Czech it is not possible to extract CLs from postponed infinitives complementing non-agreeing predicatives.<sup>12</sup> In example (34a) and its permutation (34b) which he provides, the reflexive CL *se* and the pronominal CL *mu* ‘mu’, governed by the embedded infinitive *ukazovat* ‘show’, cannot climb because the latter is a complement of the non-agreeing predicative *vhodné* ‘appropriate’.

- (34) a. [...] *mám pořád pocit že není<sub>1</sub> vhodné<sub>1</sub> ukazovat<sub>2</sub>*  
 have.1PRS always feeling that NEG.be.3SG appropriate show.INF  
*se<sub>2</sub> mu<sub>2</sub> št’astní.*  
 REFL him.DAT happy
- b. \* [...] *že se<sub>2</sub> mu<sub>2</sub> není<sub>1</sub> vhodné<sub>1</sub> ukazovat<sub>2</sub> št’astní.*  
 that REFL him.DAT NEG.be.3SG appropriate show.INF happy  
 ‘I still feel it is inappropriate to look happy in front of him.’  
 (Cz; Junghanns 2002: 77)

<sup>12</sup>This kind of infinitive complement is labelled as “rechtsextraponierter Subjektsatz” in Junghanns (2002: 77) or as “an infinitival clause being a subject” in Dotlačil (2004: 82). We shall refrain from discussing the question whether a complement clause can occupy the position of the subject.

Dotlačil (2004: 82) later examined this constraint in more detail and refined Junghanns' statement. He claims that it is only reflexive CLs *se* and *si* that are blocked from climbing out of this type of infinitive complement. In contrast, this restriction does not apply to other CLs (cf. Dotlačil 2004: 82). He supports his claims with examples featuring dative (35) and accusative CLs (36) which have climbed out of infinitive complements embedded in non-agreeing predicatives.

- (35) Myslím, že *mu*<sub>2</sub> není<sub>1</sub> možné<sub>1</sub> pomoci<sub>2</sub>.  
 think.1PRS that him.DAT NEG.be.3SG possible help.INF  
 'I think that it is not possible to help him.' (Cz; Dotlačil 2004: 82)
- (36) Myslím, že *tě*<sub>2</sub> / *ho*<sub>2</sub> není<sub>1</sub> možné<sub>1</sub> touhle zbraní  
 think.1PRS that you.ACC him.ACC NEG.be.3SG possible this weapon  
 zabít<sub>2</sub>.  
 kill.INF  
 'I think that it is not possible to kill you/him with this weapon.'  
 (Cz; Dotlačil 2004: 82)

In Junghanns' (2002: 77) example (34a), there are two CLs in the embedded infinitive, the reflexive CL *se* and the pronominal CL *mu*. Since the permutation with CC results in an unacceptable sentence (34b), Junghanns assumes that no CL can climb out of an infinitive that is a complement of a non-agreeing predicative. In contrast, both of Dotlačil (2004: 82) examples have a single CL governed by the embedded infinitive. In this way, he was able to narrow down this specific CC constraint to reflexive CLs only. The reason why Junghanns' permuted sentence presented in (34b) is unacceptable is possibly the fact that in Czech CC seems to be an all-or-nothing phenomenon (see Section 11.5.2 for more details). Hence, in example (34a) the reflexive CL *se* does not climb since it falls under the mentioned restriction and as a consequence the pronominal CL *mu* cannot climb either.

Browsing the literature on BCS, we came across the example in (37) from Ridjanović (2012: 564) which goes against claims of Junghanns (2002: 77) and Dotlačil (2004: 82). Namely, in this example the REFL<sub>LEX</sub> CL *se* does climb out of the infinitive complement embedded in the non-agreeing adjective *dobro* 'good'.

- (37) Dobro<sub>1</sub> *se*<sub>2</sub> je<sub>1</sub> nadati<sub>2</sub>.  
 good REFL be.3SG hope.INF  
 'It is good to hope.' (Bs; Ridjanović 2012: 564)

Similarly, as example (38) shows, we found examples with climbing of the reflexive CL in the same structure in the Serbian web corpus. Moreover, we found dozens of such examples in hrWaC v2.2. In (39) we provide one of them.

- (38) Dobro<sub>1</sub> se<sub>2</sub> je<sub>1</sub> raspitati<sub>2</sub> kod drugih, ali [...].  
 good REFL be.3SG ask.around.INF at others but  
 ‘It is good to ask around, but [...].’ [srWaC v1.2]
- (39) [...] no dobro<sub>1</sub> se<sub>2</sub> je<sub>1</sub> uvjeriti<sub>2</sub> da [...].  
 but good REFL be.3SG convince.INF that  
 ‘[...] but it is good to convince yourself that [...].’ [hrWaC v2.2]
- Moreover, we would like to emphasise that in all three web corpora we found examples with climbing of pronominal CLs out of infinitive complements of non-agreeing predicatives – see (40)–(42) below.
- (40) Također, dobro<sub>1</sub> ga<sub>2</sub> je<sub>1</sub> konzumirati<sub>2</sub> za [...].  
 also good him.ACC be.3SG consume.INF for  
 ‘Also, it is good to consume it for [...].’ [bsWaC v1.2]
- (41) [...] potrebno<sub>1</sub> ih<sub>3</sub> je<sub>1</sub> pokušati<sub>2</sub> osnovati<sub>3</sub> [...].  
 necessary them.ACC be.3SG try.INF establish.INF  
 ‘[...] it is necessary to try to establish them [...].’ [hrWaC v2.2]
- (42) U svakom slučaju, dobro<sub>1</sub> ih<sub>2</sub> je<sub>1</sub> imati<sub>2</sub>.  
 in every case good them.ACC be.3SG have.INF  
 ‘In any case, it is good to have them.’ [srWaC v1.2]

As our corpus data show, it seems that the constraint on CC out of infinitive embeddings of non-agreeing predicatives does not apply to BCS at all. Namely, in these varieties it is possible to extract not only pronominal CLs from infinitives complementing non-agreeing predicatives like in Czech, but also the reflexive CL *se*.

### 11.2.6 Embedded wh-infinitives

Junghanns (2002: 77), Dotlačil (2004: 83), and Rezac (2005: 8, 9) argue that although wh-infinitives generally do not present islands for syntactic movements in Czech – for instance, full prepositional phrases can be extracted from them – they do not allow the extraction of CLs. In other words, CC out of interrogative wh-infinitives is not possible.<sup>13</sup> Junghanns (2002: 77) supports his claims with example (43a) and its unacceptable permutation (43b). From the latter it is clear that the pronominal accusative CL *ho* ‘him’ cannot climb out of wh-infinitive *jak zapisovat* ‘how to record’.

<sup>13</sup>A marginal exception poses Modal Existential Construction described by Šimík (2011), where CC is possible both in Czech and BCS.

11 Constraints on clitic climbing in Czech compared to BCS

- (43) a. Ale neví<sub>1</sub> opravdu, jak ho<sub>2</sub> zapisovat<sub>2</sub>.  
 but NEG.see.1PRS really how him.ACC write.down.INF
- b. \*Ale neví<sub>1</sub> ho<sub>2</sub> opravdu, jak zapisovat<sub>2</sub>.  
 but NEG.see.1PRS him.ACC really how write.down.INF
- ‘I do not really know how to record him.’ (Cz; Junghanns 2002: 77)

Aljović (2004: 3) claims that the same constraint exists in BCS; she provides the following example with a *da*-complement and its permutation:

- (44) a. Ona nije<sub>1</sub> odlučila<sub>1</sub> kako (da li) da mu<sub>2</sub>  
 she NEG.be.3SG decide.PTCP.SG.F how that Q that him.DAT  
 pomogne<sub>2</sub> (ili ne).  
 help.3PRS or not
- b. \*Ona mu<sub>2</sub> nije<sub>1</sub> odlučila<sub>1</sub> kako (da li) da  
 she him.DAT NEG.be.3SG decide.PTCP.SG.F how that Q that  
 pomogne<sub>2</sub> ili ne.  
 help.3PRS or not
- ‘She did not decide how (/whether) to help him (or not)’.  
 (BCS; Aljović 2004: 3)

In her subsequent paper Aljović (2005: 8) provides evidence that CC out of wh-infinitives is blocked in BCS, in the form of permuted example (45b).

- (45) a. Mila je<sub>1</sub> odlučila<sub>1</sub> kome ga<sub>2</sub> preporučiti<sub>2</sub>.  
 Mila be.3SG decide.PTCP.SG.F who him.ACC recommend.INF
- b. \*Mila ga<sub>2</sub> je<sub>1</sub> odlučila<sub>1</sub> kome preporučiti<sub>2</sub>.  
 Mila him.ACC be.3SG decide.PTCP.SG.F who recommend.INF
- ‘Mila decided to whom to recommend it.’ (BCS; Aljović 2005: 8)

Our corpus-based examples (46a–48a) and their rejected permutations (46b–48b) confirm that this constraint indeed applies to wh-infinitives in Bosnian, Croatian, and Serbian as claimed by Aljović (2004, 2005).

- (46) a. “NAFAKA” – bezbroj razloga zašto ga<sub>1</sub> pogledati<sub>1</sub>.  
 NAFAKA countless reasons why him.ACC look.INF
- b. \* “NAFAKA” – bezbroj razloga ga<sub>1</sub> zašto pogledati<sub>1</sub>.  
 NAFAKA countless reasons him.ACC why look.INF
- “NAFAKA” – countless reasons to watch it.’ [bsWaC v1.2]



- (47) a. *Imate<sub>1</sub> li ideju kako mu<sub>2</sub> pomoći<sub>2</sub>?*  
 have.2PL Q idea how him.DAT help.INF  
 b. \**Imate<sub>1</sub> li mu<sub>2</sub> ideju kako pomoći<sub>2</sub>?*  
 have.2PL Q him.DAT idea how help.INF  
 ‘Do you have any idea how to help him?’ [hrWaC v2.2]
- (48) a. [...] *i potreбно<sub>1</sub> je<sub>1</sub> znati<sub>2</sub> kako ga<sub>3</sub> proceniti<sub>3</sub>.*  
 and necessary be.3SG know.INF how him.ACC estimate.INF  
 b. \* [...] *i potreбно<sub>1</sub> ga<sub>3</sub> je<sub>1</sub> znati<sub>2</sub> kako proceniti<sub>3</sub>.*  
 and necessary him.ACC be.3SG know.INF how estimate.INF  
 ‘[...] and it is necessary to know how to assess him.’ [srWaC v1.2]

This constraint is one of the clear cases of a lack of CC in both BCS and Czech.

## 11.3 Constraints related to the raising–control distinction

### 11.3.1 Object-controlled complements

There is an intensive and quite controversial debate on a possible relationship between CC and certain types of control phenomena. Most of the authors working on CC in Czech have pointed out that, unlike in raising and subject control complements, in object-controlled complements CC is highly restricted.<sup>14,15</sup> Claims have been made that CC does not completely depend on the raising–control distinction, but rather on its combination with other features like case, person, animacy, and CL type (pronominal vs reflexive), which will be discussed separately in subsequent sections.

Thorpe (1991) and Junghanns (2002) argue that in Czech CLs generally cannot climb from object-controlled infinitives, whereas Rezac (1999, 2005), Dotlačil (2004), Lenertová (2004), Hana (2007), and George & Toman (1976) do not entirely exclude this possibility. The disagreement between scholars becomes even more apparent when they quote examples which their colleagues evaluated as acceptable and mark them either as completely unacceptable (normally with \*) or as somewhat questionable (usually with ?). For instance, when arguing that CLs do not climb out of object-controlled infinitives, Junghanns (2002: 69) quotes

<sup>14</sup>For more information on the distinction between raising and control predicates see Section 2.5.2.

<sup>15</sup>For some interesting examples on restrictions on CC out of infinitive complements of reflexive subject control verbs in Czech, see Lenertová (2004: 159f).

## 11 Constraints on clitic climbing in Czech compared to BCS

Rezac's example (49), and marks it with an asterisk, although in Rezac's text the very same example was evaluated as acceptable.<sup>16</sup>

- (49) Marie *ho*<sub>2</sub> Petrovi<sub>1</sub> přikázala<sub>1</sub> poslat<sub>2</sub> domů.  
 Marie him.ACC Peter.DAT order.PTCP.SG.F sent.INF home  
 'Marie ordered Peter to send him home.' (Cz; Rezac 1999)

Hana (2007: 129), like Rezac (1999, 2005), thinks that the object control constraint does not apply to CC in Czech. He provides the following example (50) in which the pronominal CL *ho* 'him' climbs out of the infinitive complement *vyhodit* 'fire' of the object control matrix predicate *doporučila* 'recommended'.

- (50) [...] a když *ho*<sub>2</sub> i perzonalistika doporučila<sub>1</sub>  
 and when him.ACC and human.resources recommend.PTCP.SG.F  
 šéfovi<sub>1</sub> vyhodit<sub>2</sub> [...].  
 boss.DAT fire.INF  
 '[...] and when even personnel management recommended his boss to  
 fire him [...].' (Cz; Hana 2007: 129)

Although Aljović (2005: 4) does not use the term subject and object control, she indirectly comments on it when she states that in BCS CC is only possible out of complement clauses whose subject is empty and coreferential with the matrix subject. However, in a footnote she admits that CC is also possible when the subject of the embedded clause is coreferential with the matrix indirect object in the dative (cf. Aljović 2005: 4), i.e. out of object control CTPs. She provides example (51) in which the pronominal CL *ih* 'them' climbed out of the infinitive *posjetiti* 'visit' and clusterised with the dative CL *nam* 'us', which is a complement of the object control CTP *brani* '(she) forbids'.

- (51) Ona *nam*<sub>1</sub> *ih*<sub>2</sub> brani<sub>1</sub> posjetiti<sub>2</sub>.  
 she us.DAT them.ACC forbid.3PRS visit.INF  
 'She forbids us to visit them.' (BCS; Aljović 2005: 4)

Aljović's example shows that in BCS a dative object in the matrix clause does not necessarily have to block CC out of infinitive complements (cf. for Czech George & Toman 1976, Dotlačil 2004, Rezac 2005, and Hana 2007). Our corpus data also indicate that pronominal CLs can climb out of object-controlled infinitives: see the example in (52).

<sup>16</sup>Due to lack of space, in examples in all other chapters we glossed case only for personal pronouns. In this chapter and this section on the raising-control distinction we gloss the case of nominal objects in order to help readers follow the presented discussion.

- (52) [...] koji  $mi_1$   $ga_2$  pomažu<sub>1</sub> nositi<sub>2</sub>.  
 which me.DAT him.ACC help.3PRS carry.INF  
 ‘[...] which help me to carry it.’ [hrWaC v2.2]

### 11.3.2 Object control constraint related to case

Rezac (2005) and Dotlačil (2004: 79) elaborate further on the control constraint. They note that object control verbs restrict the freedom of CC through constraints which are based on case. Rezac (2005: 17) argues that there is a coherent pattern where restructuring is blocked by object control verbs. More specifically, whether a CL climbs or not depends on the one hand on the case of the controller, and on the other hand on the case of the CL governed by the embedded infinitive.<sup>17</sup> Furthermore, Rezac (2005: 7) claims that this constraint does not depend on whether the pronoun is in the full or CL form in a given sentence. Accordingly, object control CTPs with a dative controller only allow accusative CLs to climb from the infinitive (53), and block climbing by dative CLs (54).<sup>18</sup>

Example (55) shows that CC does not occur even if the controller is expressed as a NP in the dative – in this case *Martinovi* ‘(to) Martin’ (cf. Rezac 2005: 17f).<sup>19</sup>

<sup>17</sup>Rezac (2005: 17) argues that in contrast to object control CTPs, raising and subject control CTPs exhibit no case restrictions on CC out of their infinitive complements. In other words, both dative and accusative CLs can climb freely. However, the reader should bear in mind that there could be exceptions to this rule in Czech, for more information see Lenertová’s (2004: 159f) examples with CC out of infinitive complements of the reflexive subject control verb *podřít se* ‘manage’.

<sup>18</sup>Alexandr Rosen (p.c.) argues that CC and the resulting mixed cluster of two dative CLs in the following permuted example is acceptable to him:

- (i) Dana  $mi_1$   $mu_2$  přikázala<sub>1</sub> pomoct<sub>2</sub> s mytím.  
 Dana me.DAT him.DAT order.PTCP.SG.F help.INF with washing  
 ‘Dana ordered me to help him with the washing.’

<sup>19</sup>Alexandr Rosen (p.c.) points out that the string itself is acceptable, as long as *mu* ‘him’ is a matrix complement, like in the permuted example below. Note that in example (i) it is not the CL *mu* ‘him’ that climbs, but the dative NP *Martinovi* ‘(to) Martin’.

- (i) Dana  $mu_1$  Martinovi<sub>2</sub> přikázala<sub>1</sub> pomoct<sub>2</sub>.  
 Dana him.DAT Martin.DAT order.PTCP.SG.F help.INF  
 ‘Dana ordered him to help Martin.’

11 Constraints on clitic climbing in Czech compared to BCS

- (53) Dana  $ti_1$   $ho_2$  doporučila<sub>1</sub> poslat<sub>2</sub> [...].  
 Dana you.DAT him.ACC recommend.PTCP.SG.F sent.INF  
 ‘Dana recommended that you send him [...].’ (Cz; Rezac 2005: 17)
- (54) \* Dana  $jí_1$   $mu_2$  přikázala<sub>1</sub> pomoct<sub>2</sub> s mytím.  
 Dana her.DAT him.DAT order.PTCP.SG.F help.INF with washing  
 Intended: ‘Dana ordered her to help him with the washing.’  
 (Cz; Rezac 2005: 17)
- (55) \* Dana  $mu_2$  Martinovi<sub>1</sub> přikázala<sub>1</sub> pomoct<sub>2</sub>.  
 Dana him.DAT Martin.DAT order.PTCP.SG.F help.INF  
 Intended: ‘Dana ordered Martin to help him.’  
 (Cz; Rezac 2005: 18)

In addition, Rezac (2005: 18) argues that if there is an accusative controller (CL or NP), CC is even more restricted since it not only blocks the movement of dative CLs (56), but also prevents accusative CLs from climbing (57).

- (56) \* Dana  $jí_2$   $ho_1$  / Pavla<sub>1</sub> přinutila<sub>1</sub> pomoct<sub>2</sub>.  
 Dana her.DAT him.ACC Pavel.ACC force.PTCP.SG.F help.INF  
 Intended: ‘Dana forced him/Pavel to help her.’ (Cz; Rezac 2005: 18)
- (57) \* Dana  $ho_2$   $tě_1$  / Marii<sub>1</sub> přinutila<sub>1</sub> políbit<sub>2</sub>.  
 Dana him.ACC you.ACC Maria.ACC force.PTCP.SG.F kiss.INF  
 Intended: ‘Dana forced you/Maria to kiss him.’ (Cz; Rezac 2005: 18)

Here we must warn the reader that the unacceptability of the example in (56) may be related to the ordering restrictions of the CL *ho* ‘him’. According to Lenertová (2004: 153f), the Czech CL *ho* cannot appear initially in a cluster, and CC is not felicitous with pairs that have the preferred inverted order. Moreover, Lenertová (2004: 154) provides example (58), in which an accusative CL climbs to the matrix in spite of the accusative controller:<sup>20</sup>

- (58) Stejně  $jí_1$   $ho_2$  nenechali<sub>1</sub> dokončit<sub>2</sub>.  
 anyway her.ACC it.ACC NEG.let.PTCP.PL.M finish.INF  
 ‘Anyway, they did not let her finish it.’ (Cz; Lenertová 2004: 153f)

Let us now turn to BCS, which seems to show some variation between the national variants with respect to object control constructions. Regarding CC out of object-controlled infinitives in Croatian, we found examples (59)–(60) in which

<sup>20</sup>For more examples of CC in the context of dative and accusative controllers, we refer the reader to Lenertová (2004: 162).

### 11.3 Constraints related to the raising–control distinction

the dative CL complements *nam* ‘us’ and *mi* ‘me’ of the matrix verb, i.e. controller, did not prevent the accusative CLs *ga* ‘him’ and *ju* ‘her’ generated in an infinitive complement from climbing.

- (59) [...] *koji nam<sub>1</sub> ga<sub>2</sub> pomaže<sub>1</sub> upoznati<sub>2</sub>* [...].  
 which us.DAT him.ACC help.1PRS meet.INF  
 ‘[...] which helps us get to know him [...].’ [hrWaC v2.2]
- (60) [...] *koji su<sub>1</sub> mi<sub>1</sub> ju<sub>2</sub> pomogli<sub>1</sub> svladati<sub>2</sub>*.  
 which be.3PL me.DAT her.ACC help.PTCP.PL.M overcome.INF  
 ‘[...] which helped me to overcome her.’ [hrWaC v2.2]

These examples indicate that in Croatian, just like in Czech, object control CTPs with a dative CL complement do not necessarily prevent accusative CLs from climbing out of infinitive embeddings.<sup>21</sup> It is interesting to note, however, that we could not find such examples in either bsWaC or in srWaC. One reason for this could be that in Bosnian and Serbian *da<sub>2</sub>*-complements predominate with object control CTPs, and not the infinitive.<sup>22</sup> Unlike accusative CLs, dative CLs do not seem to climb out of infinitive complements in the presence of a dative controller. Permutation seems to lead to unacceptable sentences: compare example (61a) and its unacceptable permutation (61b).<sup>23</sup>

- (61) a. [...] *da nam<sub>1</sub> pomognete<sub>1</sub> naći<sub>2</sub> im<sub>2</sub> najbolje skrbnike*.  
 that us.DAT help.2PRS find.INF them.DAT best guardians
- b. \* [...] *da nam<sub>1</sub> im<sub>2</sub> pomognete<sub>1</sub> naći<sub>2</sub> najbolje skrbnike*.  
 that us.DAT them.DAT help.2PRS find.INF best guardians  
 ‘[...] that you help us find the best guardians for them.’ [hrWaC v2.2]

<sup>21</sup>Although these two sentences are not the only two sentences found in hrWaC with an accusative CL that climbs out of an infinitive in spite of a dative CL controller in the matrix clause, in psycholinguistic experiment (see Chapter 15) we could not corroborate this possibility as a general tendency. This might be due to the lack of ecological validity of our stimuli (see Sections 3.2.1 and 3.2.2) or more probably to the fact that in such a context CC is lexically restricted only to certain matrix predicates such as *pomoći* or *pomagati* ‘help’.

<sup>22</sup>For the highly restricted possibility of CC out of *da<sub>2</sub>*-complements see Jurkiewicz-Rohrbacher, Kolaković & Hansen (2017) and 13.4.

<sup>23</sup>Hansen, Kolaković & Jurkiewicz-Rohrbacher (2018: 263, 265) speak, in the context of stacked infinitives, about a same case-different governors constraint, for which they provide empirical evidence. In their example the dative reflexive CL *si* controller blocks climbing of a dative CL generated in the infinitive. Moreover, this tendency was also corroborated in our psycholinguistic experiment, see Section 15.6.4. In other words, the results of both the psycholinguistic experiment and our inspection of corpora are in line: dative CLs do not to climb out of infinitive embeddings in the presence of a dative CL controller in the matrix clause.

## 11 Constraints on clitic climbing in Czech compared to BCS

Furthermore, it seems that in Croatian, like in Czech, object control CTPs with accusative CL complements block climbing of accusative CLs generated in infinitives, as example (62a) and its unacceptable permutation (62b) suggests.

- (62) a.  $\check{Z}$ elim<sub>1</sub>    *te*<sub>2</sub>        naučiti<sub>2</sub> voljeti<sub>3</sub> *me*<sub>3</sub>.  
          want.1PRS you.ACC teach.INF love.INF me.ACC
- b. \* $\check{Z}$ elim<sub>1</sub>    *te*<sub>2</sub>        *me*<sub>3</sub>    naučiti<sub>2</sub> voljeti<sub>3</sub>.  
          want.1PRS you.ACC me.ACC teach.INF love.INF
- ‘I want to teach you to love me.’ [hrWaC v2.2]

However, some caution is called for. It may be the case that *me* cannot climb not only due to object control constraint related to case, but also due to phonological similarity (both *me* ‘me’ and *te* ‘you’ end in the vowel [ɛ]) or due to person constraint (see the next Section 11.3.3 for a detailed discussion).

We would like to emphasise that in bsWaC and srWaC we could not easily find infinitive complements of object control CTPs. Therefore, we conducted a special corpus study, presented in Chapter 13.

At the end of this subsection we would like to point out once more that scholars differ in their opinion on CC out of object-controlled infinitives in Czech. Rezac (2005: 18) claims that not only CL, but also NP accusative complements in the matrix clause block all CC. In contrast, Dotlačil (2004: 81) leaves open the possibility that two CLs in the same case will appear in one mixed cluster as a consequence of CC (for more information, see the next Section 11.3.3). Hana (2007: 123) makes it clear that two morphologically and phonetically identical CLs governed by different governors cannot appear together in the same cluster. However, he does not comment on what will happen if a sentence contains CLs in the same case with different governors which are phonetically different, i.e., if there is a difference in the grammatical category of person. Nevertheless, discussing various examples involving object controlled VPs, Hana (2007: 130) concludes that “it seems clear that for non-reflexive clitic a more fine grained distinction of verbs than that based on control is needed.”

### 11.3.3 Object control person-case constraint

It is necessary to add one more observation concerning the object control constraint related to case, made only by Dotlačil (2004) for Czech. Although he does not explicitly state that he is drawing on Bonet’s (1991, 1983) person-case con-

### 11.3 Constraints related to the raising–control distinction

straint (PCC), they clearly have some points in common.<sup>24</sup> He argues that in Czech, if the matrix clause has an object, the only CL which can climb is the third person accusative (cf. Dotlačil 2004: 79ff). He illustrates his claims with the acceptable example in (63a) and its unacceptable permutation in (63b). In the former, although the matrix clause has an indirect object in the dative *Jirkovi* ‘Jirka’, the third person CLs in the accusative *ho, ji, je* ‘him, her, them’ can freely climb from the infinitive embedding *navštěvovat* ‘visit’. In contrast, in the latter example, climbing of the first and second person accusative CLs *mě, tě, nás, vás* ‘me, you, us, you’ leads to unacceptable sentences.

(63) a. Doktoři *ho*<sub>2</sub> / *ji*<sub>2</sub> / *je*<sub>2</sub> / *Jirkovi*<sub>1</sub> *zakázali*<sub>1</sub>  
 doctors him.ACC her.ACC them.ACC Jirka.DAT forbid.PTCP.PL.M  
*navštěvovat*<sub>2</sub>.  
 visit.INF

‘The doctors forbade Jirka to visit him/her/them.’

b. \*Doktoři *mě*<sub>2</sub> / *tě*<sub>2</sub> / *nás*<sub>2</sub> / *vás*<sub>2</sub> *Jirkovi*<sub>1</sub>  
 doctors me.ACC you.ACC us.ACC you.ACC Jirka.DAT  
*zakázali*<sub>1</sub> *navštěvovat*<sub>2</sub>.  
 forbid.PTCP.PL.M visit.INF

Intended: ‘The doctors forbade Jirka to visit me/you/us/you.’

(Cz; Dotlačil 2004: 80f)

Dotlačil (2004: 81) summarises his observations as follows: the first important factor in blocking CC is arguments. However, not all arguments block CC: only objects do (Dotlačil 2004: 81). The most powerful factor in preventing CLs from climbing is an accusative object, which blocks climbing of all CLs other than the accusative third person (Dotlačil 2004: 81).<sup>25</sup>

Here we would like to point out that it is rather difficult to find examples or counterexamples for the object control person-case constraint in BCS. Example

<sup>24</sup>According to PCC rule in a combination of a direct object and an indirect object, the direct object has to be in the third person. Both the direct object and the indirect object are phonologically weak.

<sup>25</sup>As we pointed out above, Rezac (2005: 18) does not agree with this, claiming that “Accusative controllers, clitic or NP, also block the climbing of dative clitics, but in addition block the climbing of accusative clitics as well.” Furthermore, he claims that PCC does not always hold in Czech and that this kind of restriction depends on the dative type (cf. Rezac 2005: 25). PCC holds for the combination of dative and accusative CLs only in the case of argumental, benefactive or possessive dative, while in the case of dative of address the combination of dative and non-third person accusative CL is not excluded (cf. Rezac 2005: 25). However, it is important to emphasise that his example does not have an embedded infinitive complement.

(59) presented in Section 11.3.2 suggests that a third person accusative CL can climb into a matrix clause which contains a dative CL. In order to test whether non-third person CLs can climb, we permuted this example and asked native speakers to perform informal acceptability judgments of (64a) and (64b).

- (64) a. [...] koji *nam*<sub>1</sub> *te*<sub>2</sub> pomaže<sub>1</sub> upoznat<sub>2</sub>.  
           which us.DAT you.ACC help.1PRS meet.INF  
           ‘[...] which helps us get to know you.’  
       b. [...] koji *joj*<sub>1</sub> *me*<sub>2</sub> pomaže<sub>1</sub> upoznat<sub>2</sub>.  
           which her.DAT me.ACC help.1PRS meet.INF  
           ‘[...] which helps her get to know me.’

Native speakers of Croatian accepted the permuted sentences in the informal acceptability test. This result indicates that the person constraint does not hold for CC out of object-controlled infinitive complements with a dative controller. However, further testing and empirical, robust data are still necessary.

### 11.3.4 Object control and animacy of the referent of the clitic

George & Toman (1976) show that in Czech, a CL can climb from an infinitive headed by a causative. They claim that if the matrix contains an object, i.e. if the CTP is object control, only inanimate objects can climb from the infinitive complement (cf. George & Toman 1976: 241).<sup>26</sup>

They support their claims with (65a) and its unacceptable permutation (65b). In the former, the matrix clause has the direct object *Karla* ‘Karel’, but the accusative CL *ji* ‘her/it’ can climb out of the object-controlled infinitive *napsat* ‘write’ since it has an inanimate referent – application. In contrast, in (65b) climbing of the very same CL *ji* ‘her/it’ to the very same matrix leads to an unacceptable sentence, since it has an animate referent – that woman.<sup>27</sup>

<sup>26</sup>The example (65a) they provide contradicts the constraint proposed by Rezac (2005), presented in Section 11.3.2. Note that according to Rezac (2005), any kind of complement in the accusative blocks climbing of an accusative CL complement out of an infinitive, while George & Toman (1976) believe that only climbing of CLs with animate referents is blocked.

<sup>27</sup>According to George & Toman (1976: 245) CC in the context of object control matrix predicates depends wholly on the animacy of the argument of the infinitive complement. In contrast to Dotlačil (2004) and Rezac (2005), they argue that even if the controller is a NP in the dative, infinitive accusative CLs with an animate referent cannot climb into the matrix clause (cf. George & Toman 1976: 245). However, we would like to point out that this whole discussion on animacy is a bit vague. Namely, the problem is rather that they do not test the mentioned constraint with changing referents. One may wonder: would different referents change the acceptability?





### 11.3.5 Object control reflexive constraint

Regarding the two different control types, Hana (2007: 129f) observes that Czech reflexive CLs can climb from subject-controlled (69), but not from object-controlled infinitives (70b).<sup>28,29,30</sup>

- (69) Martin *se*<sub>2</sub> potřebuje<sub>1</sub> zeptat<sub>2</sub> jak [...].  
 Martin REFL need.3PRS ask.INF how  
 ‘Martin needs to ask how [...].’ (Cz; Hana 2007: 130)
- (70) a. Martin zakázal<sub>1</sub> Petrovi dívat<sub>2</sub> *se*<sub>2</sub> na televizi.  
 Martin forbid.PTCP.SG.M Peter.DAT watch.INF REFL on TV
- b. \*Martin *se*<sub>2</sub> zakázal<sub>1</sub> Petrovi dívat<sub>2</sub> na televizi.  
 Martin REFL forbid.PTCP.SG.M Peter.DAT watch.INF on TV  
 ‘Martin forbade Peter to watch TV.’ (Cz; Hana 2007: 129)

In BCS, like in Czech, a reflexive can climb from raising (71) and subject-controlled infinitives (72), but not from object-controlled infinitives (73b) (cf. Hansen, Kolaković & Jurkiewicz-Rohrbacher 2018). This finding from the corpus study on stacked infinitives was also corroborated in our psycholinguistic experiment (see Chapter 15).

- (71) Mogu<sub>1</sub> *se*<sub>2</sub> samo nasmijati<sub>2</sub> tvojem neznanju.  
 can.1PRS REFL only laugh.INF your ignorance  
 ‘I can only laugh at your ignorance.’ [hrWaC v2.2]

<sup>28</sup>Classification of the matrix predicate *potřebovat* ‘need’ as subject control was taken from Hana (2007: 130).

<sup>29</sup>Dotlačil (2004: 81) also noted that reflexives cannot climb if there is an object in the matrix clause.

<sup>30</sup>Whether a reflexive climbs or not probably does not depend only on the raising–control distinction of the CTP, but also on the type of the reflexive itself. This is suggested by Lešnerová & Malink (2008), who examine the position of the Czech reflexive CL *se* in active and passive sentences with the raising phase verb *přestat* ‘to stop’ and an infinitive complement. Their data suggest that CC is obligatory for passive sentences, i.e. CC is obligatory for the reflexive passive marker *se* in Czech. In contrast, in active structures CC can but does not have to occur (cf. Lešnerová & Malink 2008: 400f). According to them, the lack of CC in passive sentences leads to incorrect agentive interpretations of the sentence (cf. Lešnerová & Malink 2008: 396, 400f). In contrast, if the free morpheme *se* of agentive reflexive deponent verbs does not climb into the matrix clause with the phase verb *přestat*, the resulting sentence will not be ungrammatical, but will have a marked information structure (cf. Lešnerová & Malink 2008: 499f). Some observations on CC of accusative complements to a passivised matrix verb in Czech can be found in Lenertová (2004: 159).

- (72) No ne znaju<sub>1</sub> se<sub>2</sub> svi izraziti<sub>2</sub> lijepo [...].  
 but NEG know.3PRS REFL all express.INF nicely  
 ‘But not everyone knows to express themselves nicely [...].’ [hrWaC v2.2]
- (73) a. Pa tko im<sub>1</sub> brani<sub>1</sub> uključiti<sub>2</sub> se<sub>2</sub> u politiku?  
 well who them.DAT forbid.3PRS enter REFL in politics  
 b. \*Pa tko im<sub>1</sub> se<sub>2</sub> brani<sub>1</sub> uključiti<sub>2</sub> u politiku?  
 well who them.DAT REFL forbid.3PRS enter in politics  
 ‘Well, who forbids them to enter the politics?’ [hrWaC v2.2]

## 11.4 Constraints related to mixed clitic clusters

### 11.4.1 Pseudo-twins

Under this section we subsume various observations about constraints on CC in relation to mixed clusters reported by scholars who worked on Czech.<sup>31</sup> In their studies they concentrated on the relationship between two CLs which are generated by different governors and due to different constraints do not end up in a mixed CL cluster. We cover them all under one heading which we named “constraints related to mixed CL clusters”. At this point we would like to emphasise that scholars who reported constraints related to mixed CL clusters did not try to establish a meaningful connection between such clusters and the control phenomena. However, we believe that the constraints related to mixed CL clusters cannot be separated from control constraints.<sup>32</sup> Namely, whenever we analyse two CLs with two different governors, the matrix CTP is of either subject or object control type (see examples in Sections 11.4.2 and 11.4.3). Junghanns (2002: 79) speaks about what he calls a pseudo-twins constraint on CC, whose nature, however, is not completely clear. This category covers two cases in which a CL does not climb out of a complement:

1. there is a phonologically identical CL in the matrix clause;
2. the matrix clause does not contain an identical CL, but a CL with a similar syntactic function.

<sup>31</sup>For the distinction between simple and mixed clusters see Section 2.4.2.1.

<sup>32</sup>We therefore present these constraints directly after the section which was dedicated to the raising and control distinction.

However, Junghanns (2002: 80) warns that the constraints for pseudo-twins do not always apply. In a special context and with enough differentiation, the co-occurrence of similar expressions within one sentence is possible (see Lenertová 2004).

#### 11.4.2 Phonologically identical pronominal and reflexive clitics with different governors

Junghanns (2002: 79) argues that if the matrix clause and the embedded complement contain (phonologically) identical CL, there will be no CC.<sup>33</sup> He supports this claim with example (74a) and its unacceptable permutation (74b). In the latter, climbing of the reflexive CL *se* leads to formation of a mixed cluster with two reflexive CLs and consequently to an unacceptable sentence.

- (74) a. [...] a všude jsem<sub>1</sub> se<sub>1</sub> snažil<sub>1</sub> dozvědět<sub>2</sub> se<sub>2</sub>  
 and everywhere be.1SG REFL try.PTCP.SG.M find.OUT.INF REFL  
 co nejvíc [...].  
 what most
- b. \* Všude jsem<sub>1</sub> se<sub>1</sub> se<sub>2</sub> snažil<sub>1</sub> dozvědět<sub>2</sub> co  
 everywhere be.1SG REFL REFL try.PTCP.SG.M find.OUT.INF what  
 nejvíc.  
 most  
 ‘[...] and I tried to find out as much as possible [...].’  
 (Cz; Junghanns 2002: 79)

Rosen (2014: 104) agrees that two reflexive *se* CLs cannot appear in one mixed cluster, and underlies that this constraint is “blind” to different types of reflexives.<sup>34,35</sup> This is exemplified in (75a): although the reflexive CL *se* in the matrix

<sup>33</sup>Junghanns (2002: 79) does not specify what the adjective “identical” covers: phonological level, morphological level or both. As will become obvious in the following lines, CLs discussed in this section are always phonologically and sometimes also morphologically identical.

<sup>34</sup>Here we would like to emphasise that Rosen allows two reflexives in the dative case generated by two different verbs either to haploglise within the matrix cluster or to appear next to each other, like in the following example:

- (i) Netroufla<sub>1</sub> si<sub>1</sub> | si<sub>2</sub> říct<sub>2</sub> o víc knedlíků.  
 NEG.dare.PTCP.SG.F REFL REFL ask.INF about more dumplings  
 ‘She did not dare to ask for more dumplings.’ (Cz; Rosen 2014: 105)

In his words, this is possible if the two occurrences of *si* are prosodically separated (marked above with |).

<sup>35</sup>For our typology of reflexive CLs see Section 2.5.4.

clause is lexically bound ( $\text{REFL}_{\text{LEX}}$ ) and the reflexive in the complement blocks the internal argument ( $\text{REFL}_{2\text{ND}}$ ), they cannot appear in the same cluster (cf. Rosen 2014: 104f).

- (75) a. \**Děvče*  $se_1$   $se_2$  *stydělo*<sub>1</sub> *převléknout*<sub>2</sub>.  
 girl REFL REFL ashamed.PTCP.SG.N change.INF  
 b. *Děvče*  $se_{1+2}$  *stydělo*<sub>1</sub> *převléknout*<sub>2</sub>.  
 girl REFL ashamed.PTCP.SG.N change.INF  
 ‘The girl was ashamed of changing (clothes).’ (Cz; Rosen 2014: 104)

While Junghanns (2002) offers splitting of reflexives, where each reflexive CL stays with its governor (74a), as the only solution in the case of utterances containing two reflexives with different governors, Rosen (2014) allows haplology (75b). He treats the deletion of one reflexive as an instance of CC since, in his view, it is the matrix reflexive which is deleted (cf. Rosen 2014: 106, 114).

Although the previous examples contained reflexive CLs, it is important to emphasise that the constraint in question does not concern reflexive CLs only. Rather, it is a rule which applies to all pronominal and reflexive CLs, i.e., to all CLs which can theoretically undergo the process of climbing. Basing on Rosen’s example from Czech (76a), Hana (2007: 123) formulated this constraint as the following rule: “A clitic cluster cannot contain two morphologically identical clitics with different governors”. In this example, the dative pronominal CL *mi* ‘me’ does not climb from the infinitive embedding *vrátit* ‘return’ into the matrix clause because the morphologically and phonologically identical CL *mi* governed by the matrix predicate *slíbila* ‘promised’ is already there.

- (76) a. *Kamila*  $mi_1$  *slíbila*<sub>1</sub>  $mi_2$   $to_2$  *vrátit*<sub>2</sub>.  
 Kamila me.DAT promise.PTCP.SG.F me.DAT it.ACC return.INF  
 b. \**Kamila*  $mi_1$   $mi_2$   $to_2$  *slíbila*<sub>1</sub> *vrátit*<sub>2</sub>.  
 Kamila me.DAT me.DAT it.ACC promise.PTCP.SG.F return.INF  
 ‘Kamila promised me to return it to me.’  
 (Cz; Rosen 2001, as cit. in Hana 2007: 123)

Furthermore, we would like to point out that in the three examples provided by Junghanns, Hana and Rosen, CTP predicates are of the subject control type (*snažit se* ‘try hard’, *stydět se* ‘be ashamed’, *slíbit* ‘promise’, *troufnout si* ‘dare’). Although none of these scholars seems to take into account the predicate type as a relevant factor, all three of them recognize the relevance of syntax, since they argue that the constraint is not phonological. All of them support this claim

with strong arguments. In Junghanns' (2002: 80) opinion the constraint cannot be phonological in nature since the reflexive CL *se* and the homonymous preposition *se* can stand next to each other. In addition, to refute a purely phonological nature of the constraint, both Hana (2007: 124) and Rosen (2014: 105) provide examples of the verbal CL *si* 'are' and the reflexive CL *si* in contact position.

Querying {bs,hr,sr}WaC, we did not find a single occurrence of a mixed cluster containing phonologically and morphologically identical pronominal CLs with different governors. Neither did we find them in contexts of pseudodiaclisis.<sup>36</sup> This is in accordance with Hana's (2007: 123) observation made for Czech that "none of the searched corpora contain such a sentence". Conversely, in web corpora we did find examples of pseudodiaclisis of two reflexive CLs: see (77) and (79). As examples (78) and (80) suggest, haplology of one reflexive is also a possible solution. The reader has to bear in mind that the examples in (77) and (78) on the one hand, like those in (79) and (80) on the other, have identical matrix predicates and infinitive complements. In the former pair these are *truditi se* 'try' and *svidjeti se* 'be liked', while in the latter they are *bojati se* 'be afraid' and *odreći se* 'give up'.

- (77) [...] toliko *se*<sub>1</sub> *čovjek* *trudi*<sub>1</sub> *svidjeti*<sub>2</sub> *se*<sub>2</sub> *drugima* [...].  
 that.much REFL man try.3PRS like.INF REFL other  
 '[...] one tries hard to be liked by others [...]' [bsWaC v1.2]
- (78) [...] *da se*<sub>1+2</sub> *ne trudi*<sub>1</sub> *previse svidjeti*<sub>2</sub> *drugima* [...].  
 that REFL NEG try.3PRS too.much like.INF other  
 '[...] that s/he does not try too much to be liked by others [...]'  
 [bsWaC v1.2]
- (79) [...] *koji se*<sub>1</sub> *boje*<sub>1</sub> *odreći*<sub>2</sub> *se*<sub>2</sub> *grijeha* [...].  
 which REFL be.afraid.3PRS give.up.INF REFL sin  
 '[...] who are afraid to renounce their sins [...]' [hrWaC v2.2]
- (80) [...] *jednostavno se*<sub>1+2</sub> *boji*<sub>1</sub> *odreći*<sub>2</sub> *studentskog načina*  
 simply REFL be.afraid.3PRS give.up.INF student way  
*života*.  
 life  
 '[...] he is simply afraid to give up the student way of living.'  
 [hrWaC v2.2]

<sup>36</sup>We do not rule out the possibility that such sentences do exist in the queried BCS corpora. However, designing a CQL query which would be precise, would not require excessive post-hoc manual human checking, and at the same time would have a good recall is challenging.

Example (81a) and its acceptable (81b) and unacceptable (81c) permutation suggest that Rosen (2014: 106, 114) might be right when he claims that in such structures haplology of reflexives is an instance of CC. Haplology of reflexives is possible only if the pronominal CL *mi* ‘me’ climbs as well: compare (81b) and (81c), where the latter example with haplology of reflexives and without CC of the pronominal *mi* is not acceptable. In the former example the pronominal CL *mi* and the reflexive CL *se* probably climbed together, and the latter CL took the position of the reflexive CL *se* which was already present in the matrix clause.

- (81) a. Ne trebaju<sub>1</sub> se<sub>2</sub> bojati<sub>2</sub> približiti<sub>3</sub> mi<sub>3</sub> se<sub>3</sub>.  
 NEG need.3PRS REFL be.afraid.INF approach.INF me.DAT REFL
- b. Ne trebaju<sub>1</sub> mi<sub>3</sub> se<sub>2+3</sub> bojati<sub>2</sub> približiti.  
 NEG need.3PRS me.DAT REFL be.afraid.INF approach.INF
- c. \* Ne trebaju<sub>1</sub> se<sub>2+3</sub> bojati<sub>2</sub> približiti<sub>3</sub> mi<sub>3</sub>.  
 NEG need.3PRS REFL be.afraid.INF approach.INF me.DAT  
 ‘They do not have to fear approaching me.’ [hrWaC v2.2]

Since this constraint concerns not only pronominal but also reflexive CLs, we would like to formulate it slightly more accurately than Hana (2007: 123) did, namely: a mixed CL cluster cannot contain two phonologically (and sometimes morphologically) identical pronominal and reflexive CLs with different governors.

#### 11.4.3 Morphologically different clitics with similar syntactic function and different governors

There are cases in which CLs do not necessarily have the same phonological and morphological form, but CC still does not occur. Scholars dedicated the most attention to different reflexives (*se* vs *si*). Based on example (82) with the reflexive CL *se* in the matrix clause and the reflexive CL *si* in the embedding, Junghanns (2002: 79) shows that two CLs with similar syntactic functions and different governors block CC.<sup>37</sup>

- (82) Řekl jsem mu, jak jsem<sub>1</sub> se<sub>1</sub> jednou rozhodl<sub>1</sub>  
 say.PTCP.SG.M be.1SG him.DAT how be.1SG REFL once decide.PTCP.SG.M  
 trénovat<sub>2</sub> si<sub>2</sub> paměť.  
 train.INF REFL memory  
 ‘I told him how I had once decided to train my memory.’  
 (Cz; Junghanns 2002: 80)

<sup>37</sup>He does not explain what exactly is denoted by “similar syntactic function”.

Discussing the same problem, Rosen (2014: 106) agrees with Junghanns that two different reflexive CLs cannot appear in the same cluster (83a). However, he does not per se rule out the possibility of CC in such structures – CC is possible if the reflexives haploglise (cf. Rosen 2014: 106). In permuted examples (83b) and (84b) with subject control matrix predicates *bát se* ‘be afraid’ and *troufnout si* ‘dare’, the reflexive CLs *si* and *se* climbed into the matrix clause from the embeddings. In contrast to instances of haplology in which matrix reflexives are deleted, according to Rosen (2014: 106) deletion of embedded reflexives leads to sentences whose acceptability is questionable, see (83c) and (84c).

- (83) a. Jan  $se_1$   $bál_1$   $vzít_2$   $si_2$  kravatu.  
 Jan REFL be.afraid.PTCP.SG.M take.INF REFL tie
- b. Jan  $si_{1+2}$   $bál_1$   $vzít_2$  kravatu.  
 Jan REFL be.afraid.PTCP.SG.M take.INF tie
- c. ? Jan  $se_{1+2}$   $bál_1$   $vzít_2$  kravatu.  
 Jan REFL be.afraid.PTCP.SG.M take.INF tie  
 ‘Jan was afraid to put on a tie.’ (Cz; Rosen 2014: 106)
- (84) a. Troufla<sub>1</sub>  $si_1$  usadit<sub>2</sub>  $se_2$  v první řadě.  
 dare.PTCP.SG.F REFL sit.INF REFL in first row
- b. Troufla<sub>1</sub>  $se_{1+2}$  usadit<sub>2</sub> v první řadě.  
 dare.PTCP.SG.F REFL sit.INF in first row
- c. ? Troufla<sub>1</sub>  $si_{1+2}$  usadit<sub>2</sub> v první řadě.  
 dare.PTCP.SG.F REFL sit.INF in first row  
 ‘She dared to sit in the first row.’ (Cz; Rosen 2014: 106)

BCS show some variation with respect to the abovementioned constraint between the three national variants. As presented in Section 6.3.3 standard Bosnian and standard Serbian do not recognise the reflexive CL *si*. This notwithstanding, we found examples in which the reflexives *se* and *si* appear in pseudodiacclisis not only in the Croatian, but also in the Bosnian web corpora.<sup>38</sup> Permutations with CC in Croatian (85b) lead to unacceptable sentences, while permutations with CC in which the embedded reflexive *si* overrides the matrix reflexive CL *se* are marginally possible – see (85d). Permutations with haplology of unlikes and without CC are marginally possible, as are those with haplology of unlikes and with CC – compare (85c) and (85d).

<sup>38</sup>We do not rule out the possibility that such sentences exist also in srWaC, since the reflexive CL *si* is found in dialects spoken on Serbian territory – see Section 7.4.3. However, if they exist in srWaC, they must be rarer than in bsWaC and hrWaC.



- (85) a. [...] prije nego  $se_1$  odvažimo<sub>1</sub> priuštiti<sub>2</sub>  $si_2$  zeru više života.  
 before than REFL dare.1PRS afford.INF REFL little more life
- b. \* [...] prije nego  $se_1$   $si_2$  odvažimo<sub>1</sub> priuštiti<sub>2</sub> zeru više života.  
 before than REFL REFL dare.1PRS afford.INF little more life
- c. ? [...] prije nego  $se_{1+2}$  odvažimo<sub>1</sub> priuštiti<sub>2</sub> zeru više života.  
 before than REFL dare.1PRS afford.INF little more life
- d. ? [...] prije nego  $si_{1+2}$  odvažimo<sub>1</sub> priuštiti<sub>2</sub> zeru više života.  
 before than REFL dare.1PRS afford.INF little more life
- ‘[...] before we dare to allow ourselves to live life a little more fully.’  
 [hrWaC v2.2]

These data indicate that the situation in Bosnian and Croatian is quite similar to the situation in Czech. The only difference is that the examples with haplology of unlikes and with CC (85d) are just as marginally possible as examples with haplology of unlikes and without CC (85c). In addition, examples like the following call into question whether it is possible to apply haplology of unlikes in the case of different reflexives in Bosnian and Croatian.

- (86) a. Dozvoljavam<sub>1</sub>  $si_1$  opteretiti<sub>2</sub>  $se_2$  svim i svačim.  
 allow.1PRS REFL burden.INF REFL everything and anything
- b. \* Dozvoljavam  $si_1$   $se_2$  opteretiti<sub>2</sub> svim i svačim.  
 allow.1PRS REFL REFL burden.INF everything and anything
- c. \* Dozvoljavam<sub>1</sub>  $si_{1+2}$  opteretiti<sub>2</sub> svim i svačim.  
 allow.1PRS REFL burden.INF everything and anything
- d. \* Dozvoljavam<sub>1</sub>  $se_{1+2}$  opteretiti<sub>2</sub> svim i svačim.  
 allow.1PRS REFL burden.INF everything and anything
- ‘I allow myself to burden myself with everything and anything.’  
 [hrWaC v2.2]
- (87) a. [...] pa  $si_1$  dopustimo<sub>1</sub> utopiti<sub>2</sub>  $se_2$  u neke druge.  
 so REFL allow.1PRS drown.INF REFL in some others
- b. \* [...] pa  $si_1$   $se_2$  dopustimo<sub>1</sub> utopiti<sub>2</sub> u neke druge.  
 so REFL REFL allow.1PRS drown.INF in some others
- c. \* [...] pa  $si_{1+2}$  dopustimo<sub>1</sub> utopiti<sub>2</sub> u neke druge.  
 so REFL allow.1PRS drown.INF in some others
- d. \* [...] pa  $se_{1+2}$  dopustimo<sub>1</sub> utopiti<sub>2</sub> u neke druge.  
 so REFL allow.1PRS drown.INF in some others
- ‘[...] so we allow ourselves to drown in other people.’ [bsWaC v1.2]

In Bosnian and Croatian, if the reflexive CL *si* is in the matrix clause and the reflexive CL *se* is in the embedding, the only possible solution is pseudodiacclisis, since neither CC ((86b) and (87b)) nor haplology ((86c), (87c), (86d) and (87d)) lead to acceptable sentences. This is the major difference between Bosnian and Croatian on the one hand and Czech on the other.

## 11.5 How clitics climb

### 11.5.1 Clitic cannot climb over clitic

Hana (2007: 127) and Rosen (2014: 102) claim that in Czech, CC is “monotonic”. This means that a CL can climb to a given cluster only if all CLs with a less embedded governor also climb to that cluster or a higher one ((88b) and (88c)). This is because a CL cannot climb over another CL (88d).<sup>39</sup>

- (88) a. Všichni *jsme*<sub>1</sub> *se*<sub>1</sub> *snažili*<sub>1</sub> *mu*<sub>2</sub> *pomoci*<sub>2</sub> *ho*<sub>3</sub>  
 all be.1PL REFL try.PTCP.PL.M him.DAT help.INF him.ACC  
*najít*<sub>3</sub>.  
 find.INF
- b. Všichni *jsme*<sub>1</sub> *se*<sub>1</sub> *mu*<sub>2</sub> *snažili*<sub>1</sub> *ho*<sub>3</sub> *pomoci*<sub>2</sub>  
 all be.1PL REFL him.DAT try.PTCP.PL.M him.ACC help.INF  
*najít*<sub>3</sub>.  
 find.INF
- c. Všichni *jsme*<sub>1</sub> *se*<sub>1</sub> *mu*<sub>2</sub> *ho*<sub>3</sub> *snažili*<sub>1</sub> *pomoci*<sub>2</sub>  
 all be.1PL REFL him.DAT him.ACC try.PTCP.PL.M help.INF  
*najít*<sub>3</sub>.  
 find.INF
- d. \*Všichni *jsme*<sub>1</sub> *se*<sub>1</sub> *ho*<sub>3</sub> *snažili*<sub>1</sub> *mu*<sub>2</sub> *pomoci*<sub>2</sub>  
 all be.1PL REFL him.ACC try.PTCP.PL.M him.DAT help.INF  
*najít*<sub>3</sub>.  
 find.INF

‘All of us tried to help him find it.’

(Cz; Hana 2007: 127)

<sup>39</sup>We refer the reader to Lenertová (2004: 153) for examples of the lack of CC out of control constructions with CL pairs which would result in inverted CL order. In contrast, according to Lenertová (2004: 153) in Czech CC in the context of object control matrix verbs is not problematic as long as it concerns CL pairs which are never used in inverted order.



## 11.5.2 All-or-nothing constraint

Rezac (2005: 8) claims that in Czech, if CC takes place it is an all-or-nothing phenomenon, i.e. either all the CLs of an embedded verb undergo CC or none do. Diacsis of CLs which were generated in the same infinitive is claimed to lead to unacceptable sentences (cf. Rezac 2005: 8). He illustrates this with two permuted examples. In the first (92b), both pronominal CLs *ti* ‘you’ and *ho* ‘him’ generated by the infinitive *ukázat* ‘show’ climbed together to the matrix clause. In the second (92c), only the pronominal CL *ti* climbed, whereas *ho* stayed in the embedding. According to Rezac (2005: 8), only the former is acceptable.

- (92) a. Jana *chce*<sub>1</sub> *ukázat*<sub>2</sub> *ti*<sub>2</sub> *ho*<sub>2</sub> *zejtra*.  
 Jana want.3PRS show.INF you.DAT him.ACC tomorrow
- b. Jana *ti*<sub>2</sub> *ho*<sub>2</sub> *chce*<sub>1</sub> *ukázat*<sub>2</sub> *zejtra*.  
 Jana you.DAT him.ACC want.3PRS show.INF tomorrow
- c. \*Jana *ti*<sub>2</sub> *chce*<sub>1</sub> *ukázat*<sub>2</sub> *ho*<sub>2</sub> *zejtra*.  
 Jana you.DAT want.3PRS show.INF him.ACC tomorrow  
 ‘Jana wants to show him to you tomorrow.’ (Cz; Rezac 2005: 8)

Rezac does not provide further evidence; neither were we able to find corresponding hypotheses by other authors. The all-or-nothing constraint in Czech thus remains on rather shaky ground.<sup>40</sup> Furthermore, something that is quite the opposite is claimed to be possible in the case of CC out of *da*<sub>2</sub>-complements in BCS. Namely, Stjepanović (2004: 182) observes that two CLs with the same governor do not have to climb together into the matrix clause. She claims that if CLs split, then the only possibility is that the dative climbs while the accusative stays in the *da*<sub>2</sub>-complement (93c), and not vice versa (93d) (cf. Stjepanović 2004: 182).

- (93) a. Marija *želi*<sub>1</sub> *da* *mu*<sub>2</sub> *ga*<sub>2</sub> *predstavi*<sub>2</sub>.  
 Marija want.3PRS that him.DAT him.ACC introduce.3PRS
- b. Marija *mu*<sub>2</sub> *ga*<sub>2</sub> *želi*<sub>1</sub> *da* *predstavi*<sub>2</sub>.  
 Marija him.DAT him.ACC want.3PRS that introduce.3PRS

<sup>40</sup>Alexandr Rosen (p.c.) disagrees with Rezac’s claim that CC has to be an all-or-nothing phenomenon. He claims that the following example in which CLs did not climb together is completely acceptable.

(i) Jana *ti* *chce* *ho* *ukázat* *zejtra*.

Note, however, that this does not have to be an instance of CC; see Junghanns (2002: 67).

- c. Marija *mu*<sub>2</sub>    želi<sub>1</sub>    da *ga*<sub>2</sub>    predstavi<sub>2</sub>.  
 Marija him.DAT want.3PRS that him.ACC introduce.3PRS
- d. \*Marija *ga*<sub>2</sub>    želi<sub>1</sub>    da *mu*<sub>2</sub>    predstavi<sub>2</sub>.  
 Marija him.ACC want.3PRS that him.DAT introduce.3PRS  
 ‘Marija wants to introduce him to him.’

This is in line with our corpus study on CC out of *da*<sub>2</sub>-complements, where we find another example of two CLs in *da*<sub>2</sub>-complement which do not climb together. For more information on this see Section 13.4.

Furthermore, as our permuted examples show, it seems that the all-or-nothing constraint does not even apply to BCS infinitive complements. CLs generated in the same infinitive can undergo pseudodiaclisis. As long as the dative CL climbs, the sentence will stay acceptable.

- (94) a. [...] a    ja nisam<sub>1</sub>    imao<sub>1</sub>    namjeru<sub>1</sub> mijenjati<sub>2</sub> *joj*<sub>2</sub>  
 and I NEG.be.1SG have.PTCP.SG.M intention change.INF her.DAT  
*ga*<sub>2</sub> [...].  
 him.ACC
- b. [...] a    ja *joj*<sub>2</sub>    nisam<sub>1</sub>    imao<sub>1</sub>    namjeru<sub>1</sub> mijenjati<sub>2</sub>  
 and I her.DAT NEG.be.1SG have.PTCP.SG.M intention change.INF  
*ga*<sub>2</sub> [...].  
 him.ACC
- ‘[...] and I had no intention of changing it for her [...].’ [hrWaC v2.2]
- (95) a. [...] kada *su*<sub>1</sub>    *joj*<sub>2</sub>    *ih*<sub>2</sub>    zbog    opasne    infekcije  
 when be.3PL her.DAT them.ACC because dangerous infection  
 stafilokokom    morali<sub>1</sub>    izvaditi<sub>2</sub> [...].  
 staphylococcus must.PTCP.PL.M remove.INF
- b. [...] kada *su*<sub>1</sub>    *joj*<sub>2</sub>    zbog    opasne    infekcije  
 when be.3PL her.DAT because dangerous infection  
 stafilokokom    morali<sub>1</sub>    izvaditi<sub>2</sub> *ih*<sub>2</sub> [...].  
 staphylococcus must.PTCP.PL.M remove.INF them.ACC
- ‘[...] when they had to remove them from her because of a dangerous staphylococcus infection [...].’ [bsWaC v1.2]

Thus, it seems that the all-or-nothing constraint on CC, reported for Czech by Rezac (2005), is not relevant for CC in BCS. As BCS examples in this section suggest, one of the two complement CLs can climb while the other can stay in the

complement as long as it is the one which comes later in the CL cluster according to the ordering rules.<sup>41</sup>

## 11.6 Sentential negation

Sentential negation has not been discussed by scholars who researched CC in Czech, but it has been noticed in the literature on CC in BCS.<sup>42</sup> Aljović (2004: 3f, 2005: 6) was the first who claimed that sentential negation blocks CC in BCS. The permutation in (96b) demonstrates how CC out of *da*<sub>2</sub>-complements with negation leads to unacceptable sentences, whereas (97b) illustrates the same but for CC out of a negated infinitive complement.<sup>43</sup>

- (96) a. Marija želi<sub>1</sub> da ga<sub>2</sub> ne vidi<sub>2</sub>.  
 Marija want.3PRS that him.ACC NEG see.3PRS
- b. \*Marija ga<sub>2</sub> želi<sub>1</sub> da ne vidi<sub>2</sub>.  
 Marija him.ACC want.3PRS that NEG see.3PRS  
 ‘Marija wants to not see him.’ (BCS; Aljović 2004: 3)
- (97) a. Marija želi<sub>1</sub> ne sresti<sub>2</sub> ga<sub>2</sub> nigdje.  
 Marija want.3PRS NEG meet.INF him.ACC nowhere
- b. \*Marija ga<sub>2</sub> želi<sub>1</sub> ne sresti<sub>2</sub> nigdje.  
 Marija him.ACC want.3PRS NEG meet.INF nowhere  
 ‘Marija wants to meet him nowhere.’ (BCS; Aljović 2004: 4)

In her second paper, Aljović (2005: 7) extends her claims from the first paper and argues that negation in *da*<sub>2</sub>-complements always blocks CC. However, in the case of infinitives CC is blocked only if there is a negative polarity item, like *nigdje* ‘nowhere’ in example (97a). Thus, negation in the infinitive without a negative polarity item does not obligatorily block CC according to Aljović (2005: 7): compare (97b) and (98b).

- (98) a. Ona više voli<sub>1</sub> ne vidjeti<sub>2</sub> ga<sub>2</sub>.  
 she more love.3PRS NEG see.INF him.ACC

<sup>41</sup>For the relative order of CLs in the clitic cluster see Section 2.4.2.1.

<sup>42</sup>This may be due to the fact that negation seems to allow CC in Czech (Alexandr Rosen, p.c.).

<sup>43</sup>Here we would like to comment that some scholars do not agree that such sentences are possible at all. Todorović (2012: 168) claims that “In respect to negation, indicative and subjunctive *da*-complements differ in that negation can precede the embedded verb in indicative complements but cannot precede the embedded verb in subjunctive complements”. In other words, she claims that negation within the *da*<sub>2</sub>-complement is not possible.

- b. Ona  $ga_2$  više voli<sub>1</sub> ne vidjeti<sub>2</sub>.  
 she him.ACC more love.3PRS NEG see.INF  
 ‘She likes not seeing him more.’ (BCS; Aljović 2005: 7)

To explain the difference in the behaviour of infinitives and  $da_2$ -complements with respect to CC and negation, she introduces sentential negation as a constraint to CC (Aljović 2005: 7). Namely, she claims that in the case of  $da_2$ -complements there is no doubt that the negation is sentential. Moreover, the same applies to negated infinitives with negative polarity items, since only sentential negation can license them (cf. Aljović 2005: 7). In the case of negated infinitives without negative polarity items, the negative particle can be interpreted as lexical negation (cf. Aljović 2005: 7). Unlike sentential negation in (96b) and (97b), constituent negation in (98b) does not block CC (cf. Aljović 2005: 7).

We agree with Aljović that CLs can climb out of negated infinitive complements without a negative polarity item, since the permutation (99b) of the example found in hrWaC (99a) is completely acceptable to our informants.

- (99) a. Odučila<sub>1</sub> sam<sub>1</sub> ne pokazati<sub>2</sub> joj<sub>2</sub> ga<sub>2</sub>.  
 decide.PTCP.SG.F be.1SG NEG show.INF her.DAT him.ACC  
 b. Odučila<sub>1</sub> sam<sub>1</sub> joj<sub>2</sub> ga<sub>2</sub> ne pokazati<sub>2</sub>.  
 decide.PTCP.SG.F be.1SG her.DAT him.ACC NEG show.INF  
 ‘I decided to not show him to her.’ [hrWaC v2.2]

## 11.7 Constraints related to information structure

### 11.7.1 Infinitive as a whole as the topic of a sentence

Bošković (2001) was the first to notice that there is no CC in BCS if the infinitive complement is fronted. The slightly later work of Stjepanović (2004: 182f) provides examples which support this claim. In example (100) the infinitive complement *sresti* ‘meet’ is fronted. Therefore the pronominal accusative CL *ga* ‘him’ does not form a mixed cluster with the verbal CL *je* ‘is’.

- (100) Sresti<sub>2</sub> ga<sub>2</sub> u Kanadi, Dragan je<sub>1</sub> želio<sub>1</sub>.  
 meet.INF him.ACC in Canada Dragan be.3SG want.PTCP.SG.M  
 ‘Dragan wanted to meet him in Canada.’ (BCS; Stjepanović 2004: 182)

Junghanns (2002) has similar observations regarding Czech. He claims that CLs stay in their position if the embedded infinitive as a whole is the topic of a sentence (cf. Junghanns 2002: 78), for which he provides the following example (101):

## 11 Constraints on clitic climbing in Czech compared to BCS

- (101) Chovat<sub>2</sub> se<sub>2</sub> v souladu se svým svědomím nemá<sub>1</sub>  
behave.INF REFL in harmony with own conscience NEG.have.3PRS  
prý žádnou cenu [...].  
allegedly any value  
'Behave in accordance with conscience has no value [...].'  
(Cz; Junghanns 2002: 78)

### 11.7.2 Infinitive as a whole as the focus of a sentence

Junghanns (2002: 78) and Dotlačil (2004: 98) note regarding Czech that if the embedded infinitive as a whole (together with its CL complements) is the focus of the sentence or is a part of the focus, CC will not occur.

- (102) [...] kteří čas od času přicházeli<sub>1</sub> se<sub>2</sub> mu<sub>2</sub> posmívat<sub>2</sub>.  
which time from time come.PTCP.PL.M REFL him.DAT laugh.INF  
'[...] who came to mock him from time to time.' (Cz; Junghanns 2002: 79)

Junghanns (2002) further argues that climbing of the pronominal dative CL *mu* 'him' and the reflexive CL *se* in sentence (102) would not lead to an ungrammatical sentence, but would definitely change its informational structure.<sup>44</sup>

Constraints on CC which are related to information structure have been noticed both by scholars studying CC in Czech and those studying it in BCS. However, we must point out that there is relatively little literature on the phenomenon.

## 11.8 Summary

### 11.8.1 Overview

In this chapter we focused on constraints on CC which have been described in the reviewed literature on this phenomenon in Czech and/or in BCS. In our analysis, we did not take into account structures described for Czech in Junghanns (2002) which are not attested in BCS.

---

<sup>44</sup>Alexandr Rosen (p.c.) disagrees with Junghanns. He argues that there is no difference between (102) and (i):

- (i) [...] kteří se<sub>2</sub> mu<sub>2</sub> čas od času přicházeli<sub>1</sub> posmívat<sub>2</sub>.  
which REFL him.DAT time from time come.PTCP.PL.M laugh.INF  
'[...] who came to mock him from time to time.'



As already stated, our aim was to give a maximally adequate descriptive account of the possible constraints on CC in BCS. Therefore, we tried to pretest constraints on CC in the natural language environment provided by {bs,hr,sr}WaC. Furthermore, sometimes because of the problem of negative evidence we used informal acceptability judgments where sentences in each language were evaluated by at least five native speakers.

In addition, we would like to point out that even for Czech the inventory of constraints is based on very sparse natural data (normally just a minimal pair of sentences per constraint). We cannot but wonder if real empirical studies (corpus linguistic or psycholinguistic) would corroborate constraints on CC reported for Czech in the theoretical syntactic literature. Furthermore, the authors sometimes tried to offer explanations, but they had to admit that counterexamples could be found. We summarise our main findings in Tables 11.1 and 11.2. However, we are aware that these are only to be considered preliminary results. The constraints described and marked “yes” are generalisations about a potentially large set of data; however, they were created on a very small number of examples, like in the case of CC in Czech. To further validate these constraints, we have to look for appropriate, representative and bigger samples: i.e., in order to establish whether the constraints on the list marked “yes” really operate as constraints in all three South Slavonic languages or only in some of them, more robust evidence should be found.

### 11.8.2 Island constraints

Junghanns (2002) observes that in Czech CLs do not climb out of gerund phrases (see Section 11.2.3) and adjective phrases (see Section 11.2.4). However, these structures are not described in the reviewed literature on CC in BCS. Therefore we permuted sentences from {bs,hr,sr}WaC and pretested them via informal acceptability judgments. We finally came to the same conclusion as Junghanns (2002), i.e., there is no doubt that the mentioned constraints operate in both Czech and BCS for both pronominal and reflexive CLs.

Both scholars working on Czech (e.g. Junghanns 2002, Dotlačil 2004, Rezac 2005) and on BCS (e.g. Aljović 2005) recognize *wh*-infinitives as a constraint on CC (see Section 11.2.6). Our corpus-based examples and their unacceptable permutations support the claims from the theoretical literature on syntax.

Junghanns (2002) was the only one to note that CC from an infinitive which is a complement of a noun in a prepositional phrase is blocked (see Section 11.2.5.2). Since this constraint was not mentioned in the literature on CC in BCS, we first queried BCS web corpora and then conducted informal acceptability judgments

with native speakers. As our informants did not accept the permutations of sentences from corpora, we assume that the same constraint operates in BCS as well.

Junghanns (2002) notes one more island constraint for CC in Czech (see Section 11.2.1). CLs cannot climb out of infinitives in comparative sentences with *než*. We pretested this constraint via informal acceptability judgments. Our native speaker informants did not accept permuted CC versions of the sentences with CLs in *ne-go* infinitives without CC found in web corpora. Therefore, on the basis of our tentative results we can assume that this constraint on CC operates in BCS as well. This is the last of the four island constraints shared by Czech and BCS.

It seems that the infinitive as a complement of an agreeing predicative adjective (see Section 11.2.5.3) is a constraint on CC in Czech. However, Junghanns (2002), who was the first to observe this phenomenon, acknowledges that counterexamples do exist. Consequently, he admits that this restriction has to be studied more thoroughly. Based on Junghanns' examples, our assumption is that in Czech this constraint operates only in the case of reflexives. In contrast, as our tentative corpus-based research indicates, this constraint does not operate in BCS at all. This is one of many differences in CC between BCS and Czech.

The constraint termed: infinitives as complements of non-agreeing predicatives (see Section 11.2.5.4) was first observed for Czech by Junghanns (2002) and further described by Dotlačil (2004). In Czech, this constraint applies only to reflexives, while pronominal CLs can freely climb out of infinitives which are complements of non-agreeing predicatives. Our data show that unlike in Czech, in BCS this restriction does not apply to reflexive CLs. As the corpus data reveal, in all three South Slavonic varieties pronominal CLs can, like in Czech, climb out of an infinitive which is a complement of a non-agreeing predicative.

Junghanns (2002) was the first to observe that in Czech, CLs do not climb out of infinitives which are complements of nouns, i.e. light verb constructions (see Section 11.2.5.1). However, he acknowledges that there are exceptions to this constraint. Namely, in Czech CC is possible if the verbal part of a light verb construction bears no or almost no content. However, from BCS examples found in web corpora it seems that the amount of semantic content in the verbal part of light verb constructions does not play a role in blocking CC.

Clauses with an inflected verb are one more island constraint which seems to operate differently in Czech and BCS (see Section 11.2.2). Scholars unanimously agree that there is no CC out of finite clauses in Czech. However, both BCS scholarly literature and {bs,hr,sr}WaC provide sentences in which CLs climb out of a complement with a verb inflected for person, the *da<sub>2</sub>*-complement. The empirical data and discussion of this question can be found in Chapter 13.

### 11.8.3 Constraints related to the raising–control distinction

Scholars do not agree whether climbing of pronominal CLs out of object-controlled infinitives is possible in Czech (see Section 11.3.1). While Thorpe (1991) and Junghanns (2002) believe that it is impossible, George & Toman (1976), Dotlačil (2004), Rezac (2005), and Hana (2007) allow it iff some additional conditions are fulfilled. For BCS Aljović (2005) is the only scholar who claims that an indirect object in the matrix clause does not necessarily have to block climbing of pronominal CLs out of an infinitive complement. Our first tentative corpus data seem to corroborate her claim.

Dotlačil (2004) argues that it is not control by itself which blocks climbing of pronominal CLs in Czech, but that person and case are the additional features which do so (see Section 11.3.3). Namely, an accusative complement in the matrix clause blocks the climbing of all CLs except the third person accusative CL. However, it seems that Rezac (2005) does not share Dotlačil's opinion. Namely, he claims that the only additional feature which prevents CLs from climbing is case, i.e., accusative case of the complement in the matrix clause blocks all CC (see Section 11.3.2). Unlike Dotlačil (2004) and Rezac (2005), George & Toman (1976) consider animacy of the CL referent to be the only additional feature which can stop a CL from climbing out of object controlled infinitives, i.e., only CLs with inanimate referents can climb (see Section 11.3.4).<sup>45</sup>

While climbing of pronominal CLs is blocked by a combination of object control and other features according to Hana (2007) CC of reflexives out of object controlled infinitives is completely impossible in Czech (see Section 11.3.5). We pretested this constraint using permutations of examples from web corpora. The first tentative results of informal acceptability judgments made by our informants confirm that the constraint in question operates in BCS as well. However, it is important to emphasise that in Czech, according to Hana (2007), and in BCS,

---

<sup>45</sup>The striking difference between Dotlačil (2004) and George & Toman (1976) becomes more obvious if we compare their examples presented in (63a) and in (65b) respectively, the latter marked by George & Toman (1976: 245) as incorrect. First, in both cases the embedded infinitive is *navštívit/navštěvovat* 'visit', once in perfective and once in its imperfective form. Second, the complements of the embedded infinitive are accusative CLs with animate referents. Why do George & Toman (1976: 245) evaluate their sentence as unacceptable and Dotlačil (2004: 80) as acceptable? This may be one of the best examples showing that syntactic theories lie on shaky grounds, being based on informal acceptability judgments made by linguists and probably deliberately chosen to support their theories. We admit that it is possible that differences in evaluations are due to differences in authors' dialects or idiolects. However, this is exactly why we see the necessity of a serious empirical approach to syntactic problems as argued in Chapter 3.

according to our tentative exploration of {bs,hr,sr}WaC, raising and subject control CTPs do not block reflexive CLs from climbing.

#### 11.8.4 Constraints related to mixed clitic clusters

The pseudo-twins constraint (see Section 11.4.1) was first described by Junghanns (2002) on examples with reflexive CLs and later elaborated on by Hana (2007) and Rosen (2014). These scholars unanimously agree that the nature of the constraint in question is not phonological.

In contrast to Junghanns (2002), Rosen (2014) argues that in Czech, CC is possible in the case of reflexives since the more embedded reflexive overrides the less embedded one. If reflexives are phonologically identical and have different governors, CC is possible in the form of haplology (see Section 11.4.2). Similarly, if reflexives are morphologically different and have different governors, CC is possible in the form of haplology of unlikes (see Section 11.4.3).

Unlike for reflexive CLs, as Hana (2007) observes, CC is not possible in Czech if pronominal CLs are phonologically (and morphologically) identical and have different governors.

So it seems that there are differences among CLs. This goes well with the hypothesis put forward by Dotlačil (2004: 82f): that with respect to CC, CLs cannot be treated as one homogeneous class, since they do not behave in the same way.<sup>46</sup> The data from Czech and BCS discussed in Sections 11.2.5.3, 11.2.5.4 and 11.3.5 indicate exactly that: reflexive CLs seem to behave differently than pronominals. Furthermore, from the work of Lešnerová & Malink (2008) and Rosen (2014) it has become clear that different types of reflexives do not behave in a uniform way. We agree with the scholars who pointed out that not only do pronominal and reflexive CLs differ from each other, but also that reflexive CLs form a heterogeneous group. These differences are important factors which cannot be neglected in the research on CC in BCS. Therefore, we included them as variables in our empirical studies on CC in Croatian described in Chapters 14 and 15.

Since these constraints were not recognised in the literature on CC in BCS, we pretested them through informal acceptability judgments of permutations of examples from web corpora. Our tentative results suggest that this constraint does apply to BCS. However, in contrast to Czech, haplology of reflexives is not always a possible solution. According to our first results for BCS, haplology can be applied only to phonologically identical reflexive CLs.

---

<sup>46</sup> “[...] all clitics were treated as one homogenous class and were expected to behave same. The point of both subsections is precisely against this treatment of clitic climbing.” (Dotlačil 2004: 82f).

In addition, we would like to emphasise that the situation of two CLs with different governors has to be observed coherently with respect to the distinction of two different control predicate types: subject and object. See Section 11.3 on features which are strongly correlated with the control constraint.

### 11.8.5 How clitics climb

Two constraints are related to the way in which CLs climb (see Section 11.5). According to Hana (2007) and Rosen (2014) CC in Czech is monotonic, i.e., the more embedded CLs cannot climb unless the less embedded CLs climb as well (see Section 11.5.1). Rezac (2005) claims that in Czech CC is an all-or-nothing phenomenon, i.e. either all CLs governed by the embedded infinitive climb or none do (see Section 11.5.2).

However, it is important to note that the latter constraint might apply only to Czech. Namely, according to data in the literature on CC out of *da*<sub>2</sub>-complements (e.g. Stjepanović 2004) and the results of our corpus study in Section 13.4 and permutations of examples of CC out of infinitive complements found in {bs,hr,sr}WaC, the all-or nothing constraint does not apply in BCS.

### 11.8.6 Sentential negation

Aljović (2004, 2005) is the only scholar who elaborates on the sentential negation constraint on CC in BCS. She claims that negation in *da*<sub>2</sub>-complements always blocks CC (see Section 11.6). However, it is not completely clear whether such constructions are possible at all, since Todorović (2012) claims that negation of *da*<sub>2</sub>-complements is not possible. In contrast, as Aljović (2005) and our corpus-based examples show, negation can be found within infinitive complements. However, negation does not necessarily function as a constraint on CC from infinitive embeddings in BCS. CC is blocked only if there is a negative polarity item within the infinitive clause: otherwise CLs can climb.

### 11.8.7 Constraints related to information structure

Two constraints linked to the information structure of a sentence are reported in the literature on CC in Czech and BCS. Bošković (2001), Stjepanović (2004), and Junghanns (2002) agree that CLs do not climb out of fronted infinitive complements, i.e. out of an infinitive which is the topic of a sentence. BCS and Czech share this constraint (see Section 11.7.1). In addition, Junghanns (2002) and Dotlačil (2004) argue that in Czech CLs cannot climb out of an embedded infinitive which is the focus of a sentence (see Section 11.7.2).

11 Constraints on clitic climbing in Czech compared to BCS

Table 11.1: Overview of tentative constraints for Czech and BCS

Constraint type	Subtype	Subsubtype	Czech	BCS	
2. island constraints	2.1. infinitives in comparative sentences with <i>než/nego</i>		yes	yes	
	2.2 clauses with inflected verbs		yes	partially	
	2.3 phrases with gerunds		yes	yes	
	2.4 adjective phrases		yes	yes	
	2.5 depth and kind of embeddedness of infinitive phrases		2.5.1. infinitives as complements of nouns	not clear	no
			2.5.2. infinitives as complements of nouns in prepositional phrases	yes	yes
			2.5.3. infinitives as complements of agreeing predicative adjectives	not clear	no
			2.5.4. infinitives as complements of non-agreeing predicatives	only for REFL	no
		2.6. embedded interrogative wh-infinitives		yes	yes
	3. constraints related to the raising-control distinction	3.1. object-controlled complements	3.1.1. object control constraint related to case	yes	yes
3.1.2. object control person-case constraint			not clear	no	
3.1.3. object control and animacy of the referent of the CL constraint			not clear	no	
3.1.4. object control reflexive constraint			yes	yes	

(Continuation in Table 11.2)

Table 11.2: Continuation of Table 11.1

Constraint type	Subtype	Subsubtype	Czech	BCS
4. constraints related to mixed CL clusters	4.1. pseudo-twins	4.1.1. phonologically identical pronominal and reflexive CLs with different governors	yes	yes
		4.1.2. morphologically different CLs with similar syntactic function and different governors	yes	yes
5. how CLs climb	5.1. CL cannot climb over CL		yes	yes
	5.2. all-or-nothing constraint		not clear	no
6. sentential negation			not described	not clear
7. constraints related to information structure	7.1. infinitive as a whole as the topic of a sentence		yes	yes
	7.2. infinitive as a whole as the focus of a sentence		yes	not described

## 11.9 Further Perspectives

In this chapter we showed that Czech and BCS share the following island restrictions on CC: gerunds or adverbial participles respectively, adjective phrases, infinitives as complements of nouns in prepositional phrases and embedded *wh*-infinitives, as well as one constraint caused by comparative sentences with *než/nego*. Other island constraints reported for Czech, such as finite clauses, infinitives as complements of nouns (i.e. light verb CTPs), infinitives as complements of agreeing predicative adjectives, and infinitives as complements of non-agreeing predicatives seem mostly – in the small scope that is covered by literature and according to our first tentative results – not to operate in BCS. However, apart from the previous five relatively clear constraints on CC in BCS, there are some less clear cases. The case of the reflexive CL *se/si* reported for Czech as the pseudo-twins constraint and the control constraint turns out to be particularly intriguing. While the former constraint has to be systematically linked to the difference between subject and object control predicate types, the latter should be investigated in the context of other features such as case, person, animacy, CL

type (reflexive vs pronominal). According to our first tentative results, some of these features could be important for CC in BCS as well.

Additionally, we saw that some features relevant to CC seem to interact with each other, but we do not know exactly how. These features are: predicate type (control vs raising), CL type (pronominal vs reflexive) and those related to the mixed CL cluster under the label pseudo-twins. The last set of features has not been systematically described under the control distinction, but we believe they cannot be separated. Namely, in the case of pseudo-twins, if one reflexive CL is in the matrix clause and the other in the infinitive clause, then the matrix verb can be either of the subject control type (*veseliti se* ‘look forward to’, *odlučiti se* ‘decide’) or object control type (*prisiliti se* ‘force oneself’). Hence, if a pseudo-twins constraint is present, there is still the factor of object control, which cannot be neglected. Furthermore, since differences are reported in climbing of different CL types, even within the group of reflexive CLs, for valid results this feature has to be tested in combination with different control predicates as well.

This discussion of Czech and BCS shows that putative blocking effects on CC in object-controlled complements in BCS are worth investigating (see Chapter 15). As we have already stated, blocking effects seem to arise from the combination of control and some other features. We will discuss in more detail the link between the control (and raising) distinction and CC in Chapters 13–15.



# 12 Introductory remarks to corpus studies on clitic climbing

## 12.1 Corpus-driven studies on clitic climbing

In the next two chapters we present two corpus studies designed to investigate some tentative constraints on CC in BCS, formulated in the previous chapter where we compared CC in BCS and in Czech. The two studies are methodologically similar. As explained in Section 3.3, our approach to CC is empirical and inductive. Instead of assuming any universal rules, we are interested in statistical tendencies and regularities. We do not reject any constructions before examining real instances of them in corpora, and potentially testing them further on informants. We suspect that structures have often been rejected due to their infrequency. Hence, we first try to retrieve all permuted structures from large corpora, also the supposedly incorrect ones.<sup>1</sup>

In order to do that, in Section 12.2 we define the construction which we investigated in terms of CC and show how we formulated the appropriate queries. In Section 12.3, we argue which sources are the most suitable for retrieving the studied constructions. Section 12.4 explains how complement-taking predicates (CTPs) were sampled for the study. Data analysis is presented in the next chapters; in this chapter we explain only the data collection process.

## 12.2 Operationalising the constructions in question

As explained in Chapter 10, we study constructions containing two verbal elements. These are *da*<sub>2</sub>-constructions in Serbian (Chapter 13) and infinitive complement constructions in Croatian (Chapter 14).<sup>2</sup> In both constructions, several positions of the CL complement in relation to the complement-taking predicate and verbal complement are potentially possible, as shown in Table 12.1.<sup>3</sup>

---

<sup>1</sup>For more information on large corpora see Section 4.4.

<sup>2</sup>See Section 2.5.3 for more information on *da*<sub>2</sub>-constructions.

<sup>3</sup>For more information on complement-taking predicates see Section 2.5.1.

Table 12.1: Permuted constructions with verbal complements.

Complement type	noCC		CC	
	Variant 1	Variant 2	Variant 3	Variant 4
<i>da</i> <sub>2</sub>	CTP DA COMP CL	CTP DA CL COMP	CTP CL DA COMP	CL CTP DA COMP
infinitive	CTP INF CL	NA	CTP CL INF	CL CTP INF

Occurrences of each variant of the constructions in question can be retrieved from the corpora by means of CQL queries in which we can combine morphosyntactic tags with word form and lemma-based attribute search.<sup>4</sup> To explain the logic behind the CQL queries in our studies, we provide an example of variant 4 with an infinitive complement from Chapter 14, where only the third person pronominal and reflexive CLs were retrieved, while CTPs were restricted to the present tense form. Following the template from Table 12.1, the basic CQL query, according to the current MSD index used for hrWaC, should be as follows:<sup>5</sup>

```
[ (word=" ([mj]u) | (joj) | (i[hm]) | (ga) | (se) | (je) | (si) " ] [tag="Vmr.*"]
[tag="V.n"]
```

The first segment in the query encodes the expression for third person pronominal and reflexive CLs via their word forms, while the query for a present tense indicative form of a CTP (second segment) and an infinitive (third segment) is performed via their morphosyntactic tags.<sup>6</sup>

This, however, is not enough, as for example the forms of two CLs, the pronominal CL *je* and the reflexive CL *si*, are ambiguous since they are homographic with verbal CLs.<sup>7</sup> The correction shown below decreases the likelihood that non-target CLs, i.e. verbal instead of pronominal or reflexive, will appear in the result of a query:

```
[ (word=" ([mj]u) | (joj) | (i[hm]) | (ga) | (se) " ) | (word=" (je) | (si) "&
tag!="V.*" ) ]
```

This is, naturally, under the assumption that the tagger has high accuracy. Prior to using tag attributes we examined frequency lists to ensure that a given tag is

<sup>4</sup>For a basic description of CQL see <https://www.sketchengine.eu/documentation/cql-basics/>.

<sup>5</sup>For the current MSD index used for hrWaC visit <http://nl.ijs.si/ME/V6/msd/html/msd-hbs.html#sd.msds-hbs>.

<sup>6</sup>Each segment is specified in square brackets.

<sup>7</sup>A list of all CL forms used in BCS standard varieties can be found in Section 6.3.

## 12.2 Operationalising the constructions in question

not too prone to error. Still, errors cannot be eliminated completely. For example, the tagger sometimes interprets the verbal CL *je* ‘is’ as the accusative form of the third person feminine pronominal, as in (1). A similar error occurs for the verbal CL *si* ‘are’, which is interpreted in (2) as the reflexive CL in the dative. In order to better explain the problem, we introduce an additional line of glosses containing morphosyntactic tags from hrWaC.

- (1) Prva stvar koju trebate zapamtiti je ne paničariti.  
 first thing which have.to.2PRS remember.INF be.3SG NEG panic.INF  
 Mlofsn Ncfsn Pi-fsa\* Vmr2p Vmn Pp3fsa\* QZ Vmn  
 ‘The first thing you have to remember is not to panic.’ [hrWaC v2.2]
- (2) Onoga, koga treba banirati si upravo ti [...].  
 that which have.to.3PRS ban.INF be.3SG exactly you  
 Pd-nsg\* Pi3m-a\* Vmr3s Vmn Px-sd\* Rgp Pp2-sn  
 ‘You are the one who should be banned [...].’ [hrWaC v2.2]

The first two words in example (1) are tagged correctly, whereas the relative pronoun *koju* ‘which’ is misclassified as an indefinite pronoun. It is worth pointing out that although the relative pronoun class does exist in the tagset description, the query [tag="Pr.\*"] returns an empty result: hence, the tagger does not recognise this class. The misclassified relative pronoun is followed by two correctly tagged words. However, the verbal CL *je* ‘is,’ which has the function of copula, is due to its homographic form and its position misclassified as a pronominal CL. This is because the automatic tool does not perform a syntactic analysis and the infinitive *zapamtiti* ‘remember’ is more likely to be followed by an accusative phrase (a direct object) than by a copula. Thus, it is wrongly tagged as the third person singular feminine personal pronoun in the accusative. The last two words in example (1) are tagged correctly.

In example (2), three out of seven words are mistagged. The tagger assigned the wrong case and gender to the demonstrative pronoun: *onoga* ‘that’ is masculine and not neuter. It is in the accusative and not in the genitive case. Similarly to *koju* in example (1), the relative pronoun *koga* ‘which’ is misclassified as indefinite. Whereas the infinitive, adverb and personal pronoun are tagged correctly, the verbal CL *si* ‘are’ is misclassified due to its homographic form and its position, since the infinitive *banirati* ‘ban’ is more likely to be followed by a dative phrase (the indirect object). The verbal copula is falsely interpreted as the reflexive CL *si* in the dative.

The basic query shown above, consisting only of CORE ELEMENTS, that is, the elements belonging to the target construction, gives low recall, particularly in the



## 12.2 Operationalising the constructions in question

either to the preceding, or to the following predicate. An example of the former is (4), in which the CL *mu* ‘him’ is part of the structure *su mu u pripremi* and is not governed by the infinitive *pročitati* ‘read’. In (5) the CL *ih* ‘them’ is a complement of *pronaći* ‘find’ and not of *potruditi se* ‘try’; that is, it is governed by the following predicate, and not by the target predicate.<sup>9</sup>

- (4) Pitanja koja *su*<sub>1</sub> *mu*<sub>1</sub> već u pripremi<sub>1</sub> možete<sub>2</sub>  
 questions which be.3PL him.DAT already in preparation can.2PRS  
 pročitati<sub>3</sub> pod komentarima [...].  
 read.INF under comments  
 ‘You can read the questions for him which are already in preparation in  
 the comments [...].’ [hrWaC v2.2]
- (5) Mislim da *se*<sub>2</sub> za tratorie i restorane treba<sub>1</sub> malo  
 think.1PRS that REFL for trattorias and restaurants have.to.2PRS little  
 potrudit<sub>2</sub> *ih*<sub>3</sub> pronac<sub>3</sub> [...].  
 try.INF them.ACC find.INF  
 ‘I think that for trattorias and restaurants one has to try a little bit to find  
 them [...].’ [hrWaC v2.2]

As a result, we obtained more complex but better performing queries (in this particular case, for variant 4 from Table 12.1). This allowed us to extract more and better data for our studies. Regardless of the improvements, one should bear in mind that some level of error is unavoidable. The full query is shown below:

```
[!(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}
```

```
[(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)"))){0,4}
```

<sup>9</sup>Both infinitives in that example are written without the final vowel *-i*, which is a feature of colloquial BCS. Furthermore, the infinitive *pronaći* ‘find’ is also written without a diacritic, which is a feature of the language used in user generated content.

```
[lemma="sramiti"␣&␣tag="V.r.*"]
```

```
[!(tag="C.*" | lemma="\Z" | tag="P[iq].*" | tag="V.*" | tag="Rr" | word=
".*\..*" | lemma="što" | word="(me)|([mj]u)|([oj])|([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}
```

```
[tag="V.n"␣&␣lemma!="biti"]
```

```
[!(tag="C.*" | lemma="\Z" | tag="P[iq].*" | tag="V.*" | tag="Rr" | word=
".*\..*" | lemma="što" | word="(me)|([mj]u)|([oj])|([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}within<s/>
```

In our last study on infinitive complements in Croatian, we compared data from the Forum subcorpus of hrWaC with data from corpora of standard Croatian: Riznica and CNC. The latter corpus uses an older tag set, and for some reason does not allow comparably complex queries.<sup>10</sup> Therefore, the results from CNC were obtained via 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 results, we filtered out all instances containing a CL within seven words of the CTP. After that we identified instances of embedded complements which were no further than ten tokens after the CTP, and then excluded the occurrences of *da* up to 10 tokens after the CTP to avoid the *da*-complements. This simplified procedure of data collection with multiple filtering was also the reason why CNC was used only as a complementary source of standard Croatian.

### 12.3 Choice of corpora

A consequence of using big data is the necessity of relying on search engine efficiency and precision of queries. As explained in Chapter 4, the next prerequisite of a good corpus after size is availability of a search mechanism. This boils down to a search engine, filtering options and extensive, precise morphosyntactic annotation of structures in the database.

<sup>10</sup>For the tag set used in CNC visit <http://nl.ijs.si/ME/V4/msd/html/msd-hr.html>.

Finally, as the focus of the study is not on one relatively homogeneous language, but on three closely related languages, the examined material should be comparable in at least some aspects, such as age of texts, size of data, and possibly text type. Bearing these factors in mind, in Section 4.6.3 we conclude that the most suitable source of data for studying CC in BCS is, in our view, available web corpora. Therefore, in both studies we used the WaC family.

In Serbian, the  $da_2$ -complement is a construction that competes against infinitive complements. This is why in the case of  $da_2$ -complements of CTPs, CC and noCC structure variants were retrieved from srWaC. Additionally, we used the Serbian version (Adamovičová & Vavřín 2020) of InterCorp (Čermák & Rosen 2012) (which has an identical tag set) to establish the set of matrix verbs with which this construction appears the most often.<sup>11</sup>

Our decision concerning the data source for CC out of infinitive complements was based on the relative frequency of infinitive complements in comparison to  $da_2$ -complements. Unlike in Serbian, in Croatian infinitive complements dominate with raising and subject control CTPs, and are possible to a certain extent even in the case of object control matrix predicates.<sup>12</sup> Therefore, the data were taken from hrWaC. Additionally, in the second corpus study we collected data from Riznica and CNC so that we could test whether diaphasic variation as factor has an impact on the frequency of CC. As explained in Section 4.6.3, in 2018 the accessibility of Riznica improved, which allowed us to use this corpus. However, as described in the previous section, CNC was treated only as a complementary source of standard Croatian, since the simplified queries with multiple filtering do not allow for full equivalence of queries.

## 12.4 Choice of matrix verbs

When studying  $da_2$ -constructions and infinitive complements, we first constructed CTP frequency lists from which we chose CTPs. In the chronologically first study, we distinguished only three types of CTPs, basing on the raising-control distinction existing in Czech for constraints on CC.<sup>13</sup>

Each set of queries was performed separately for each matrix verb. One position in the query thus became more specific, which had a positive impact on the precision of queries.

<sup>11</sup>A detailed discussion of what motivated this choice can be found in Jurkiewicz-Rohrbacher, Kolaković & Hansen (2017).

<sup>12</sup>For more information on the difference between raising, subject and object control matrix predicates see Section 2.5.2.

<sup>13</sup>For more information on this topic see Section 11.3.

In the case of  $da_2$ -constructions, the list of CTPs was based on the core version of the Serbian InterCorp (Čermák & Rosen 2012), which contains only original Serbian literary works and is manually aligned. In order to retrieve the list of CTPs, we constructed four CQL queries based on the four possible variants given in Table 12.1. We obtained a frequency list with the 42 matrix verbs from InterCorp which have  $da$ -complements. From this list, we first removed matrix predicates which have  $da_1$  complements.<sup>14</sup> Next, we excluded two unwanted types of  $da_2$ -predicates: reflexive and polyfunctional. This allowed us to avoid impersonal and passive constructions in the case of reflexive verbs. We also wanted to avoid sentences with pseudo-twins, which usually lead to pseudodiaclisis or haplology/haplology of unlikes.<sup>15</sup> Polyfunctionality influences the type of complement on one hand and the syntactic type of the matrix on the other. Since CTPs such as *znati* ‘know’, *ht(j)eti* ‘want/will’ and *morati* ‘must’ may take both  $da_2$ - and  $da_1$ -complements, they were not excluded in the first step with other CTPs which take  $da_1$ -complements. However, although they can take  $da_2$ -complements, due to their polyfunctionality we decided to exclude them in order to avoid excessive manual filtering of unwanted utterances with  $da_1$ -complements. Further, to eliminate polyfunctionality with respect to syntactic type, we avoided CTPs such as *učiti*, which can mean both ‘learn’ and ‘teach’. Whereas in the former meaning it is a simple subject control predicate, in the latter it is an object control predicate.

In general, raising verbs are much more frequent than object control verbs. Therefore, we avoided the most frequent raising verbs and included two of their subtypes: modal and phasal. This resulted in 15 lemmata, 5 per each syntactic type.

However, the number of object control predicates that met our query conditions in InterCorp was small and in comparison to raising and subject control verbs their frequencies were lower. The latter might be due to the fact that the population of object control lemmata is bigger than the one of raising lemmata, so each token appears with lower frequency. To increase the recall for object control verbs from srWaC we added two object control verbs which are not present in InterCorp, but are quite frequent in srWaC. The list of CTPs which emerged as a result of this procedure and which was used to collect data on CC out of  $da_2$ -complements for Chapter 13 is given in Table 12.2.<sup>16</sup> Basing on the results of the four CQL queries for the variants from Table 12.1 in srWaC, we calculated the estimated frequencies of CTPs in srWaC.<sup>17</sup>

<sup>14</sup>For more information on the difference between  $da_1$ - and  $da_2$ -complements see Section 2.5.3.

<sup>15</sup>For more information on haplology and pseudodiaclisis see Sections 2.4.2.2 and 2.4.5.

<sup>16</sup>Since Serbian orthography allows for both the ekavian (*smeti*) and the ijekavian (*smjeti* ‘be allowed’) pronunciation, we took this into account when querying srWaC.

<sup>17</sup>For more information on calculating estimated frequencies in that study see Section 13.3.



Table 12.2: CTPs selected for study of CC out of *da*<sub>2</sub>-complements

No.	Verb	Translation	Frequency of <i>da</i> <sub>2</sub>		Syntactic type
			InterCorp	srWaC <sup>a</sup>	
1.	<i>moći</i>	‘can’	69	37526	
2.	<i>nastaviti</i>	‘continue’	2	1028	
3.	<i>početi</i>	‘start’	26	6546	raising
4.	<i>prestati</i>	‘stop’	5	1203	
5.	<i>sm(j)eti</i>	‘be allowed’	7	2027	
6.	<i>nam(j)eravati</i>	‘intend’	3	465	
7.	<i>nastojati</i>	‘strive’	7	721	simple
8.	<i>pokušati</i>	‘try’	7	4794	subject
9.	<i>um(j)eti</i>	‘be able to’	8	1209	control
10.	<i>usp(j)eti</i>	‘succeed’	9	4331	
11.	<i>dozvoliti</i>	‘allow’	7	2528	
12.	<i>narediti</i>	‘order’	5	1174	object
13.	<i>nat(j)erati</i>	‘force’	2	502	control
14.	<i>zamoliti</i>	‘ask’	3	1584	with dative
15.	<i>pustiti</i>	‘let’	3	534	and accusative
16.	<i>primorati</i>	‘compel’	0	248	controllers
17.	<i>pomoći</i>	‘help’	0	331	

<sup>a</sup>(estimated)

To prepare the study of infinitive complements presented in Chapter 14 we used frequency lists from hrWaC. Since Hansen, Kolaković & Jurkiewicz-Rohrbacher (2018: 266) suggested that the reflexivity of the CTP might be a constraint on CC out of stacked infinitives, we decided to include a group of reflexive subject control CTPs in that study. The same CTPs, presented in Table 12.3, were also used in our psycholinguistic experiment (see Chapter 15) for easier comparison of the results of the two studies.

The list of all verbs which have an infinitive as a complement was extracted from hrWaC v2.2 in three steps. First, we applied the query [tag="Vm.\*"], which let us find all examples with a main verb in all their possible forms. After that, the Filter function was applied. We used the query [tag="V.n"] as a positive filter, which allowed us to extract only those examples with the main verbs which have

12 Introductory remarks to corpus studies on clitic climbing

Table 12.3: CTPs selected for study of CC out of infinitive complements. Lemma frequency is taken from hrWaC v2.2 and expressed per million.

No.	Verb	Translation	Frequency/mil. words	Syntactic type
1.	<i>moći</i>	‘can’	4056.5	
2.	<i>trebati</i>	‘have to’	1761.4	
3.	<i>morati</i>	‘must’	1224.8	
4.	<i>smjeti</i>	‘be allowed’	208.0	
5.	<i>počinjati</i>	‘start’	125.0	raising
6.	<i>kretati</i>	‘go/start’	101.2	
7.	<i>nastavljati</i>	‘continue’	77.7	
8.	<i>prestajati</i>	‘stop’	24.9	
9.	<i>znati</i>	‘know/can’	1584.6	
10.	<i>željeti</i>	‘want/will’	859.1	
11.	<i>pokušavati</i>	‘try’	139.6	simple
12.	<i>planirati</i>	‘plan’	105.7	subject
13.	<i>nastojati</i>	‘strive’	76.2	control
14.	<i>odlučivati</i>	‘decide’	54.9	
15.	<i>odbijati</i>	‘refuse’	32.6	
16.	<i>uspijevati</i>	‘succeed’	31.9	
17.	<i>bojati se</i>	‘be afraid’	106.0	
18.	<i>sjetiti se</i>	‘remember’	79.1	
19.	<i>truditi se</i>	‘try’	53.3	reflexive
20.	<i>sramiti se</i>	‘be ashamed’	12.4	subject
21.	<i>usuđivati se</i>	‘dare’	5.6	control
22.	<i>stidjeti se</i>	‘be ashamed’	4.2	
23.	<i>libiti se</i>	‘hesitate’	3.0	
24.	<i>ustručavati se</i>	‘hesitate’	2.5	

an infinitive to their right (within 1 to 5 positions). In the next step we applied the Frequency function in NoSketchEngine and sorted the verbs according to their lemma form. The list obtained was downloaded as a .txt file, and opened for further editing in Excel. Next to each lemma we noted the predicate type, with three major groups: raising, subject control and object control. In the next step we classified subject control predicates into two further groups: simple subject control predicates such as *planirati* ‘plan’ and reflexive subject control predicates

such as *bojati se* 'be afraid', and formed separate lists of predicates according to this classification.<sup>18</sup>

## 12.5 Data collection

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 retrieved examples would have exceeded our human capacities, we decided to work with random samples of maximally 100 examples per structure variant per CTP.<sup>19</sup> Maximally one hit from one web page or text was taken for the sample. To achieve this, we first applied the 1<sup>st</sup> hit in doc function of NoSketch-Engine and then its Sample function. The samples were downloaded as .txt files and revised manually in Excel. The clean and manually revised data were then used in the analyses described in the next two chapters.

---

<sup>18</sup>The same procedure applies also to object control predicates which were later subclassified to object control predicates into four groups according to their controllers: object control predicates with pronominal controllers in dative and accusative and object control predicates with REFL<sub>2nd</sub> controllers *se* and *si*. The predicates from the four object control groups are not listed in this chapter because they were not used in the study presented in Chapter 14. They can be found in Section 15.3.1.

<sup>19</sup>In the case of less frequent simple and reflexive subject control predicates when entered query resulted in less than 100 examples all retrieved examples were checked manually.



# 13 A corpus-based study on clitic climbing out of *da*<sub>2</sub>-construction and the raising–control distinction (Serbian)

## 13.1 Introduction

In this chapter we address CC out of *da*-complements.<sup>1</sup> In many languages, CC is only attested in clauses with infinitive complements; cross-linguistically, CC out of complements with inflected verbs is a rare phenomenon.<sup>2</sup>

In Serbian, infinitive complements compete with the so-called *da*-complement, that is, a verb marked for person and number which is introduced by an element usually treated as a complementiser.<sup>3</sup> As an illustration, compare the sentence presented in (1) with the infinitive complement *naći* ‘find’, and the sentence with the *da*-complement *nađete* ‘(you) find’ in (2).

- (1) Ako ne možete naći udoban položaj [...].  
if NEG can.2PRS find.INF comfortable position  
‘If you cannot find a comfortable position [...].’ [srWaC v1.2]
- (2) [...] na celoj toj teritoriji ne možete da nađete 500 stanovnika.  
on whole that territory NEG can.2PRS that find.2PRS 500 inhabitants  
‘[...] on that whole territory you cannot find 500 inhabitants.’ [srWaC v1.2]

However, it is rather unclear to what extent and under what circumstances CC out of *da*-complements is possible. The present chapter approaches this problem empirically using corpus-based methods. Section 13.2 refers to the discussion on

---

<sup>1</sup>Some results from this chapter were previously discussed in Jurkiewicz-Rohrbacher, Kolaković & Hansen (2017).

<sup>2</sup>A comparison of CC out of complements with inflected verbs in Czech and BCS can be found in Section 11.2.2.

<sup>3</sup>See Section 2.5.3 for basic information on complement types in BCS.

CC out of *da*-complements in Serbian. Section 13.3 describes the results of the corpus queries in detail, while in Section 13.4 we analyse and discuss them. The final Section 13.5 draws conclusions from the main results and offers a suggestion for further research.

## 13.2 The *da*-complement and CC in Serbian

As discussed in Section 1.2, the research on the syntax of BCS is divided into descriptive empirical studies on the one hand, and works with a formal theoretical orientation on the other. Therefore it comes as no surprise that in the literature we find largely contradictory statements concerning CC out of *da*-complements. Stjepanović (2004: 174ff) argues that *da*-complements and infinitives allow CC in a similar way. Nevertheless, discussing examples of CC out of *da*-complements, she writes imprecisely that those “are acceptable sentences, however, they are short of perfect” (cf. Stjepanović 2004: 201). Similarly, according to Franks & King (2000: 243), movement out of the finite complement is only “marginally possible”. On the opposite side of the spectrum, Ćavar & Wilder (1994: 41) and Browne (2003: 41) argue that CC out of finite complements is completely impossible. Moreover, Ćavar & Wilder (1994: 448) claim that CC is not possible even out of semi-finite complements of subject control verbs like *ht(j)eti* ‘will’.<sup>4</sup> Finally, Progovac (2005: 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.<sup>5</sup>

However, as explained in Section 2.5.3 we have to bear in mind that *da*-complements do not behave in a uniform way since they differ with regard to tense marking. Based on Todorović (2012), we assume that if CC is possible, this is only the case for *da*<sub>2</sub>-complements, which are marked only for person and number. One hypothetical reason why some scholars reject the possibility of CC out of *da*<sub>2</sub>-complements is its extreme rarity in comparison to equivalent constructions without CC. An early empirical work concerning CC is Marković (1955), which assumes that the variation in CL positioning is closely related to the (at that time) new and growing tendency to replace infinitives with *da*<sub>2</sub>-complements.<sup>6</sup> Marković addresses the question of diatopic and diaphasic variation with respect

---

<sup>4</sup>Ćavar & Wilder (1994: 448) address causative constructions as the only exception to that rule.

<sup>5</sup>For more information on the traps of studies based on self-constructed examples and intuition, see Section 3.1.

<sup>6</sup>Marković (1955) does not use the term clitic climbing.

to CC out of such constructions.<sup>7</sup> As to the former dimension he claims that ekavian Serbian speakers, who (at that time) had already almost completely replaced the infinitive with *da*<sub>2</sub>-complements, preferred keeping the pronominal CL directly after the *da* particle, i.e. no CC.<sup>8</sup> Regarding the two types of variation, he stated that at that time, CC was common in journalistic ijekavian texts published in Sarajevo (cf. Marković 1955: 35–40).

Very few papers recognise the importance of the raising–control dichotomy for CC in BCS. Aljović (2005) observes that in BCS, CC is only possible out of complements whose subject is empty and coreferential with the matrix subject, although in a footnote she acknowledges that CC is also possible when the subject of the embedded complement is coreferential with the matrix indirect object in the dative, i.e. out of object-controlled infinitives.<sup>9</sup> For Czech a range of various constraints on CC closely connected with object control was described in the theoretical literature (Rezac 2005, Dotlačil 2004, Hana 2007).

As explained in Section 10.1, in this study we focus exclusively on Serbian, and not on Bosnian or Croatian, because *da*-complements are much more frequently used in Serbian than in Bosnian and Croatian, especially in the context of modal verbs in non-epistemic meanings as in example (2) above.

Therefore, in this chapter we address the following research question:

Q1: To what extent is CC out of *da*<sub>2</sub>-complements possible in Serbian?

If CC out of *da*<sub>2</sub>-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 for categorising different types of sentences with verbal complements.<sup>10</sup> The point of departure for the present study is divergent statements on the link between CC and the raising–control dichotomy in Czech.<sup>11</sup>

Hansen, Kolaković & Jurkiewicz-Rohrbacher (2018) demonstrate that CC out of stacked infinitives, that is, multiply embedded infinitives, is not obligatory in BCS. However, empirical data for CC out of *da*-complements are still lacking. Based on this, we formulate the second research question:

<sup>7</sup>More information on diaphasic and diatopic variation can be found in Section 2.3.

<sup>8</sup>In BCS languages speakers of ekavian, ikavian and ijekavian dialects may be distinguished. More information on this can be found in Section 7.2.

<sup>9</sup>Bear in mind that complements whose subject is empty and coreferential with the matrix subject are complements of raising and subject control CTPs.

<sup>10</sup>For more information on the distinction between control and raising predicates see Section 2.5.2.

<sup>11</sup>For more information on this topic see Section 11.3.

Q2: Does CC out of  $da_2$ -complements in Serbian depend on verb type with respect to the raising–control distinction?

We approach the questions stated above by investigating 17 matrix verbs whose choice is explained in Section 12.4. The data come from srWaC, which is due to its size the most reliable source for tracking rare phenomena in BCS, such as CC out of  $da_2$ -complements in Serbian (Jurkiewicz-Rohrbacher, Kolaković & Hansen 2017).<sup>12</sup> Accordingly, although Marković (1955) argued that CC out of  $da_2$ -complements is more frequent on Bosnian than on Serbian language territory, we decided to conduct the study on Serbian material and extract data from srWaC since it is almost twice as big as bsWaC.

### 13.3 Results

We present the results of corpus queries in detail in Tables 13.1–13.3. The results still posed problems due to the number of retrieved queries and their precision. Since we noticed that not all retrieved sentences correspond to the target structures, we decided to conduct a manual check.<sup>13</sup>

As no gold standards have been broadly acknowledged, we decided to follow some suggestions by Wallis (2014), and accordingly we estimated the precision of queries through sampling. From all sentences retrieved, with the Sample function in NoSketch Engine we took random samples of 100 sentences and checked all of them manually. The number of correct target structures can be seen in Tables 13.1–13.3. A sample of this size should usually give no more than a 10% margin of error at a confidence level of 95% regardless of the population size. We calculated the binomial probability confidence interval (“conf. interval” column in Tables 13.1–13.3) using the Clopper-Pearson exact method. On the basis of the worst-case scenario for the obtained confidence intervals, we recalculated raw frequencies (“retrieved sentences CQL” column in Tables 13.1–13.3) into estimated frequencies (“estimated frequency” column in Tables 13.1–13.3). The relative frequency of CC out of  $da_2$ -complements in these tables refers to the proportion of the estimated frequency of CC out of  $da_2$ -complements to the estimated frequency of all  $da_2$ -complements for the given CTP. These are analysed in the next section.

---

<sup>12</sup>For more information on the corpora selected and our argumentation for choosing those and not other corpora, see Section 4.6.3. See Section 12.2 for the queries used and Section 3.3 for an exhaustive discussion of our methodological approach.

<sup>13</sup>The target structures can be found in Table 12.1.



Table 13.1: Position of CL with respect to *da*-complementiser in sentences with raising predicates

Construction	Retrieved sentences CQL	Correct target structures in sample	Estimated frequency	Conf. interval
<i>moći</i> ‘can’				
CTP da PRS CL	2527	0	0	0
CTP da CL PRS	44691	91	37363	0.83602
CTP CL da PRS	1843	2	5	0.00243
CL CTP da PRS	9637	5	158	0.01643
relative frequency of CC out of $da_2$			0.0043	
<i>nastaviti</i> ‘continue’				
CTP da PRS CL	83	1	1	0
CTP da CL PRS	1120	97	1025	0.91482
CTP CL da PRS	24	1	1	0
CL CTP da PRS	255	2	2	0.00243
relative frequency of CC out of $da_2$			0.0029	
<i>početi</i> ‘start’				
CTP da PRS CL	427	1	1	0.00025
CTP da CL PRS	6724	100	6480	0.96378
CTP CL da PRS	80	5	5	0
CL CTP da PRS	956	12	61	0.06357
relative frequency of CC out of $da_2$			0.01	
<i>prestati</i> ‘stop’				
CTP da PRS CL	82	2	2	0
CTP da CL PRS	1298	97	1187	0.91482
CTP CL da PRS	13	2	2	0
CL CTP da PRS	213	11	12	0.05621
relative frequency of CC out of $da_2$			0.0116	
<i>sm(j)eti</i> ‘be allowed’				
CTP da PRS CL	86	1	1	0
CTP da CL PRS	2134	99	2018	0.94554
CTP CL da PRS	56	0	0	0
CL CTP da PRS	338	6	8	0.02233
relative frequency of CC out of $da_2$			0.01	

13 A corpus-based study on clitic climbing out of  $da_2$ -construction

Table 13.2: Position of CL with respect to  $da$ -complementiser in sentences with subject control predicates.

Construction	Retrieved sentences CQL	Correct target structures in sample	Estimated frequency	Conf. interval
<i>nam(j)eravati</i> 'intend'				
CTP da PRS CL	12	1	1	0
CTP da CL PRS	506	97	463	0.91482
CTP CL da PRS	1	0	0	0
CL CTP da PRS	34	1	1	0
relative frequency of CC out of $da_2$			0.0021	
<i>nastojati</i> 'strive'				
CTP da CL PRS	23	1	1	0
CTP da CL PRS	797	96	718	0.90074
CTP CL da PRS	0	0	0	0
CL CTP da PRS	55	2	2	0
relative frequency of CC out of $da_2$			0.0027	
<i>pokušati</i> 'try'				
CTP da PRS CL	126	2	2	0.00243
CTP da CL PRS	5044	99	4769	0.94554
CTP CL da PRS	33	4	4	0
CL CTP da PRS	296	12	19	0.06357
relative frequency of CC out of $da_2$			0.0047	
<i>um(j)eti</i> 'be able to'				
CTP da PRS CL	59	0	0	0
CTP da CL PRS	1277	99	1207	0.94554
CTP CL da PRS	53	0	0	0
CL CTP da PRS	381	2	2	0.00243
relative frequency of CC out of $da_2$			0.0016	
<i>usp(j)eti</i> 'succeed'				
CTP da PRS CL	173	0	0	0
CTP da CL PRS	4655	98	4327	0.92962
CTP CL da PRS	74	0	0	0
CL CTP da PRS	673	2	2	0.00243
relative frequency of CC out of $da_2$			0.009	

Table 13.3: Position of CL with respect to *da*-complementiser in sentences with object control predicates

Construction	Retrieved sentences CQL	Correct target structures in sample	Estimated frequency	Conf. interval
<i>dozvoliti</i> 'allow'				
CTP da PRS CL	125	1	1	0.00025
CTP da CL PRS	2805	96	2527	0.90074
CTP CL da PRS	1932	0	0	0
CL CTP da PRS	3142	0	0	0
relative frequency of CC out of $da_2$			0	
<i>narediti</i> 'order'				
CTP da PRS CL	22	1	1	0
CTP da CL PRS	1242	99	1174	0.94554
CTP CL da PRS	721	0	0	0
CL CTP da PRS	814	0	0	0
relative frequency of CC out of $da_2$		0		
<i>nat(j)erati</i> 'force'				
CTP da PRS CL	39	0	0	0
CTP da CL PRS	558	96	503	0.90074
CTP CL da PRS	64	0	0	0
CL CTP da PRS	3930	0	0	0
relative frequency of CC out of $da_2$			0	
<i>primorati</i> 'force'				
CTP da PRS CL	10	0	0	0
CTP da CL PRS	257	100	248	0.96378
CTP CL da PRS	209	0	0	0
CL CTP da PRS	474	0	0	0
relative frequency of CC out of $da_2$			0	
<i>pomoći</i> 'help'				
CTP da PRS CL	68	0	0	0
CTP da CL PRS	874	48	331	0.37901
CTP CL da PRS	1963	0	0	0
CL CTP da PRS	7528	0	0	0
relative frequency of CC out of $da_2$			0	
<i>pustiti</i> 'let'				
CTP da PRS CL	76	2	2	0
CTP da CL PRS	685	86	532	0.77627
CTP CL da PRS	2087	0	0	0
CL CTP da PRS	2271	0	0	0
relative frequency of CC out of $da_2$			0	
<i>zamoliti</i> 'ask'				
CTP da PRS CL	24	1	1	0
CTP da CL PRS	1784	95	1583	0.88717
CTP CL da PRS	1380	0	0	0
CL CTP da PRS	2116	0	0	0
relative frequency of CC out of $da_2$			0	

### 13.4 Discussion: Constraints on CC from $da_2$ -complements

Although the following discussion is based on a worst-case scenario, our material provides empirical evidence that CC out of  $da_2$ -complements into matrix clauses is indeed possible, but it is most likely a marginal phenomenon.<sup>14</sup>

Our samples yielded 69 correct sentences with CC originating from 42 different top-level domains. From that we estimated a worst-case scenario of 286 CC cases in the whole examined population in srWaC. The frequencies of CC normalised to the frequency of a  $da_2$ -complement for a particular verb are presented as part of Tables 13.1–13.3 and in Figure 13.1. Analysis of the frequencies shows that CC out of  $da_2$ -complements occurs with verbs of different frequencies. The Chi-square test of dependence between syntactic type and CC yields a significant result ( $p < 0.001$ ), so the null-hypothesis that there is no relation between CC and the type of CTP can be rejected.

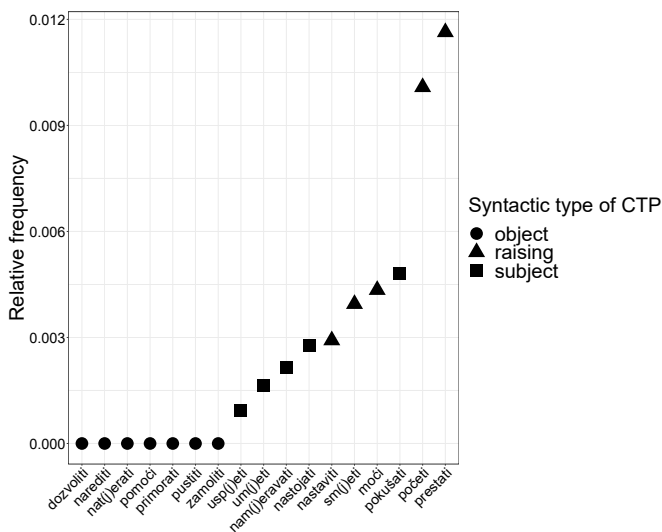


Figure 13.1: Relative frequencies of CC for the retrieved CTPs

Figure 13.1 shows that the two phasal verbs *prestati* ‘stop’ and *početi* ‘start’ have the highest relative frequency of CC out of  $da_2$ -complements, followed by the subject control predicate *pokušati* ‘try’, and raising verbs *moći* ‘can’, *sm(j)eti*

<sup>14</sup>CC out of  $da_2$ -complements into matrix clauses is also attested in dialects: for more information and examples see Section 7.7.

‘be allowed’ and *nastaviti* ‘continue’. An interesting finding is that object control CTPs with both dative and accusative controllers are highly unlikely to allow CC. We did not find a single example for the predicates we selected.

Although the probability that CC out of a *da*<sub>2</sub>-complement will occur is generally low, we can conclude that it is additionally influenced by the syntactic type of the CTP. Thus, it is lower for subject control verbs than for raising predicates, and not retrievable from corpus data for object control predicates.

As explained in Chapter 12, in the case of CC out of *da*<sub>2</sub>-complements we distinguish four different CL positions. In Table 12.1, it may be seen that as orientation points for the CL positions we use the particle *da* and matrix predicates. Tables 13.1–13.3 show that sentences in which the CL is placed to the right of the verb of the *da*<sub>2</sub>-complement are extremely rare (3)–(5), albeit possible for all three investigated types of CTPs (pace Browne 2003: 41, Ćavar & Wilder 1994: 41).<sup>15</sup> In example (3) with the raising matrix predicate *pretao* ‘(I) stopped’ the pronominal CL *te* ‘you’ is placed to the right of its governing semi-finite verb *volim* ‘(I) love’. The same CL positioning can be observed in examples with subject (4) and object (5) control matrix predicates.

- (3) [...] ali nikada nisam<sub>1</sub> pretao<sub>1</sub> da volim<sub>2</sub> te<sub>2</sub> [...].  
 but never NEG.be.1SG stop.PTCP.SG.M that love.1PRS you.ACC  
 ‘[...] but I never stopped loving you [...].’ [srWaC v1.2]
- (4) Mihajlo htede<sub>1</sub> da pokuša<sub>2</sub> da urazumi<sub>3</sub> ga<sub>3</sub> [...].  
 Mihajlo want.3AOR that try.3PRS that bring.to.reason.3PRS him.ACC  
 ‘Mihajlo wanted to bring him to reason [...].’ [srWaC v1.2]
- (5) [...] i dozvoli<sub>1</sub> da osetim<sub>2</sub> te<sub>2</sub> [...].  
 and allow.2IMP that feel.1PRS you.ACC  
 ‘[...] and let me feel you [...].’ [srWaC v1.2]

It is also very clear that regardless of the CTP type, CLs tend to be placed directly after the *da* particle, as in the examples with raising (6), subject (7) and object control (8) CTPs below. This is the CL position which some scholars (e.g. Browne 2003: 41, Ćavar & Wilder 1994: 41) assumed to be the only possible and correct one.<sup>16</sup>

<sup>15</sup>For basic information on CL placement after complementisers in BCS standard varieties see Section 6.5.3.

<sup>16</sup>In varieties which are not under direct influence of prescriptive norms, i.e. dialects and spoken varieties, CLs do not always follow the *da* particle. For more information and examples, see Section 7.6.2.

13 A corpus-based study on clitic climbing out of *da*<sub>2</sub>-construction

- (6) Možete da *ga*<sub>2</sub> podignete<sub>2</sub> [...].  
 can.2PRS that him.ACC lift.2PRS  
 ‘You can lift it [...].’ [srWaC v1.2]
- (7) Nastojim da *ih*<sub>2</sub> razumem<sub>1</sub> [...].  
 try.1PRS that them.ACC understand.1PRS  
 ‘I try to understand them [...].’ [srWaC v1.2]
- (8) Ova okolnost pomogla<sub>1</sub> je<sub>1</sub> Pavlu da *ga*<sub>2</sub> uspešno  
 this circumstance help.PTCP.SG.F be.3SG Pavle that him.ACC successfully  
 prati<sub>2</sub> [...].  
 follow.3PRS  
 ‘This circumstance helped Pavle to follow him successfully [...].’  
 [srWaC v1.2]

Furthermore, in the case of CC, CLs tend to be placed left of the matrix verb as in (9). However, they can appear between the CTP and the *da* particle as well, as in (10). Both examples contain the raising CTP form *mogu* ‘(I) can’.

- (9) [...] i zato *te*<sub>2</sub> ne mogu<sub>1</sub> da primim<sub>2</sub>.  
 and therefore you.ACC NEG can.1PRS that welcome.1PRS  
 ‘[...] and therefore I cannot welcome you.’ [srWaC v1.2]
- (10) [...] ne mogu<sub>1</sub> *ih*<sub>2</sub> da napišem<sub>2</sub> [...].  
 NEG can.1PRS them.ACC that write.1PRS  
 ‘[...] I cannot write them.’ [srWaC v1.2]

If auxiliaries belonging to the CTP appear, the climbing CLs can form mixed clusters with them.<sup>17</sup> In example (11), the pronominal CL *im* ‘them’ climbed out of the *da*<sub>2</sub>-complement ‘(he) speaks’ and formed a mixed cluster with the auxiliary CL *je* ‘is’ which was present in the matrix clause. We observe a similar situation in example (13) where the matrix auxiliary CL *je* formed a mixed cluster with the pronominal CL *mi* ‘me’ which climbed out of the *da*<sub>2</sub>-complement. These examples allow us to reject Todorović (2012: 166) 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”.

- (11) [...] počeo<sub>1</sub> *im*<sub>2</sub> je<sub>1</sub> da govori<sub>2</sub> o dolasku ove  
 start.PTCP.SG.M them.DAT be.3SG that speak.3PRS about arrival this  
 grupe.  
 group  
 ‘[...] he began to speak to them about the arrival of this group.’  
 [srWaC v1.2]

<sup>17</sup>For more information and examples of simple and mixed CL clusters, see Section 2.4.2.1.

A reflexive CL *se* can either climb with the pronominal CL, as in (12), or it can stay in the *da*<sub>2</sub>-complement, as in (13).

- (12) U poslednje vreme *mi*<sub>2</sub> *se*<sub>2</sub> pocelo<sub>1</sub> da desava<sub>2</sub> da  
 in past time me.DAT REFL start.PTCP.SG.N that happen.3PRS that  
 cujem<sub>3</sub> [...].  
 hear.1PRS  
 ‘Recently, it has started happening to me that I hear [...].’ [srWaC v1.2]
- (13) [...] i pocelo<sub>1</sub> *mi*<sub>2</sub> *je*<sub>1</sub> da *se*<sub>2</sub> vrti<sub>2</sub> u glavi.  
 and start.PTCP.SG.N me.DAT be.3SG that REFL spin.3PRS in head  
 ‘[...] and I started to feel dizzy.’ [srWaC v1.2]

The fact that two CLs that were generated by the same verb do not have to climb together out of *da*<sub>2</sub> complement was observed already by Stjepanović (2004: 182). Her examples, however, concern only two pronominal CLs and not the reflexive CL *se* in combination with a pronominal CL. Stjepanović (2004: 182) concludes that in the case of pseudodiacclisis only a dative CL climbs, while an accusative CL stays in the *da*<sub>2</sub>-complement. We additionally argue that if two CLs are generated in the *da*<sub>2</sub>-complement and occur in pseudodiacclisis, it is the pronominal that climbs, while the reflexive tends to stay in the *da*<sub>2</sub>-complement, like in example (13). In addition, since in the latter example the two CLs do not climb together, we can conclude that in Serbian there is no all-or-nothing constraint on CC out of *da*<sub>2</sub>-complements (pace Rezac 2005: 8).

Moreover, it is worth mentioning that when the reflexive CL *se* climbs with a pronominal CL in the matrix clause, the auxiliary CL *je* ‘is’ from the matrix clause is omitted. In other words, haplology of unlikes occurs.<sup>18</sup> Since we did not find examples with three CLs (auxiliary, pronominal and reflexive) in a cluster, we may speculate that whenever there are three CLs in a sentence, the reflexive tends to stay in the *da*<sub>2</sub>-complement.

Finally, it is worth mentioning that CC has not been attested for the third person accusative/genitive singular feminine CL *je*. This needs further investigation, but could be due to error in tagging, i.e., if the pronominal CL *je* was tagged as the verbal CL *je* ‘is’.

<sup>18</sup>See Section 2.4.2.2 for basic information and examples of haplology of unlikes. It is claimed that haplology of unlikes is obligatory in the standard Serbian variety – see Section 6.4.2.2. However, in dialects spoken on Serbian territory, haplology of unlikes is not obligatory – for more information and examples see Section 7.5.1.

## 13.5 Conclusions

In this chapter we addressed the syntactic mechanism of CC in the context of  $da_2$ -complements. These complements are characterised by the presence of a verb inflected for person and number. This is an interesting topic because it is claimed for Czech, for instance, that finite complements block CC. The point of departure of our study was the observation that there is large disagreement as to the acceptability of CC out of  $da_2$ -complements. Whereas Stjepanović (2004) allows the grammaticality of CC out of  $da_2$ -complements mainly within a unified formal theory of CC in BCS, most other authors reject the grammaticality of this structure outright. Our data allow the following answers to be given to our research questions from Section 14.3:

- A1: Serbian  $da_2$ -complements do marginally allow CC. In these cases, the CL that climbed can form a mixed cluster with the auxiliary CL of the matrix verb. We thus in principle agree with Stjepanović (2004), but have to point out that we are dealing with a highly marginal construction. Examples supporting the occurrence of CC for all CL forms were not found. In contrast, CC out of infinitive complements in Croatian (see next chapter) and even out of stacked infinitives in Bosnian, Croatian, and Serbian (Hansen, Kolaković & Jurkiewicz-Rohrbacher 2018) is not a marginal phenomenon at all.
- A2: 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. Moreover, the results of our corpus study on CC out of  $da_2$ -complements in Serbian are in line with our psycholinguistic study on CC out of infinitive complements in Croatian presented in Chapter 15 and with the corpus study on CC out of stacked infinitives in Bosnian, Croatian, and Serbian by Hansen, Kolaković & Jurkiewicz-Rohrbacher (2018).

In addition we would like to comment on some further evidence for the following two constraints. First, it seems that the reflexive CL *se* does not climb out of the  $da_2$ -complement if there is an auxiliary CL in the matrix clause. Second, if the  $da_2$ -complement is reflexive and governs the pronominal CL and if those CLs appear in pseudodiacyclisis, it is the pronominal one that climbs and the reflexive that stays in the complement. First, this suggests that the pronominal CL and reflexive *se* behave differently, which leads to the conclusion that CC is not



a unified syntactic mechanism. Second, examples of pseudodiaclisis in the context of CC out of the  $da_2$ -complement indicate that CC is not an all-or-nothing phenomenon, which is in line with Stjepanović's (2004: 182) observations (pace Rezac 2005: 8). Finally, we were able to reject Todorović's (2012) hypothesis that past tense or future auxiliaries block CC.



# 14 A corpus-based study on clitic climbing out of infinitive complements in relation to the raising–control dichotomy and diaphasic variation (Croatian)

## 14.1 Introduction

The present chapter is a further empirical study on CC out of infinitive complements with a specific focus on the raising–control distinction, which we showed to be a relevant factor for CC out of *da*<sub>2</sub>-constructions in Serbian in Chapter 13.<sup>1,2,3</sup> Here, we broaden the empirical base for the investigation of this dichotomy by specifically examining Croatian infinitive constructions. In addition, we zoom in on the diaphasic dimension of variation as a factor influencing the probability of CC occurring. This type of variation, as explained in Section 2.3, reflects different modes of language use in different situations. To illustrate, the examples provided below contain the same CTP *morati* ‘must’ and infinitive *odlučiti se* ‘decide’. However, whereas in example (2) extracted from the corpus of the standard Croatian variety the REFL<sub>LEX</sub> CL *se* climbs out of the infinitive complement, in example (1) extracted from the Forum subcorpus of the Croatian web corpus the very same CL stays in situ.<sup>4</sup> In this chapter we thus examine whether these differences in CL positioning are due to chance or whether they can be ascribed to diaphasic variation.

- (1) Pa ako već morate<sub>1</sub> odlučiti<sub>2</sub> se za jaslice [...].  
well if already must.2PRS decide.INF REFL for nursery  
‘Well if you have to opt for nursery [...].’ [hrWaC v2.2]

---

<sup>1</sup>Some results from this chapter have been previously discussed in Kolaković, Jurkiewicz-Rohrbacher & Hansen (2019).

<sup>2</sup>For basic information on the raising–control dichotomy see Section 2.5.2.

<sup>3</sup>See also Jurkiewicz-Rohrbacher, Hansen & Kolaković (2017).

<sup>4</sup>For more information on our typology of reflexives see Section 2.5.4.2.

- (2) A da se morate<sub>1</sub> odlučiti<sub>2</sub>?  
 and if REFL MUST.2PRS decide.INF  
 ‘And if you have to decide?’ [Riznica]

The rest of this chapter is structured as follows: Section 14.2 describes the importance of diaphasic variation for CC in Spanish and Portuguese. Spanish and Portuguese are of interest because their CL systems have many features in common with Croatian and show diaphasic variation. Next we present basic information on diaphasic variation in Croatian. Our research questions are presented in Section 14.3. The choice of data and the collection process are explained in Section 14.4, while Section 14.5 describes the results in detail. It is followed by the final Section 14.6, which draws conclusions.

## 14.2 Clitic climbing and diaphasic variation

### 14.2.1 Clitic climbing and diaphasic variation in Romance languages

Although in Chapter 10 we avoid comparison between BCS CLs and those in Romance languages, we do make it here. As the relationship between CC and diaphasic variation has never been the topic of any study on a Slavonic language, it is worth looking at variationist work on Romance languages. All the more so as Spanish and Portuguese are languages with CLs which can climb. In the literature on variation in Spanish CC, several authors (e.g. Davies 1995, Cacoulios 1999) point out the relevance of register: generally it can be said that CC is less frequent in Spanish written texts than in Spanish spoken texts.<sup>5</sup>

Davies (1995) investigated CC on the basis of a corpus composed of texts from ten Spanish-speaking countries. He reported a consistent difference between registers. His data show that the distance between registers with respect to CC can be as high as 30% (cf. Davies 1995: 373f). Cacoulios (1999) studied CC in Mexican Spanish using similar methodology. Register once again turned out to be an important factor: sociolinguistic interviews had higher rates of CC than essays (89% versus 68%) (cf. Cacoulios 1999).

de Andrade (2010) replicated the results of studies on Spanish CC for European Portuguese data. He analysed CC in 1000 Portuguese sentences, which were annotated as formal (newspaper interviews and novels) or informal (sociolinguistic interviews). Using basic statistical correlation testing de Andrade (2010: 99) showed that the CC rates in those two registers differ significantly.

<sup>5</sup>A deeper look at those papers reveals that authors who worked on the impact of register on CC in Spanish and Portuguese use the concept of register to refer to different things. In its broadest sense, register is a language variety defined by the context of usage (Čolak 2015: 31).

It is worth mentioning that he also analysed, but only on the data from the formal register, how language-internal factors such as CL type and grammatical function, syntactic context, the presence of intervening elements between CTP and infinitive, and the frequency of the CTP influence CC. His results on the importance of CL type and grammatical status are in accordance with claims concerning those factors made in the theoretical literature on CC in Czech.<sup>6</sup> Like in Czech, in European Portuguese CL type and grammatical function are important factors for CC. Specifically, in Portuguese CC is much more frequent in the case of datives than in the case of accusatives. While the CC rate is 51.6% for ethic and possessive datives and 50.7% for argumental datives, for accusatives it is only 32.6% (cf. de Andrade 2010: 101).

#### 14.2.2 Diaphasic variation in Croatian

Although we are aware of the differences in the stratification of the languages mentioned above, we treat 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. We only refer to Frančić et al. (2006: 10–17), who distinguish the following diatopic strata in Croatian: local idioms or Croatian dialects; urban idioms (substandard idioms or jargon) and the Croatian standard language. The latter is an abstract system based on three dialects – not only Štokavian but also Čakavian and Kajkavian (cf. Frančić et al. 2006: 22f). Furthermore, the literature (e.g. Frančić et al. 2006: 230) acknowledges the following diaphasic strata in Croatian: scientific (scholarly), administrative, journalistic, literary and colloquial.

As we can see, on the one hand we have standard Croatian and on the other, non-standard Croatian conventionally labelled as spoken, colloquial, dialectal, rural, etc. (cf. Murelli 2011: 32f). The latter variety comprises various idioms with elements which are not codified or are rejected in the standard.

However, in this particular study of CC, we are interested only in the standard Croatian variety and in the non-standard variety termed “everyday colloquial language”, “conversational standard” or “informal spoken standard”. Because everyday colloquial Croatian as a non-standard idiom is in fact a sub-variety of standard language with elements which are not a part of the norm (cf. van Marle 1997: 13–17, Langston & Peti-Stantić 2014: 30), this non-standard variety shares more similarities with the Croatian standard than Croatian local idioms (i.e. dialects) do. In the remainder of our paper, the term *COLLOQUIAL CROATIAN* (variety) will be used to refer to this particular non-standard variety of Croatian.

---

<sup>6</sup>For more information see Sections 11.3.3 and 11.3.5.

### 14.3 Research questions

Based on the considerations presented in Sections 14.2 and 11.3, we explore the claim that CC varies with respect to both the raising–control dichotomy and register. As already mentioned in Section 12.4 we expand the typology of CTPs to include reflexive subject control verbs, as suggested by Hansen, Kolaković & Jurkiewicz-Rohrbacher (2018: 266), and address the following research questions:

- Q1: Does clitic climbing out of single infinitive complements in Croatian depend on CTP type with respect to the raising–control distinction?
- Q2: Does CC out of single infinitive complements in Croatian depend on CL type?
- Q3: Does CC out of single infinitive complements in Croatian depend on CL case?
- Q4: Does CC out of single infinitive complements appear equally frequently in the standard and colloquial Croatian variety if the type of CTP verb (raising vs control) as a variable remains constant?

### 14.4 Methods

In order to answer the research questions we quantitatively analysed data obtained from three corpora. Forum, a subcorpus built from a hrWaC subdomain forum.hr (Forum) represents the informal register, while CNC and Riznica (Standard) are used as the source of formal data strongly influenced by prescriptive norms.<sup>7</sup> We present information on the matrix verbs chosen for the study as well as details of data retrieval in Chapter 12.

As mentioned in that chapter, we investigated variants with and without CC for 24 CTPs in two types of corpora representing standard and colloquial language varieties.<sup>8</sup> We limit ourselves to the analysis of raising and subject control verbs only. The study could have been extended to object control verbs. However, finding appropriate observations in corpora is very costly for several reasons, such as the frequency of particular lexemes in comparison to the size of the population of all object control lexemes, as well as the higher grade of complexity of object control predicates (in comparison to other CTPs), which necessarily

---

<sup>7</sup>For more information on corpora available for Croatian and their detailed descriptions see Chapter 4, where we also explain our choices concerning particular studies.

<sup>8</sup>For more information on structure variants see Table 12.1.

encode two arguments: subject and object.<sup>9</sup> We obtained 96 samples, upon which we built a logistic regression.<sup>10</sup> The model contains the following variables to be investigated with respect to the research questions: type of corpus, type of CTP, type and case of infinitive CL. These variables and their levels are summarised in Table 14.1 below.

Table 14.1: Variables used in the regression model

Variable name	Levels	Class
presence of CC	1 (CC present) 0 (CC absent)	dependent
corpus type	forum standard	independent
CTP type	raising subject control reflexive subject control	independent
case of infinitive CL	accusative dative no case	independent
type of infinitive CL	pronominal REFL <sub>LEX</sub> REFL <sub>2ND</sub>	independent

In the study we did not control for infinitive complements, but we ensured they did not influence the results (see more in Section 14.5).

<sup>9</sup>The manual revision of data for Chapter 13 taught us that CQL queries for object control matrices perform poorly in retrieval of CC. The CLs that appear in the matrix are not CLs which climb out of infinitive complements, but CLs which are complements of object control matrix predicates. In other words, manual revision of CC structure variants with object control matrix predicates would be extremely time consuming and would ultimately result in a small number of observations which we would not be able to analyse using the logistic regression model. These kinds of predicates are extensively studied in Chapter 15 since the experimental approach allows fast collection of the necessary amount of the observations.

<sup>10</sup>Our aim was to obtain fully crossed data (144 samples), that is, samples of the size of 100 for three variants (see Sections 12.2 and 12.5) from two corpora, but this task turned impossible for some CTPs.

## 14.5 Results and discussion

### 14.5.1 Data distribution

We now discuss the distributions of the independent variables in the context of the studied dependent variable, which was the presence of CC. The analysed data set comprised 2337 observations in total. CC occurred in 1850 cases, while in 477 cases, that is 20%, it did not occur. 1566 observations originated from Forum, and 761 from Riznica and CNC. This difference in the number of retrieved examples also corresponds to the difference in the size of the corpora used.

From the list of CTPs, we did not retrieve occurrences of *stidjeti se* ‘be ashamed’ in any of the three queried patterns, while the verbs *sramiti se* ‘be ashamed’ and *kretati* ‘go, to start’ were identified only in Forum. The size of samples obtained for different CTPs differed drastically. We identified 1027 observations of raising CTPs, 1118 of simple subject control and only 182 of reflexive subject control predicates. The frequencies for individual lexemes are shown below in Figure 14.1.

The distribution is quite proportional to the absolute frequency of the lexemes in the whole hrWaC presented in Table 12.3, but it does not precisely follow the same order. Simple subject control predicates are generally less frequent (with the exception of the verbs *željeti* ‘wish’ and *znati* ‘know’) than raising predicates, and reflexive subject control predicates are even less frequent than simple subject control predicates (with the exception of the verb *truditi se* ‘try’). Because of the overall differences in the frequencies of particular syntactic types, it is completely impossible to build frequency triplets with the three types of predicates. We elaborate further on that problem in Section 15.3.1.

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

Figure 14.2 shows the CL position for all studied CTPs. The plots on the left represent Forum, the plots on the right, Standard (Riznica + CNC). Raising and simple subject control predicates show a strong tendency to appear in CC constructions, in contrast to reflexive subject control predicates, which show the opposite trend. This is particularly visible in the case of the only well-represented verb, *truditi se* ‘try’, which has the same strong preference for not climbing in



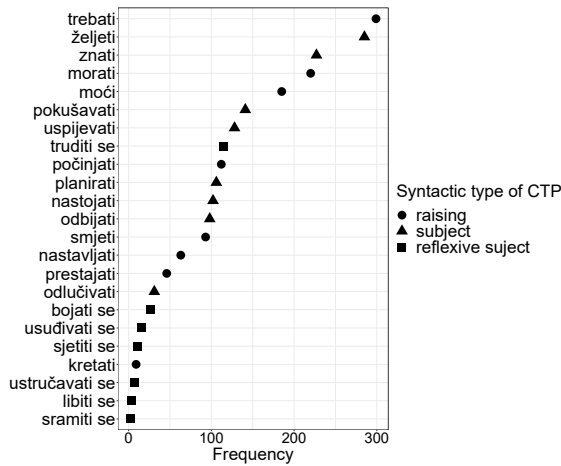


Figure 14.1: Distribution of different CTP lexemes in the data set: abscissa – number of observations per CTP lexeme, ordinate – CTP lexemes chosen for study.

both types of corpora. Also, at first glance the climbing seems more frequent in the observations from standard language corpora for all three types of predicates. The only exceptions seem to be the reflexive subject control matrix predicate *truditi se* ‘try’ and *libiti se* ‘hesitate’, which occur more often in noCC than in CC structures even in standard corpora. Furthermore, it is worth pointing out that structures without CC in the Forum subcorpus are distinctly more frequent in the case of the raising CTPs *trebati* ‘have to’ and *moći* ‘can’, and the simple subject control CTP *željeti* ‘wish/want’.

In standard corpora, the raising CTP *smjeti* ‘be allowed’ and simple subject control CTP *znati* ‘know’ are attested only in CC structures, as in examples (3) and (5) below. In Forum, both variants are attested for these verbs. Examples (4) and (6) show the noCC structures.

- (3) [...] ne smije<sub>1</sub> se<sub>2</sub> više baviti<sub>2</sub> čokoladom.  
 NEG be.allowed.3PRS REFL more occupy.3SG chocolate  
 ‘[...] he is not allowed to be in the chocolate business anymore.’ [Riznica]
- (4) Apsolutno smiju<sub>1</sub> baviti<sub>2</sub> se<sub>2</sub> znanošću [...].  
 absolutely be.allowed.3PRS occupy.INF REFL science  
 ‘They are absolutely allowed to pursue science [...]’ [hrWaC v2.2]
- (5) [...] koja se<sub>2</sub> ne zna<sub>1</sub> vratiti<sub>2</sub> u svoje toplo gnijezdo.  
 which REFL NEG know.3PRS come.back.INF in own warm nest  
 ‘[...] which does not know how to come back to its warm nest.’ [Riznica]

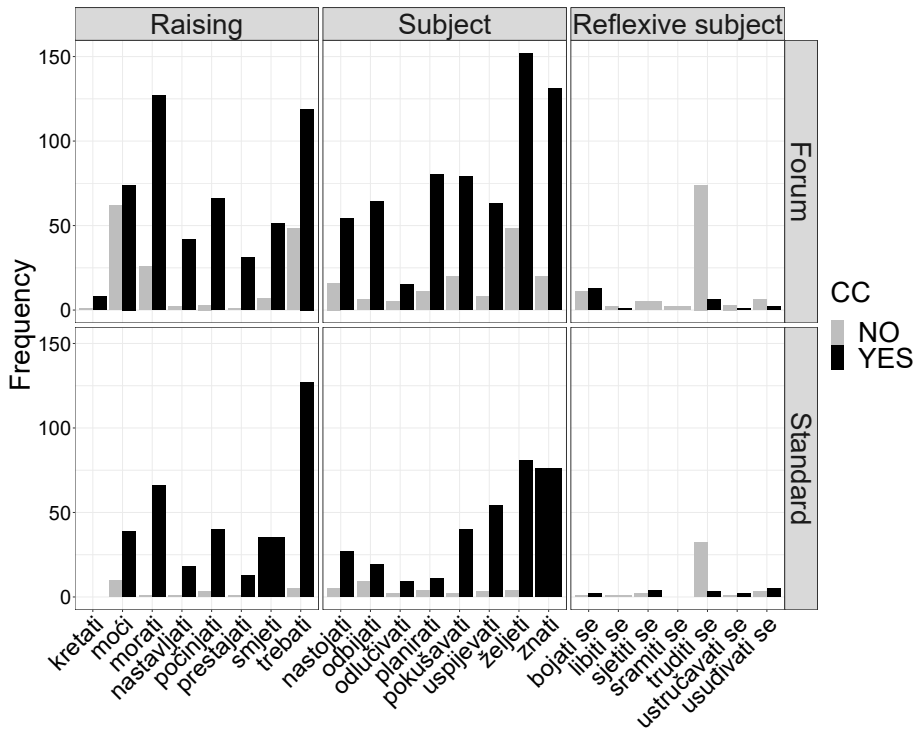


Figure 14.2: Verb-specific CL positioning across different CTP types and corpora

- (6) Mačke obično znaju<sub>1</sub> vratiti<sub>2</sub> se<sub>2</sub> kući [...].  
 cats usually know.3PRS come.back.INF REFL home  
 ‘Cats usually know how to come back home [...].’ [hrWaC v2.2]

Reflexive subject control predicates *sramiti se* ‘be ashamed’ and *sjetiti se* ‘remember’ were attested only in noCC structures in Forum, as in examples (7) and (8). The CTP *sramiti se* ‘be ashamed’ was attested only twice in our data, and only in Forum. The two utterances with *sramiti se* represent pseudo-twin structures – both matrix predicate and infinitive complement are reflexive (see example (7)). As explained in some detail in Section 11.4.1, these structures allow only pseudo-diaclisis or haplogy, and not a mixed cluster with two reflexive CLs.<sup>11,12,13</sup> The

<sup>11</sup>For basic information on pseudodiaclisis see Section 2.4.5.

<sup>12</sup>For basic information on haplogy see Section 2.4.2.2.

<sup>13</sup>This was also confirmed in our psycholinguistic study presented in Chapter 15.

second predicate, *sjetiti se* ‘remember’, was also attested in a pseudo-twins structure in Forum, and additionally for infinitive complements with pronominal CLs, as in (8), but not in constructions with mixed clusters containing reflexive matrix and pronominal infinitive CLs. Mixed clusters were attested with this CTP in the standard Croatian variety, see example (9). Moreover, in the case of *sjetiti se* ‘remember’, in standard corpora of Croatian mixed clusters like (9) are more frequent than pseudodiaclisis structures.

- (7) [...] ili  $se_1$  ti sramiš<sub>1</sub> maknuti<sub>2</sub>  $se_2$  iz [...].  
 or REFL you be.ashamed.2PRS move.away.INF REFL from  
 ‘[...] or you are ashamed to move away from [...].’ [hrWaC v2.2]
- (8) [...] a sjetim<sub>1</sub>  $se_1$  upisati<sub>2</sub>  $ga_2$  na forum [...].  
 and remember.1PRS REFL register.INF him.ACC on forum  
 ‘[...] and I remember to register it on the forum [...].’ [hrWaC v2.2]
- (9) [...] i rijetko  $ga_2$   $se_1$  tko sjeti<sub>1</sub> pozvati<sub>2</sub> na  
 and rarely him.ACC REFL who remember.3PRS invite.INF on  
 premijeru.  
 premiere  
 ‘[...] and people rarely remember to invite him to a premiere.’ [CNC]

We retrieved only one occurrence of the reflexive subject control predicate *libiti se* ‘hesitate’ from standard corpora. It was attested as a noCC structure in CNC: see example (10). However, in Forum this predicate was attested not only in noCC, but also in CC structures, as shown in example (11).

- (10) [...] da  $se_1$  ne libi<sub>1</sub> upotrijebiti<sub>2</sub>  $ga_2$  [...].  
 that REFL NEG hesitate.3PRS use.INF him.ACC  
 ‘[...] that he should not hesitate to use it [...].’ [CNC]
- (11) [...] pa da  $mu_2$   $se_1$  ne libe<sub>1</sub> staviti<sub>2</sub> brnjicu.  
 so that him.DAT REFL NEG hesitate.3PRS put.INF muzzle  
 ‘[...] so that they should not hesitate to muzzle him.’ [hrWaC v2.2]

We now move to the type and case of infinitive complement CL. Figure 14.3. presents CL type distribution across CTP types and corpora. In all, the  $REFL_{LEX}$  CL *se* is the most frequent ( $N = 1026$ ), while the pronominal ( $N = 691$ ) and the  $REFL_{2ND}$  CLs *se* and *si* have similar distributions ( $N = 610$ ).

If we consider the size of the retrieved samples, all three types of CLs are used similarly frequently in both types of corpora. The main difference in distribution concerns predicate type. In the sample of reflexive subject CTPs, we

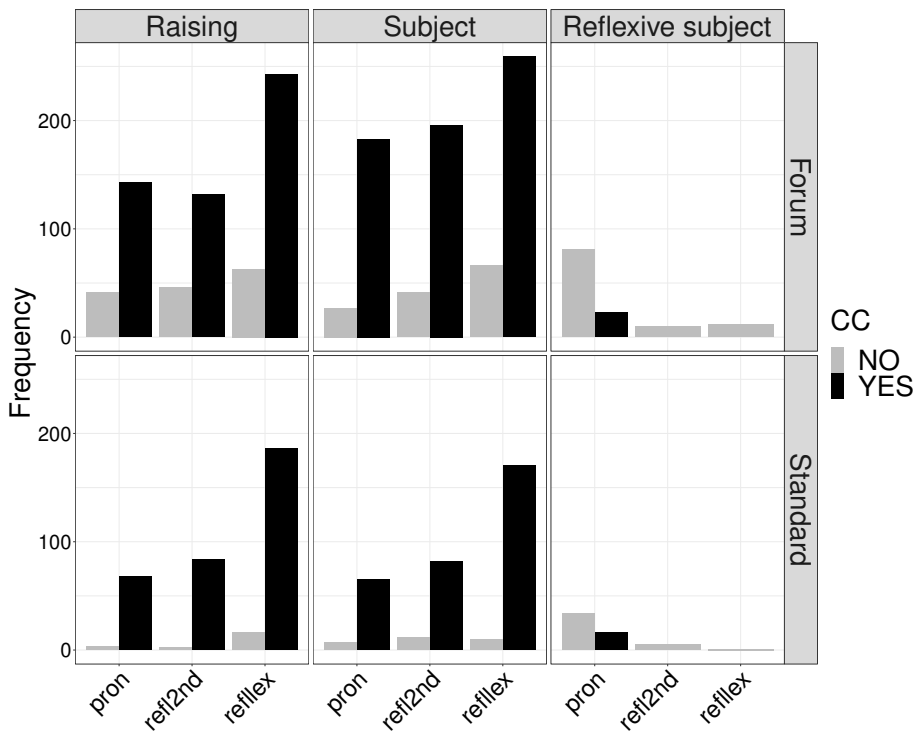


Figure 14.3: Type-specific CL positioning across different CTP types and corpora

retrieved mainly structures with pronominal infinitive CLs. Sentences with a reflexive subject CTP and reflexive infinitive CL such as the one in (12) are very rare in our data. This example contains the only occurrence of the  $REFL_{LEX}$  infinitive CL *se* appearing in pseudodiaclisis in the reflexive subject control sentences retrieved from standard corpora.

- (12) [...] *te se<sub>1</sub> usđuje<sub>1</sub> oprijeti<sub>2</sub> se<sub>2</sub> Tvom pozitivnom*  
 and REFL dare.3PRS withstand.INF REFL your positive  
*nalogu [...].*  
 ordering  
 ‘[...] and it (council) dares to oppose your express orders [...]’ [Riznica]

The possible reasons for this may be the syntactic rarity of the combination of a reflexive subject control predicate with a lexical reflexive complement or the usage of a competing construction such as haplology or *da<sub>2</sub>*-construction.

CC dominates for all types of CLs in the case of raising and simple subject control predicates. The noCC structures seem to be slightly more pronounced in Forum. In the sample of reflexive subject control predicates, pronominal CLs tend not to climb (only 39 of 154 CLs climb, that is, 25%), see examples (8) and (10). However, unlike pronominal CLs which can climb out of infinitive complements of reflexive subject control predicates, reflexive CLs do not climb at all: compare examples (9), (11), (13), and (14) on the one hand with examples (7) and (12) on the other hand.

- (13) A ja *ih*<sub>2</sub> se<sub>1</sub> bojim<sub>1</sub> drzati<sub>2</sub> skupa.  
and I them.ACC REFL be.afraid.3PRS keep.INF together  
'And I am afraid to keep them together.' [hrWaC v2.2]
- (14) [...] ali bojim<sub>1</sub> joj<sub>2</sub> se<sub>1</sub> jos davati<sub>2</sub> ljudsku hranu [...].  
but be.afraid.3PRS her.DAT REFL still give.INF human food  
'[...] but I am still afraid to give her food for humans [...].' [hrWaC v2.2]

These differences suggest that CL type is a constraint on CC in the case of reflexive subject control predicates. This is tested further in the next section

Finally, we present the distribution of case across predicate types and corpora. In general, accusative CLs ( $N = 963$ ) appeared three times as often as dative CLs ( $N = 338$ ). Since reflexive CLs seem to be distributed differently for reflexive subject control verbs, in the plot we distinguish the case of pronominal CLs and of REFL<sub>2ND</sub> CLs too. This is shown in Figure 14.4.

When examining CL type, we see that cases are not distributed equally across types – dative is more frequent as a pronominal case ( $N = 252$ ) than as the case of REFL<sub>2ND</sub> CLs ( $N = 86$ ). Accusative is used 439 times as the case of pronominal CLs, and 524 times for REFL<sub>2ND</sub> CLs. In general, the usage of CL cases is quite similar in both corpora. Nevertheless, the prevailing part of observations concerning dative reflexive CLs is from Forum ( $N = 78$ ), whereas the standard corpora yielded only 8 occurrences. Closer inspection of the data reveals further interesting differences. While in the standard corpora the REFL<sub>2ND</sub> CL *si* is a complement of infinitives which have an obligatory dative argument, such as *dopustiti* 'allow' (15) and *priušiti* 'afford' (16), the same CL is used in Forum as a complement of infinitives such as *kupiti* 'buy' (17) and *obnoviti* 'renew' (18). In the Croatian standard variety a dative complement of these infinitives is not obligatorily expressed when it refers to the subject itself, but it is usually inferred.<sup>14</sup>

<sup>14</sup>Petar Vuković (p.c.) claims that precisely such constructions with overtly expressed dative complement are features of the non-standard variety.

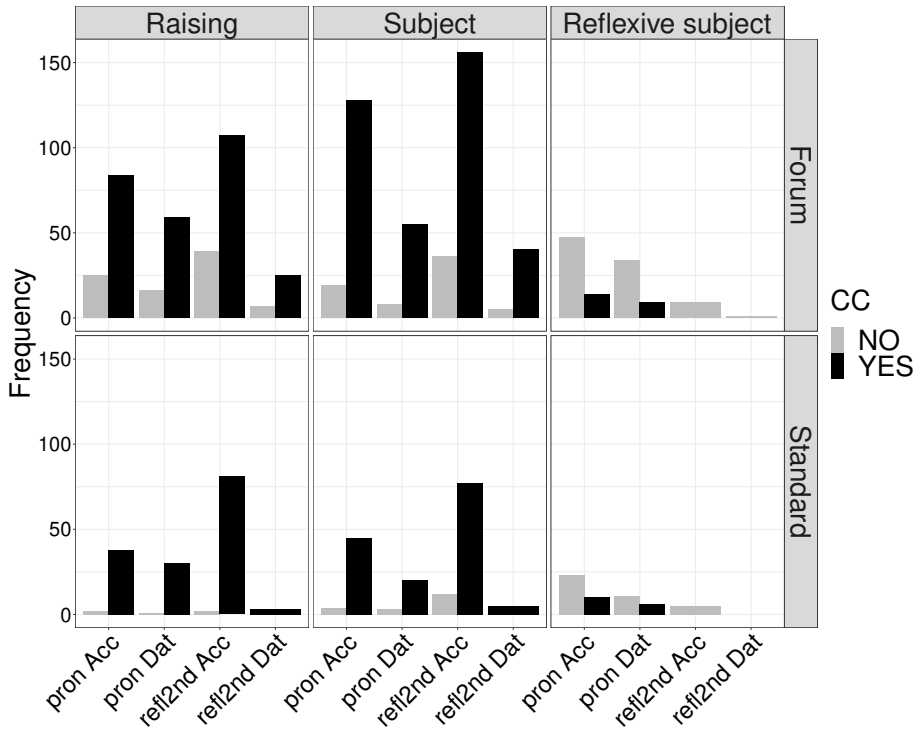


Figure 14.4: Case-specific CL positioning across different CTP types and corpora

- (15) [...] ne mogu<sub>1</sub> si<sub>2</sub> vise dopustiti<sub>2</sub> luksuz [...].  
 NEG can.3PRS REFL more allow.INF luxury  
 ‘[...] they cannot allow themselves the luxury [...].’ [Riznica]
- (16) [...] Hrvatice si<sub>2</sub> smiju<sub>1</sub> priustiti<sub>2</sub> poraz [...].  
 Croatians REFL be.allowed.3PRS afford.INF defeat  
 ‘[...] Croatians can afford to be defeated [...].’ [Riznica]
- (17) [...] a i ja si<sub>2</sub> planiram<sub>1</sub> kupiti<sub>2</sub> jedan [...].  
 and and I REFL plan.1PRS buy.INF one  
 ‘[...] and I am planning to buy myself one too [...].’ [hrWaC v2.2]
- (18) [...] i moram<sub>1</sub> si<sub>2</sub> malo obnoviti<sub>2</sub> garderobu.  
 and must.1PRS REFL little renew.3PRS wardrobe.INF  
 ‘[...] and I have to renew my wardrobe a little bit.’ [hrWaC v2.2]

Further, a closer look at our data reveals that in standard corpora the REFL<sub>2ND</sub> CL *si* was not attested at all with reflexive subject control predicates, and with

subject control and raising predicates it was attested only in CC structures (for the latter see example (15)). In Forum this CL was attested in both CC and noCC structures with raising and subject control predicates. Additionally, in Forum the REFL<sub>2ND</sub> CL *si* was also attested in a sentence with a reflexive subject control predicate, but, as expected, in a noCC structure: see example (19) below.

- (19) [...] tko *se*<sub>1</sub> *usudjuje*<sub>1</sub> *dati*<sub>2</sub> *si*<sub>2</sub> *nadimak* Master of [...].  
           who REFL dares.3PRS give.INF REFL alias       Master of  
           ‘[...] who dares call himself Master of [...].’                   [hrWaC v2.2]

Moreover, we would like to point out that Figure 14.4 does not reveal any striking differences between the accusative and the dative in relation to CC. CC is the more frequent construction for both cases, in both types of corpora for raising and simple subject control predicates, and the less frequent construction for both cases in both corpora types for reflexive subject control predicates. Summing up, we observe that the behaviour of CLs belonging to complements of reflexive subject control predicates shows an opposite trend as to CC than the other two types of CTPs. Reflexive CLs are generally rare and do not climb to the matrix at all. We see that CC is a slightly more unified mechanism in corpora representing standard language than in Forum. CL case does not seem to make any difference to CC as long as the CTP type is held constant, but climbing of reflexive CLs in the group of reflexive subject control CTPs does not seem to occur.

## 14.5.2 Testing correlations with a logistic regression model

### 14.5.2.1 Complement-taking predicate type and corpus type

In order to statistically test the relationships between CC, CTP type, infinitive CL type and case, and corpus type discussed in the previous subsection, we used logistic regression models with CTP lexemes as random effects. For our calculations we used the generalised linear mixed model fit by maximum likelihood from the lme4 R-package (Bates, Kliegl, Vasishth & Baayen 2015). The first model covered CTP type and corpus type. The remaining variables, type and case of the infinitive CL, were tested separately for two reasons. First, a model that includes case should include only CLs marked for case to avoid interaction with CL type. Second, we have very few observations for reflexive CLs of infinitive complements in sentences with reflexive subject control CTPs. The results are reported in Table 14.2.<sup>15</sup>

<sup>15</sup>The model formula is:  $CC \sim CtpType * CorpusType + (1|CtpVerb)$ ; for explanation of statistical measures in Table 14.2 and significance codes see Appendix B.

Table 14.2: Generalised mixed effects regression model

Random effects				
Groups Name	Var.	SD		
CtpVerb (Intercept)	0.4854	0.6967		
Number of obs: 2327, groups: CtpVerb, 23				
Fixed effects	Est.	SE	z	Pr(>  z )
(intercept CC-yes, raising, forum)	1.7498	0.2918	5.997	< 0.001 ***
simple subject	-0.1784	0.3969	-0.449	0.653
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	0.021 *
reflexive subject and standard	-0.7246	0.4978	-1.456	0.146

The results of the first model confirm our preliminary observations – corpus type and predicate type (reflexive subject control versus others) influence the probability of CC occurring in a sentence. We elaborate shortly on them.

The intercept in the model is CC occurring for raising CTPs in Forum and is used as a reference level for effects. The estimate of the intercept which is log odd can be recalculated to probability.<sup>16</sup> That is, the chance of CC occurring when a raising CTP is used in colloquial Croatian is 0.85. The other estimates refer to the change in log odds when particular effects are compared with the intercept. Thus, in colloquial Croatian there is no substantial difference between raising and simple subject CTPs, but there is only a 0.23 chance of CC occurring with a reflexive subject control CTP.

The change from colloquial to standard Croatian is significant, and has a positive effect on CC in the presence of raising CTPs. Namely, chances of CC increase to 0.96. This increase in probability also applies to simple subject control verbs, but the increase is significantly lower than for raising CTPs: the probability of CC is only 0.94. The change from the Forum subcorpus to standard Croatian has little impact on probability of CC in sentences with reflexive subject control verbs.

<sup>16</sup>The formula looks as follows:  $P = e^{\log O} / (1 + e^{\log O})$ , where P – probability, O – odds.



### 14.5.2.2 Infinitive clitic case and type

We built separate models for type and case of infinitive CLs; however, neither of them yielded any significant differences. Thus, for raising and simple subject control predicates neither the case of pronominal and reflexive infinitive CLs nor the type of infinitive CL appears to be a relevant factor influencing CC. The small number of observations for infinitive reflexive CLs in constructions with reflexive subject predicates leads to the conclusion that in the case of infinitive complements these CLs are haplogotised (i.e., omitted), or that an alternative construction, for example, with a *da*<sub>2</sub>-complement, is used.

## 14.6 Conclusions

This study gives the following answers to our research questions:

- A1: The difference between raising and simple subject control CTPs is statistically significant for CC only in standard Croatian, while CC with reflexive subject control CTPs is significantly less frequent than with the other two types of CTPs regardless of diaphasic variety.
- A2: The type of infinitive CL does not influence CC out of single infinitive complements either in standard or in colloquial Croatian.
- A3: The case of the infinitive CL is not a relevant factor for CC out of single infinitive complements either in standard or in colloquial Croatian.
- A4: Diaphasic variation is a significant factor influencing the probability of CC. Unlike in Romance languages, in Croatian CC appears more frequently in the standard than in the colloquial variety.

These findings allow some tentative observations to be made, which should feed into future research. Although in standard Croatian CC out of a single infinitive complement appears highly probable with raising CTPs, it does not seem to be absolutely obligatory (pace Aljović 2005, in accordance with Hansen, Kolaković & Jurkiewicz-Rohrbacher 2018). Colloquial language in particular allows the lack of CC to a certain degree. This tendency, however, is not universal to CC languages since Romance languages exhibit the opposite trend – CC is significantly more frequent in colloquial language, whereas in formal language CLs are more likely to appear in noCC constructions.

Furthermore, our assumption that a differentiation of simple and reflexive subject control CTPs hitherto neglected in theoretical syntactic research on CL could actually shed new light on mechanisms of CC is justified. In order to get more data, we further explore the possibilities of CC in the context of reflexive subject control CTPs in Chapter 15, where we report a psycholinguistic experiment. As in the case of reflexive subject control CC inevitably leads to mixed CL clusters, we might conjecture that there could exist a strategy to avoid such mixed clusters. Therefore, in order to broaden the database to include structures which might lead to such mixed clusters, in the next chapter object control predicates are studied in addition to reflexive subject control predicates.

# 15 Experimental study on constraints on clitic climbing out of infinitive complements (Croatian)

## 15.1 Introduction

As we have already pointed out, some of the data on CLs based on linguists' informal judgments have turned out to be flawed.<sup>1</sup> Therefore our goal is to provide data which do not suffer from bias, unreliability, and narrowness by testing a part of the constraints on CC previously discussed in the syntactic literature.<sup>2</sup> To supplement the results of our corpus linguistic studies on CC, we decided to broaden the available information on CC out of infinitive complements to include empirical data collected through acceptability judgment tasks.<sup>3</sup>

This study complements our investigations described in the previous chapters, in accordance with the principle of triangulation of methods described in Section 3.2.1. As pointed out, we follow the scheme: intuition/theory – observation – experiment. Chapter 11 presents the first step in this procedure, in which we give an exhaustive account of the constraints scattered across the literature, and pretest them for BCS by retrieving naturally occurring constructions and performing informal acceptability judgment tasks on them. We further tested some of the constraints on CC mentioned in that chapter in more exhaustive corpus linguistic studies presented in Hansen, Kolaković & Jurkiewicz-Rohrbacher (2018), Chapters 13, and 14. These studies belong to the second step. We are aware that corpus studies cannot provide negative evidence and that control over influencing factors in corpus studies is limited.<sup>4</sup> To some extent, we can overcome these

---

<sup>1</sup>For more information see Chapters 3, 9, 11, and 13.

<sup>2</sup>We warn our readers that the reception of this chapter requires an acquaintance with various phenomena which are closely related to CC. Therefore, we advise our readers to at least become closely acquainted with the contents of Chapters 2, 3, 10, and 11 before reading this chapter. Ideally, one should have first read all the chapters dedicated to CC in this part III of the book before reading this chapter.

<sup>3</sup>For more information on the corpora and queries used in our corpus linguistic studies on CC see Sections 4.6.3 and 12.2. The results are presented in Hansen, Kolaković & Jurkiewicz-Rohrbacher (2018), Chapters 13, and 14.

<sup>4</sup>For more information on the drawbacks of corpus studies see 3.3.2.4.

problems with experimental manipulation and presentation of stimuli. This is the third step in our triangulation of methods. The acceptability judgment experiment allow us to test the hypotheses formulated during the intuition/theory and observation (corpus) steps with a high level of control over individual factors.

As mentioned in Section 3.3.3.5, there are two main advantages of judgment experiments. First, they can provide negative data and data which cannot be collected otherwise. In other words, introspection experiments such as acceptability judgments make possible the investigation of rare phenomena that fail to appear even in a very large corpus (such as web corpus). Low acceptability of a structure is considered negative evidence, i.e., it indirectly indicates that such a structure is very probably not used by native speakers. Second, if the test is designed properly, judgment data have internal validity, i.e. these kinds of studies allow unambiguous causal inferences.

In what follows we describe the process of systematic collection of data which fulfil all the requirements of inferential statistical methods. This allows more robust generalisations on some of the constraints on CC. In Section 15.2 we generate research questions concerning CC basing on constraints previously put forward in this book. Section 15.3 brings exhaustive information on the test setup: selection of matrix verbs (CTPs), test design, production of stimuli, and participants. The experimental procedure together with data preprocessing is explained in Section 15.4. Our results with respect to each research question are thoroughly discussed in Sections 15.5 and 15.6. In Section 15.7 we analyse the reaction time, our control measure, for accepted sentences. Section 15.8 puts forward an overview of the results summed up in general conclusions about CC in BCS.

## 15.2 Research questions

Building on the previous research on CC in Czech and BCS summarised in Chapter 11, we turn to the present study, in which we further explore the impact of the raising–control distinction and selected mechanisms mentioned in Sections 11.3 and 11.4.<sup>5,6</sup>

Our first research question addresses the raising–control distinction, which has been reported as crucial for CC in Czech. Our next six research questions

---

<sup>5</sup>We exclude the constraint tightly connected to object control and animacy of the CL referent described in Section 11.3.4, which lies beyond the scope of this study. However, bearing in mind that this factor may be important, we kept it constant through all experimental situations.

<sup>6</sup>For basic information on different predicate types with respect to the raising–control dichotomy see Section 2.5.2.

concern fine differences between CTP (sub)types. The last two research questions address the type and case of the infinitive CL. Some of those differences have only been discussed in the literature on CC in Czech, while others have been partially addressed in some of our papers or in other chapters of this book.

In Chapter 13 we empirically show that the raising–control dichotomy plays an important role in CC out of *da*<sub>2</sub>-complements in Serbian.<sup>7</sup> We formulate our first research question with the aim of experimentally investigating whether matrix predicate types are relevant also with respect to CC out of infinitive complements.

To the best of our knowledge, we are the first to compare differences in CC rates of simple subject control predicates on the one hand and reflexive subject control predicates on the other in corpora of standard and colloquial Croatian (cf. Kolaković et al. 2019, see Chapter 14). Moreover, the study by Hansen, Kolaković & Jurkiewicz-Rohrbacher (2018) on CC out of stacked infinitives also showed that reflexivity of the matrix predicate influences CC. Our second research question is intended to test differences in CC rates of simple subject and reflexive subject control predicates via acceptability judgment tasks.

Authors working on CC in Czech mention the object control case constraint. Namely, object control predicates with a dative controller block only the climbing of dative pronominal CLs, while object control predicates with an accusative controller build even stronger barriers, and block not only the climbing of dative but also of accusative pronominal CLs. Since obtaining evidence from corpora is rather difficult and would require excessive manual filtering and checking, an acceptability judgment task seems to be the most suitable approach for examining this topic empirically. Our third research question thus addresses this constraint in the context of CC out of infinitive complements in Croatian.

Scholars who work on the object control case constraint on CC in Czech base their discussion only on object control predicates with pronominal or NP controllers. Since reflexivity is recognised as an important factor in CC and since object control predicates with a reflexive controller have not previously been included in the discussion of CC, we formulate this as our fourth research question. Moreover, we are the first to directly compare the behaviour of the two *REFL*<sub>2ND</sub> CLs, *se* and *si*.

The literature review shows that object control predicates with pronominal or NP controllers in the dative trigger restrictions only on the climbing of dative CLs. Furthermore, in this case reflexivity might be an important factor in limiting

---

<sup>7</sup>For more information on *da*-complements see Section 2.5.3.

the range of CC. Our fifth research question involves a comparison of object control predicates with pronominal CL controllers in the dative and object control predicates with the reflexive controller *si* in the dative. Analogously, the sixth research question is dedicated to a comparison of pronominal CL controllers in the accusative with the reflexive controller *se*.

Next, reflexive subject control predicates and object control predicates with the REFL<sub>2ND</sub> controller *se* have not been compared with respect to CC in the previous works. Moreover, the mentioned reflexives are not only of different types (REFL<sub>LEX</sub> vs REFL<sub>2ND</sub>), but they also appear with different matrix predicates (subject vs object control). Our seventh research question addresses CC in the context of these differences.<sup>8</sup>

The literature on CC in Czech indicates that the type of infinitive CL (pronominal vs reflexive) plays an important role in CC. It has been claimed that unlike pronominal CLs, reflexives cannot climb out of object-controlled infinitives.<sup>9</sup> These claims motivated us to formulate our eighth research question.

Scholars working on CC in Czech indicate that pronominal infinitive CL complements in the accusative are less restricted in the climbing than pronominal infinitive CL complements in the dative, provided that the matrix predicate is of the object control type.<sup>10</sup> We address this in our ninth research question.

Junghanns (2002) and Rosen (2014) investigate the problem of phonologically identical/different and morphologically identical/different CLs with different governors in respect of CC.<sup>11</sup> Rosen (2014) offers haplology as a solution to CC in such contexts.<sup>12</sup> However, due to the design of our experiment we cannot test such sentences. Addressing this phenomenon properly would require systematic investigation of the factors that influence haplology, i.e. manipulating the position of CLs, understanding which of two CLs is being eliminated etc. As a consequence, the number of sentences on the list would increase beyond a size which is reasonable for participants. Therefore, we made an informed decision to leave haplology for separate, future research.

Our set of research questions thus targets the following variables:

- type of matrix verb (including the reflexivity of the predicate),
- number of CLs in a sentence,

---

<sup>8</sup>For more information on types of reflexives see Section 2.5.4.

<sup>9</sup>For Czech examples see Section 11.3.5.

<sup>10</sup>For Czech examples see Section 11.3.2.

<sup>11</sup>For more information and Czech examples see Section 11.4.1.

<sup>12</sup>For more information on haplology see Section 2.4.2.2.

- type of infinitive CL,
- case of infinitive CL,
- position of the infinitive CL (CC vs no CC).

An exhaustive description of dependent and independent variables and their levels can be found in Sections 15.3.1 and 15.3.2.

In this chapter we address the following nine research questions:

- RQ1: Do raising, subject and object control matrix predicates behave the same with respect to CC out of their infinitive complements?
- RQ2: Do simple subject control predicates (such as *planirati* ‘plan’) and reflexive subject control predicates (such as *bojati se* ‘be afraid’) behave the same with respect to CC out of their infinitive complements?
- RQ3: Do object control predicates with a pronominal CL controller in the dative and those with a pronominal CL controller in the accusative behave the same with respect to CC out of their infinitive complements?
- RQ4: Do object control predicates with a  $\text{REFL}_{2\text{ND}}$  CL *si* controller and those with a  $\text{REFL}_{2\text{ND}}$  CL *se* controller behave the same with respect to CC out of their infinitive complements?
- RQ5: Do object control predicates with a pronominal CL controller in the dative and those with a  $\text{REFL}_{2\text{ND}}$  CL *si* controller behave the same with respect to CC out of their infinitive complements?
- RQ6: Do object control predicates with a pronominal CL controller in the accusative and those with a  $\text{REFL}_{2\text{ND}}$  CL *se* controller behave the same with respect to CC out of their infinitive complements?
- RQ7: Do reflexive subject control predicates with the reflexive CL *se* and object control predicates with a  $\text{REFL}_{2\text{ND}}$  CL *se* controller behave the same with respect to CC out of their infinitive complements?
- RQ8: Do pronominal and reflexive ( $\text{REFL}_{\text{LEX}}$  CL *se* and  $\text{REFL}_{2\text{ND}}$  CL *se* and *si*) infinitive CLs behave the same with respect to CC if the type of matrix predicate is constant?
- RQ9: Do dative and accusative infinitive CLs behave the same with respect to CC provided that the type of matrix predicate is constant?

## 15 Experimental study on clitic climbing out of infinitive complements

These ten research questions are operationalised in the form of null hypotheses as follows:

- H<sub>0.1</sub>: Raising, subject and object control predicates behave the same with respect to CC out of their infinitive complements.
- H<sub>0.2</sub>: Simple subject control predicates (such as *planirati* ‘plan’) and reflexive subject control predicates (such as *bojati se* ‘be afraid’) behave the same with respect to CC out of their infinitive complements.
- H<sub>0.3</sub>: Object control predicates with a pronominal CL controller in the dative and those with a pronominal CL controller in the accusative behave the same with respect to CC out of their infinitive complements.
- H<sub>0.4</sub>: Object control predicates with a REFL<sub>2ND</sub> CL *si* controller and those with a REFL<sub>2ND</sub> CL *se* controller behave the same with respect to CC out of their infinitive complements.
- H<sub>0.5</sub>: Object control predicates with a pronominal CL controller in the dative and those with a REFL<sub>2ND</sub> CL *si* controller behave the same with respect to CC out of their infinitive complements.
- H<sub>0.6</sub>: Object control predicates with a pronominal CL controller in the accusative and those with a REFL<sub>2ND</sub> CL *se* controller behave the same with respect to CC out of their infinitive complements.
- H<sub>0.7</sub>: Reflexive subject control predicates with the reflexive CL *se* and object control predicates with a REFL<sub>2ND</sub> CL *se* controller behave the same with respect to CC out of their infinitive complements.
- H<sub>0.8</sub>: Pronominal and reflexive (REFL<sub>LEX</sub> CL *se* and REFL<sub>2ND</sub> CLs *se* and *si*) infinitive CLs behave the same with respect to CC if the type of matrix predicate is constant.
- H<sub>0.9</sub>: Dative and accusative infinitive CLs behave the same with respect to CC provided the type of matrix predicate is constant.

### 15.3 The test set-up

Since the value of acceptability judgment data depends on the validity of the experimental procedures (cf. Myers 2017), we decided to take all necessary steps in



following all recommendations possible with respect to the test set-up, test design, stimulus production, selection of participants, and procedure. These aspects will be discussed in the following sections.

### 15.3.1 Selection of matrix verbs

The research questions presented in the previous section form the main guidelines for designing the experiment and later for analysis of the data. The question of potential constraints is explored through sentence processing, i.e. an acceptability judgment task.<sup>13</sup>

Each sentence is a carefully developed stimulus in the experiment. Basing on the research questions from Section 15.2 we will now discuss the elements which the stimuli must contain. For the raising–control constraint the following three major predicate types are used in the stimuli: raising (e.g. *moći* ‘can’), subject (e.g. *pokušavati* ‘try’) and object control (e.g. *pomagati* ‘help’). Furthermore, since we investigate the role reflexivity plays in CC, the latter two groups must be further divided.<sup>14</sup> Thus, in the subject control group we have:

- simple subject control verbs (e.g. *planirati* ‘plan’),
- reflexive subject control verbs, i.e. verbs with the REFL<sub>LEX</sub> CL *se* (e.g. *bojati se* ‘be afraid’).

The object control group includes predicates which have a pronominal CL as a controller and those which have a reflexive CL as a controller. Since reflexivity and case of the matrix complement (i.e. controller) are addressed, the group of object control predicates is divided into four subgroups:

- object control matrix predicates with a pronominal CL controller in the dative (e.g. *pomagati* ‘help’),
- object control matrix predicates with a pronominal CL controller in the accusative (e.g. *poticati* ‘encourage’),
- object control matrix predicates with the REFL<sub>2ND</sub> CL *si* controller (e.g. *dozvoljavati si* ‘allow oneself’),

<sup>13</sup>An explanation of why this and not some other psycholinguistic test was chosen can be found in Section 3.3.4.

<sup>14</sup>For more information on the role of reflexivity see Sections 11.2.5.3, 11.2.5.4, 11.3.5, 11.4.2, 11.4.3, the results of our studies presented in Hansen, Kolaković & Jurkiewicz-Rohrbacher (2018), and in Chapter 14.

## 15 Experimental study on clitic climbing out of infinitive complements

- object control matrix predicates with the REFL<sub>2ND</sub> CL *se* controller (e.g. *prisiljavati se* ‘force oneself’).

Summing up, the following seven types of matrix verbs are distinguished for the purpose of stimulus preparation:

- raising,
- simple subject control verbs,
- reflexive subject control verbs, i.e. verbs with the REFL<sub>LEX</sub> CL *se*,
- object control verbs with a pronominal CL controller in the dative,
- object control verbs with a pronominal CL controller in the accusative,
- object control verbs with the REFL<sub>2ND</sub> CL *si* controller,
- object control verbs with the REFL<sub>2ND</sub> CL *se* controller.

We chose the verbs according to the procedure described in Section 12.4. Since the use of tenses other than the present tense implies the use of auxiliary CLs in the matrix clause, we constructed the stimuli using matrix predicates in the present tense only.<sup>15</sup> We avoided stimuli with auxiliary CLs since we are not sure of their impact on CC. Further we narrowed our list down to imperfective verbs only.<sup>16</sup> In the case of verbs with the same stem but different prefixes such as *po-*, *za-*, *od-* in *počinjati*, *započinjati*, *otpočinjati* ‘begin’, we take the one with the least complex lexical meaning. In the case of *započinjati* and *otpočinjati* prefixes put additional emphasis on the beginning, that is, on the first phase of the situation expressed by those verbs. Therefore, we decided on the most neutral variant *počinjati*. That was usually also the most frequent variant. The group of raising CTPs is very small and contains only imperfective predicates with obligatory raising of the subject (i.e. modal and phasal verbs). Its eight members are listed in Table 15.1.

---

<sup>15</sup>In BCS there are other simple tenses besides the present tense, such as the aorist and imperfect. However, their usage in everyday language is stylistically restricted and it would not make much sense to construct stimuli for acceptability judgment tasks with them. In fact, it might even be dangerous, since those tenses could influence the evaluation of the stimuli.

<sup>16</sup>The use of present tense as actual present in simple and main clauses requires imperfective aspect. In a few cases we used perfective verbs, but we then used temporal adverbials implying habituality (see Dickey 2000 on the aspect of BCS habituales).

Table 15.1: Raising predicates selected for the acceptability judgment experiment. The frequencies of the lemmas are taken from hrWaC v2.2 and expressed per million words.

Nr.	Verb	Translation	Frequency
1.	<i>moći</i>	‘can’	4056.5
2.	<i>trebati</i>	‘have to’	1761.4
3.	<i>morati</i>	‘must’	1224.8
4.	<i>smjeti</i>	‘be allowed’	208.0
5.	<i>počinjati</i>	‘start’	125.0
6.	<i>kretati</i>	‘go/start’	101.2
7.	<i>nastavljati</i>	‘continue’	77.7
8.	<i>prestajati</i>	‘stop’	24.9

During the preparatory phase various decisions had to be taken. First, in all seven groups of predicates the same number of matrix verbs had to be selected. Since the number of raising predicates extracted to design stimuli on the first experimental list was eight, it defined the maximal number of predicates on each of the other six experimental lists.

It is commonly known from psycholinguistic studies that the frequency of a word has a wide impact (Baayen et al. 2016, Brysbaert et al. 2018). Therefore, in an ideally designed experiment the frequencies of selected verbs should match across different predicate type groups. Unfortunately in our case this was not possible, since subject control predicates and object control predicates which can take infinitive complements are much less frequent than raising predicates.<sup>17</sup> Therefore, the most frequent verbs are usually taken. We justify our decision below.

Table 15.2 shows the subject control predicates selected for the experiment; non-reflexive on the left side of the table and reflexive on the right.

Some frequent subject control verbs do not appear on the list. Although *htjeti* ‘will/ want’ is the most frequent subject control verb, we exclude it from the study since it is predominantly used as a future tense auxiliary. Instead we take *željeti* ‘want/wish’. Since the list of potential CTP candidates is long, we avoid partial synonyms. For example, the choice of *planirati* rules out *namjeravati* ‘intend/ plan’ as these verbs have very similar meanings. Furthermore, we

<sup>17</sup>For more information on complement types in BCS, see Section 2.5.3. Most object control predicates in Croatian actually take  $da_2$ -complements.

15 Experimental study on clitic climbing out of infinitive complements

Table 15.2: Subject control predicates selected for the acceptability judgment experiment

Nr.	Verb	Translation	Freq.	Verb	Translation	Freq.
1.	<i>znati</i>	'know/can'	1584.6	<i>bojati se</i>	'be afraid'	106.0
2.	<i>željeti</i>	'want/will'	859.1	<i>sjetiti se</i>	'remember'	79.1
3.	<i>pokušavati</i>	'try'	139.6	<i>truditi se</i>	'try'	53.3
4.	<i>planirati</i>	'plan'	105.7	<i>sramiti se</i>	'be ashamed'	12.4
5.	<i>nastojati</i>	'strive'	76.2	<i>usuđivati se</i>	'dare'	5.6
6.	<i>odlučivati</i>	'decide'	54.9	<i>stidjeti se</i>	'be ashamed'	4.2
7.	<i>odbijati</i>	'refuse'	32.6	<i>libiti se</i>	'hesitate'	3.0
8.	<i>uspijevati</i>	'succeed'	31.9	<i>ustručavati se</i>	'hesitate'	2.5

exclude the quite frequent verb *misliti* 'intend', since it is more common with *da*<sub>1</sub>-complements in its other meaning 'think'. We also exclude all verbs of motion such as *ići* 'go', *dolaziti* 'come' and *ostajati* 'stay' as they are often used with final subordinate clauses and with a *da* complementiser.

Table 15.3 shows the object control predicates selected for the experiment.<sup>18</sup> The list includes both object control predicates with a dative controller (left side

<sup>18</sup>In certain contexts, the verb *učiti* in table 15.3 can mean 'learn'. The problem of polysemy was solved through context, i.e. from a given sentence it was clear that the meaning 'teach' was employed.

Table 15.3: Object control predicates selected for the acceptability judgment experiment

Nr.	Verb (DAT)	Translation	Freq.	Verb (ACC)	Translation	Freq.
1.	<i>pomagati</i>	'help'	112.5	<i>učiti</i>	'teach'	104.4
2.	<i>omogućivati</i>	'enable'	64.7	<i>poticati</i>	'encourage'	59.2
3.	<i>savjetovati</i>	'advise'	38.8	<i>tjerati</i>	'force'	29.9
4.	<i>preporučivati</i>	'recommend'	34.5	<i>puštati</i>	'let/allow'	26.4
5.	<i>dopuštati</i>	'allow'	33.2	<i>obvezivati</i>	'oblige'	9.9
6.	<i>dozvoljavati</i>	'allow'	19.8	<i>prisiljavati</i>	'force'	6.2
7.	<i>zabranjivati</i>	'forbid'	12.1	<i>požurivati</i>	'hurry up'	1.2
8.	<i>naređivati</i>	'order'	3.5	<i>primoravati</i>	'compel'	0.6

of the table) and object control predicates with an accusative controller (right side of the table).

The object control predicates with a REFL<sub>2ND</sub> controller selected for the experiment are presented in Tables 15.4 and 15.5.<sup>19</sup> The former contains object control predicates with the REFL<sub>2ND</sub> controller *si*, while the latter is object control predicates with the REFL<sub>2ND</sub> controller *se*.

Table 15.4: Object control predicates with the REFL<sub>2ND</sub> CL *si* controller selected for the acceptability judgment experiment

Nr. Verb (DAT)	Translation	Frequency
1. <i>braniti si</i>	‘forbid oneself’	81.6
2. <i>omogućivati si</i>	‘enable oneself’	64.7
3. <i>dopustiti si</i>	‘allow oneself’	33.2
4. <i>priuštiti si</i>	‘allow oneself’	21.5
5. <i>dozvoljavati si</i>	‘allow oneself’	19.8
6. <i>olakšavati si</i>	‘make something easier for oneself’	11.7
7. <i>naređivati si</i>	‘assign oneself’	3.5
8. <i>uskraćivati si</i>	‘deprive oneself’	3.3

Object control predicates with the REFL<sub>2ND</sub> CL *si* controller occur quite rarely. Although some of them might sound slightly odd, like *naređivati si* ‘assign oneself’ and *braniti si* ‘forbid oneself’, all are attested in corpora. A similar problem is encountered on the list of object control predicates with a REFL<sub>2ND</sub> CL *se* controller, although most of the verbs in Table 15.5 are used in everyday language. Only *ovlašćivati se* ‘authorise’ is an exception: it is typical of the administrative register. Nevertheless, also the verbs in this group are all attested in corpora.

### 15.3.2 Experiment design

In stimulus design, a fully crossed factorial design is usually aimed for. This implies that each level of an independent variable is crossed with each level of other independent variables. Such a design provides the highest level of methodological rigour. However, since our stimuli are extracted from natural language, applying a fully crossed design was not possible. In other words, certain combinations

<sup>19</sup>In Tables 15.2–15.5, frequency refers to the frequency of lemmas without reflexive markers since in hrWaC v2.2 all reflexives are annotated separately and it is not possible to extract the exact data on the frequency of the lemma with a reflexive. The high frequency of *braniti* is a result of homonymy. There are actually two lemmas: *braniti* ‘defend’ and *braniti* ‘forbid.’

## 15 Experimental study on clitic climbing out of infinitive complements

Table 15.5: Object control predicates with a REFL<sub>2ND</sub> CL *se* controller selected for the acceptability judgment experiment

Nr. Verb (ACC)	Translation	Frequency
1. <i>učiti se</i>	‘teach oneself’	104.4
2. <i>poticati se</i>	‘encourage oneself’	59.2
3. <i>spremati se</i>	‘prepare oneself’	39.6
4. <i>prisiljavati se</i>	‘force oneself’	6.2
5. <i>ohrabrivati se</i>	‘encourage oneself’	5.1
6. <i>navikavati se</i>	‘accustom oneself’	1.7
7. <i>ovlašćivati se</i>	‘authorise oneself’	0.7
8. <i>primoravati se</i>	‘compel oneself’	0.6

of factor levels do not exist in language, or are too rare for enough examples to be found and build a fully fledged list of stimuli. Although it is sometimes possible to artificially construct critical examples, this is not the rule but rather an exception.

For example, the variables of highest interest to us are type of matrix verb (raising, subject control, object control), number of CLs (one, two), type of matrix and infinitive CL (personal pronoun, REFL<sub>2ND</sub>, REFL<sub>LEX</sub>), and case of matrix and infinitive CL (dative, accusative). However, achieving a fully crossed factorial design with all these variables is not possible, as for example due to their argument structure matrix verbs of the raising type do not have CL complements. Therefore, they do not appear in constructions with two CLs. In contrast, object control matrix verbs always form such constructions as they have their own CL complements (i.e. controllers) and their infinitive complements also have CL complements.<sup>20</sup> Similarly, some subject control matrix predicates have the REFL<sub>LEX</sub> CL *se*, whereas object control matrix verbs have either pronominal or REFL<sub>2ND</sub> complements (i.e. controllers).<sup>21</sup> Furthermore, even if all the combinations were present in language, given the number of our variables of interest, permuting them would give us a very high number of stimuli. Such a large number of sentences would be too demanding for experiment participants. This would not only decrease the

<sup>20</sup>We are aware that object control predicates can have a NP instead of CL complement/controller. But the fact that they have one more complement than raising predicates still remains.

<sup>21</sup>We are aware that some subject control predicates like *obećati* ‘promise’ are polyvalent and that they can have a NP or CL complement in the dative, for more information see Section 2.5.2. But the fact that subject control predicates differ from the raising and object control predicates with respect to their complements still remains.

reliability of the observed data (given the fatigue level), but also present ethical issues. Therefore, our design was a compromise between methodological rigour, availability of language material, and operational capabilities of participants. At the same time, we could also call it an optimal solution for tackling the research questions given the language structure and operational capabilities of participants.

With all this in mind, we developed a design which enables us to examine the relationship between the dependent variable (sentence acceptability) and the four independent variables mentioned at the end of Section 15.2. These are summarised in Table 15.6.

In accordance with the explanations from Section 15.3.1, the first independent variable is predicate type with seven levels (raising verbs, simple subject control verbs, reflexive subject control verbs with the  $REFL_{LEX}$  CL *se*, object control verbs with a pronominal CL controller in the dative, object control verbs with a pronominal controller in the accusative, object control verbs with the  $REFL_{2ND}$  CL *se* controller, object control verbs with the  $REFL_{2ND}$  CL *si* controller). For the reasons discussed above, this independent variable was introduced as a between-participants (different participants were presented with different predicate types) and between-items factor (as one verb cannot belong to multiple predicate types).

The second independent variable was type of the infinitive CL, which had three levels (pronominal,  $REFL_{2ND}$ ,  $REFL_{LEX}$ ). The third independent variable was case of the infinitive CL, with two levels (dative, accusative). This factor was nested in two levels of the second independent variable (pronominal and  $REFL_{2ND}$ ), as  $REFL_{LEX}$  does not have grammatical case. Within a given matrix verb type, both the second and the third factor were introduced as within-participant and between-item factors.

Finally, we manipulated the position of the infinitive CL, i.e., we introduced the fourth independent variable, CC, which had two levels (CC present, CC absent). This variable was introduced as both a within-participant and a within-item factor. Given that the phenomenon of CC is central to our study, we found it crucial to allow for the comparison of acceptability scores of the same sentence in two conditions: with and without CC. We counterbalanced the position of the critical CL by applying the Latin square design, as we will describe in more detail in the next section.<sup>22</sup>

In order to control for the effects of the additional variables that are not subject to manipulation in this research, we introduced some additional restrictions.

---

<sup>22</sup>The critical CL is the CL of interest, i.e. the infinitive CL complement whose climbing is being tested.

Table 15.6: List of variables

Variable	Type	Levels	Class
sentence acceptability	binary	0 (unacceptable)	dependent
		1 (acceptable)	
		raising	independent
		simple subject control reflexive subject control object control with pronominal CL	
predicate type	categorical	controller in dative	independent
		object control with pronominal CL	
		controller in accusative	
		object control with REFL <sub>2ND</sub> CL <i>si</i> controller object control with REFL <sub>2ND</sub> CL <i>se</i> controller	
type of infinitive CL	categorical	pronominal	independent
		REFL <sub>2ND</sub> REFL <sub>LEX</sub>	
case of infinitive CL	categorical	dative	independent
		accusative	
CC	binary	0 (CC absent)	independent
		1 (CC present)	



First, we controlled for the animacy of the CL referents and constructed sentences from which it is clear that the CL referents are animate.<sup>23</sup> Next, we controlled for the person of the critical pronominal CL and constructed only sentences with third person pronominal CLs.<sup>24,25</sup> Additionally, we controlled for the length of the sentences across conditions, grammatical number and gender of the CL where applicable, as we will describe in the next section.<sup>26,27</sup>

### 15.3.3 Stimuli

#### 15.3.3.1 Stimulus design

The stimuli, that is, sentences evaluated in the experiment, were designed with the matrix verbs listed in Section 15.3.1 in the present tense and supplemented with infinitives which had pronominal and reflexive CLs as complements.<sup>28</sup> In contrast to matrix verbs, which were the independent variable of the greatest interest to us, infinitives were not treated as variables. Both the matrix predicate verbs and the infinitives were extracted from hrWaC v2.2 using CQL queries and the Frequency function. Whenever we were unable to find infinitives with a given pronominal or reflexive CL complement in a given case in hrWaC v2.2, we turned to the Institute of Croatian Language and Linguistics where an e-dictionary of verb valencies is being developed (cf. Birtić et al. 2017). We paid a lot of attention to the creation of stimulus sentences, as it is well known that

[...] an informant's response to an individual sentence may be affected by many different lexical, syntactic, semantic and pragmatic factors, together

<sup>23</sup>For more information on the importance of this factor see Section 11.3.4.

<sup>24</sup>For more information on the importance of this factor see Section 11.3.3.

<sup>25</sup>As the third person singular feminine accusative pronominal CL, we had both *ju* and *je* forms in our stimuli, because some speakers prefer *ju* while others prefer *je*. In this way we tried to avoid their personal preferences based on their dialects or idiolects affecting the rating of our stimuli. For the status of the third person singular accusative feminine CL *ju* and *je* in standard Croatian see Section 6.3.1 and for its status in Štokavian dialects see Section 7.4.1.1.

<sup>26</sup>This type of design on the one hand allows the researcher to test the effect of each independent variable separately and on the other it allows the researcher to look at possible interactions between the independent variables. For these reasons it is more cost-effective than conducting various separate experiments on each independent variable. In addition, using this type of design also allows the researcher to determine if the effect of one independent variable depends on the value of another independent variable (cf. Abbuhl et al. 2013: 121).

<sup>27</sup>A design like ours with two or more independent variables (factors) is called a factorial design. One of the main advantages of such designs is that they help control for unintended differences between the conditions (Stowe & Kaan 2006: 14).

<sup>28</sup>We prepared core elements of target sentences as Cowart (1997: 50) recommends.

## 15 Experimental study on clitic climbing out of infinitive complements

with an assortment of extralinguistic influences that become haphazardly associated with linguistic materials and structures. (Cowart 1997: 46)

To deal with all confounding factors, scholars recommend paradigm-like token sets as a safe strategy. This ensures that all the abovementioned unwanted and hazardous factors are uniformly spread across all tested sentences. This in turn guarantees that the differences in ratings can be attributed exclusively to the phenomenon under investigation (Cowart 1997: 13, 47, 52). Furthermore, we created multiple lexical encodings of each condition to minimise the effects of particular lexical items on the results, as recommended in the methodological literature (cf. Schütze & Sprouse 2013: 39).<sup>29</sup>

We created seven experimental lists, each containing only one matrix predicate subtype (raising, simple subject control, reflexive subject control, etc.). The structure of stimuli without CC (henceforth noCC stimuli) is presented in Table 15.7.<sup>30,31</sup>

Table 15.7: Structure of noCC stimuli

Item	Position					
	Adverb	CL <sub>1</sub>	Matrix.PRS <sub>1</sub>	Infinitive <sub>2</sub>	CL <sub>2</sub>	PP Complement/Adjunct
I15.1	Posve	NA	prestajemo <sub>1</sub>	prigovarati <sub>2</sub>	<i>mu</i> <sub>2</sub>	zbog lošeg društva.
I15.2	Čak	NA	nastojim <sub>1</sub>	pozivati <sub>2</sub>	<i>ga</i> <sub>2</sub>	na mjesečne sastanke.
I15.3	Zaista	<i>se</i> <sub>1</sub>	ustručavam <sub>1</sub>	ugađati <sub>2</sub>	<i>si</i> <sub>2</sub>	u svakom pogledu.
I15.4	Oduvijek	<i>mu</i> <sub>1</sub>	dozvoljavaš <sub>1</sub>	skrivati <sub>2</sub>	<i>se</i> <sub>2</sub>	od znatiželjnih pogleda.
I15.5	Vidno	<i>je</i> <sub>1</sub>	požurujem <sub>1</sub>	očitovati <sub>2</sub>	<i>se</i> <sub>2</sub>	o iznesenim prijedlozima.

For each experimental list, eight different matrix verbs (see position Matrix.PRS<sub>1</sub>) were used multiple times.<sup>32</sup> Since two of our independent variables

<sup>29</sup>We avoid the term “lexicalization” used by Schütze & Sprouse (2013: 39) due to its ambiguity.

<sup>30</sup>More examples of noCC stimulus sentences for each matrix predicate type can be found in the Appendix A.

<sup>31</sup>Here are the item translations from Table 15.7:

(I15.1) ‘We are entirely stopping complaining about the bad company he keeps.’

(I15.2) ‘I am even trying to invite him to the monthly meetings.’

(I15.3) ‘I truly hesitate to please myself in every way.’

(I15.4) ‘You have been allowing him to hide from curious glances since always.’

(I15.5) ‘I visibly hurry her to voice her opinion on the presented suggestions.’

<sup>32</sup>Lists with the eight verbs for each of the seven experimental lists can be found in Tables 15.1–15.5.

are critical (infinitive) CL type (pronominal vs reflexive) and case (dative vs accusative), in each of the seven experimental lists we had:

- eight sentences with dative pronominal CLs,
- eight sentences with accusative pronominal CLs,
- eight sentences with the REFL<sub>2ND</sub> CL *si*,
- eight sentences the REFL<sub>2ND</sub> CL *se* and
- sixteen sentences with the REFL<sub>LEX</sub> CL *se* (see position CL<sub>2</sub>).

As governors, we used eight different infinitives (see Position Infinitive<sub>2</sub>) per critical CL subtype.<sup>33</sup> Each sentence on an experimental list had a unique adverb at the beginning and a unique (prepositional) complement/adjunct at the end (see Positions Adverb and PP Complement/Adjunct). The first CL, which is generated by the matrix predicate, was not present on the first two experimental lists, on which we presented the raising and simple subject control predicates (see position CL<sub>1</sub>).<sup>34</sup>

As we explain in the section below, each sentence was rated in its CC and its noCC version. In the CC version, the critical CL (CL<sub>2</sub>) climbs and takes 2P directly following the adverb. If the matrix predicate has its own CL, the matrix CL (CL<sub>1</sub>) and critical CL (CL<sub>2</sub>) clusterise. CL<sub>1</sub> appears in the cluster first, and is followed by CL<sub>2</sub>.<sup>35</sup>

We now briefly present the stimuli. The comparison of stimuli is based on the different CL<sub>2</sub> subtypes. In the first two items, I16.1 and I16.2 presented in Table 15.7, the infinitive governs the third person pronominal CL in the dative and accusative, while in the second two items, I16.3 and I16.4, the infinitive governs

<sup>33</sup>In other words, due to the recommendation to use different lexical materialisations as mentioned above, each of the eight accusative pronominal CLs was governed by a different infinitive. The same applies to dative pronominal CLs and to REFL<sub>2ND</sub> and REFL<sub>LEX</sub> CLs, with the exception that the latter CL depended on 16 different infinitives.

<sup>34</sup>As we already pointed out in Section 15.3.2, unlike reflexive subject control and object control predicates, raising and simple subject control predicates do not have their own CLs, as will become obvious from examples (4a) and (6a).

<sup>35</sup>The exceptions to that CL cluster sequence are examples in which CL<sub>1</sub> is reflexive (for instance in the case of reflexive subject control predicates or object control predicates with a REFL<sub>2ND</sub> CL controller). In that case the pronominal CL<sub>2</sub> appears in the cluster first, and is then followed by the reflexive CL<sub>1</sub>. This was done in order to follow the patterns of CL ordering in a cluster – for the relative order of CLs in the CL cluster in standard BCS varieties see Section 2.4.2.1. In sentences with two reflexive CLs, the order in the cluster was as usual: CL<sub>1</sub> followed by CL<sub>2</sub>.

15 *Experimental study on clitic climbing out of infinitive complements*

the REFL<sub>2ND</sub> CLs *si* and *se*. In the last item, I16.5, the REFL<sub>LEX</sub> CL *se* is the critical CL. For presentation purposes we deliberately chose stimuli with matrix predicates which belong to different types to show how some of them have their own CL<sub>1</sub> (see items I15.1 and I15.2), while others do not (see items I15.3–I15.5).

Table 15.8: Comparison of noCC stimuli across seven experimental lists

Item	Position							
	Adverb	CL <sub>1</sub>	Matrix.PRS <sub>1</sub>	Infinitive <sub>2</sub>	CL <sub>2</sub>	PP	Complement/Adjunct	
I15.6	Stoga	NA	krećem	pozivati	ga	na	mjesečne	sastanke.
I15.7	Čak	NA	nastojim	pozivati	ga	na	mjesečne	sastanke.
I15.8	Nekako	se	ustručavam	pozivati	ga	na	mjesečne	sastanke.
I15.9	Uporno	mi	naređuju	pozivati	ga	na	mjesečne	sastanke.
I15.10	Javno	ih	obvezujem	pozivati	ga	na	mjesečne	sastanke.
I15.11	Ujedno	si	dopuštam	pozivati	ga	na	mjesečne	sastanke.
I15.12	Nevoljko	se	prisiljavam	pozivati	ga	na	mjesečne	sastanke.

As can be seen in Table 15.8 (compare items I15.6–I15.12) and Appendix A, sentences are designed in such a way that they differ only as to adverbs and matrix predicates (and consequently also as to matrix CLs if available), and at the same time they contain the same infinitives, critical CLs and PP complements/adjuncts.<sup>36</sup>

As we already pointed out in Section 3.3.3.3 it is important for participants to be exposed to polarised sentences, otherwise they will start to evaluate acceptable sentences as unacceptable. Therefore, in each experiment, besides the 48 target sentences, participants had to evaluate 48 target-like syntactically and morphologically ill-formed sentences. Those sentences were deliberately constructed via disruption of obvious grammatical rules unrelated to our study. Furthermore, the stimuli must be counterbalanced.

<sup>36</sup>Here are the item translations from Table 15.8:

(I15.6) ‘Therefore, I am starting to invite him to the monthly meetings.’

(I15.7) ‘I am even trying to invite him to the monthly meetings.’

(I15.8) ‘We kind of hesitate to invite him to the monthly meetings.’

(I15.9) ‘They are persistently ordering me to invite him to the monthly meetings.’

(I15.10) ‘I publicly oblige them to invite him to the monthly meetings.’

(I15.11) ‘At the same time I am allowing myself to invite him to the monthly meetings.’

(I15.12) ‘We begrudgingly force ourselves to invite him to the monthly meetings.’

Counterbalancing aims to distribute both the idiosyncratic and the systematic structural effects that arise in a single sentence across the whole experiment in such a way that the systematic effects can be reliably discriminated from the background blur of idiosyncratic effects. (Cowart 1997: 93)

The first rule of counterbalancing is that a participant is never to see a sentence twice, i.e. s/he is never exposed to more than one member of a token set (Cowart 1997: 50f, 93).<sup>37,38</sup> Latin square design helped us fulfil this requirement, i.e., it enabled us to distribute items across participants' lists properly (cf. Stowe & Kaan 2006: 49, Abbuhl et al. 2013: 121). In our experiment, each list contains one sentence with each of the conditions, and no list contains more than one version of each sentence. Moreover, the application of Latin square design means that for each sentence, half of the participants saw a noCC version (like the one presented in (1a)), while the other half saw a CC version (like the one presented in (1b)), and that each participant saw both CC and noCC sentences.

- (1) a. Potpuno  $ih_1$  primoravam<sub>1</sub> angažirati<sub>2</sub>  $je_2$  u političkoj  
 categorically them.ACC compel.1PRS hire.INF her.ACC in political  
 kampanji.  
 campaign
- b. Potpuno  $ih_1$   $je_2$  primoravam<sub>1</sub> angažirati<sub>2</sub> u političkoj  
 categorically them.ACC her.ACC compel.1PRS hire.INF in political  
 kampanji.  
 campaign
- 'I am categorically compelling them to hire her in the political campaign.'

The second rule of counterbalancing is that one should obtain a subject's judgments on all the relevant factor combinations (cf. Cowart 1997: 50, 93). The third rule of counterbalancing is that every sentence in every token set should be judged by a participant (Cowart 1997: 93).

Since participants can form implicit hypotheses on the aim of the experiment, which could potentially distort or affect their judgments (cf. Cowart 1997: 51f, 93f), syntactically and morphologically well- and ill-formed fillers were included

<sup>37</sup>This recommendation should be followed because the second (or any further) encounter with the same sentence will be influenced by the first one; this danger exists even in the case of similar sentences (cf. Cowart 1997: 50).

<sup>38</sup>In our case a token set is one of the 48 target sentences structured as in Tables 15.7 and 15.8. A member of a token set is the CC or the noCC version of a particular sentence.

in the experiment. Schütze & Sprouse (2013: 39) name two more roles of fillers in addition to this important one. First, they can help us to ensure that all the possible responses are used about equally often. Second, they can be used to collect data for other research questions. For the latter reason, we used target sentences from the research of Dóra Vuk on agreement (80 grammatically well-formed and 65 syntactically and morphologically ill-formed sentences) as fillers.<sup>39</sup> Since Vuk could not provide us with enough filler sentences, they were supplemented by an additional 20 syntactically and morphologically well-formed sentences from hrWaC and an additional 35 syntactically and morphologically ill-formed sentences, which had the structure of her target sentences (see Cowart's 1997: 52 recommendations). The latter were obtained via permutation of sentences attested in the aforementioned corpus. The filler-stimulus ratio was 2:1. Below are examples of syntactically and morphologically well-formed (2) and ill-formed (3) filler sentences.

- (2) Sestra i mama su slične.  
 sister and mother be.3PL similar  
 'Sister and mother are similar.'
- (3) \* Ruka i noga su mu podignuta.  
 arm and leg be.3PL him.DAT elevated  
 Intended: 'His arm and leg were elevated.'

All target sentences in the experiment have similar length in order to control for this extraneous variable, so that differences in judgments can be attributed solely to differences in structure (cf. Cowart 1997: 45).

In order to avoid effects of fatigue, boredom and response strategies which participants develop during the experiment, the order of sentences presented to participants was randomised (cf. Cowart 1997: 51, 94, Krug & Sell 2013: 82). Furthermore, randomisation is important because the preceding sentence can influence the judgment of the following sentence (cf. Cowart 1997: 51f). It was carried out with the algorithm of the software we used – OpenSesame version 3.1.9 *Jazzy James* (Mathôt et al. 2012).<sup>40</sup> The order of stimulus presentation was

<sup>39</sup>Dóra Vuk's PhD thesis *Kongruenz in der kroatischen Herkunftssprache in Ungarn und Österreich* 'Agreement in Croatian heritage language in Hungary and Austria' was financially supported by the Graduate School for East and Southeast European Studies. In her research she concentrated on gender agreement in conjoined phrases and its realisation in adjectives in nominal predicate and in past participle. Her sentences were also constructed for use in an acceptability judgment task.

<sup>40</sup>For more information on OpenSesame visit <http://osdoc.cogsci.nl/>.

shuffled in the experimental part and in the practise session (see Cowart's (1997: 96) recommendations for randomisation).

The greatest advantages of computer-based acceptability judgment tasks are that two measures can be taken at the same time (reaction time and acceptability rating) and that participants cannot go back and change previous answers. However, as in all experiments, there is the problem that only a few members of a speech community are willing to participate in such studies since it means that they have to come to a certain place at a certain time.<sup>41</sup> In other words, it is harder to bring the participants to the lab than to give them a paper questionnaire that they can fill in on the spot.

### 15.3.3.2 Ecological validity of stimuli in our study

We tried to improve the ecological validity of our stimuli as far as possible.<sup>42</sup> When constructing the stimuli, we used corpora to make them sound more natural. For instance, we always searched them for adverbials or (prepositional) complements/adjuncts often appearing to the right of the infinitive used in our stimuli: see position PP Complement/Adjunct in Table 15.8, examples (I15.6)–(I15.12). Further, we looked for the most frequent adverbs to appear left of matrix verbs – see position Adverb in example (I15.6)–(I15.12). Additionally, in order to be sure that the adverbs at the beginning of the sentence can serve as hosts for CLs, we checked how well the chosen adverbs were attested with pronominal CLs such as *ga*. All adverbs which had less than 100 hits with pronominal CLs in the whole hrWaC were replaced with adverbs more likely to appear as hosts.

Some may object that the object control matrix verbs (see Tables 15.3–15.5) chosen for the study do not sound natural with the infinitive and they might prefer the *da*<sub>2</sub>-construction instead. However, we emphasise that the stimuli were constructed exclusively with object control verbs which were attested with infinitive complements in hrWaC.<sup>43</sup>

<sup>41</sup>At the time when we were collecting these data, online solutions were still being developed, and they were neither as wide-spread nor as well-tested as they are today: the pandemics gave these solutions an additional boost. Currently, reaction time can quite reliably be collected online, and there is even a way to present participants with a reward; for more information, see Filipović Đurđević (2021). However, one advantage of in-person testing is control over testing conditions. For us, in-person testing was also crucial because we could check whether the participants were really speakers of the Neo-Štokavian dialect.

<sup>42</sup>Ecological validity is a problem of experimental data; for more information see Sections 3.2.1 and 3.2.2.

<sup>43</sup>We compared our 16 object control verbs with the verbs listed in Gnjatović & Matasović (2013), a study on verbs with obligatory control in Croatian. Only three of them (*savjetovati* 'advise', *preporučivati* 'recommend', *požurivati* 'hurry') were not mentioned in this article.

## 15 Experimental study on clitic climbing out of infinitive complements

Moreover, *naređivati si* ‘order yourself/give yourself a command’ may be considered objectionable and odd-sounding by some. Indeed, not all of the eight object control matrix predicates with the REFL<sub>2ND</sub> *si* (see Table 15.4) selected for one of the seven experimental lists are completely satisfactory. However, it was not possible to find eight more appropriate representative verbs with the REFL<sub>2ND</sub> *si* controller that were attested with infinitive complements. The verbs chosen were therefore a compromise that enabled us to fully cover the experimental design.

As a last step, we conducted a pilot study where we asked native speakers to evaluate our target sentences. The results of their feedback were used to improve stimuli to sound as natural as possible.

We are aware that the constructed stimuli can never achieve the ecological validity of data produced spontaneously. However, the abovementioned steps, i.e. the double check in the corpus, which provided us with model sentences for the acceptability experiment and the pilot study, allow us to reject the claim that the constructed examples are entirely artificial. In other words, they are likely to be similar to sentences which appear in real-life situations.

### 15.3.4 Participants

Methodological literature recommends avoiding linguists as potential participants for several reasons.<sup>44</sup> First of all, they have been exposed to a great deal of language contact and therefore may have different intuitions than non-linguists. On the one hand, their linguistic knowledge may lead them to under- or over-report on marginal structures or features (cf. Krug & Sell 2013: 78). On the other hand, since they are probably aware of the theoretical impact of their judgments, they may be consciously or subconsciously biased to judge in accordance with their theoretical viewpoints (cf. Ferreira 2005: 372, Wasow & Arnold 2005: 1483, Gibson & Fedorenko 2010: 233, Gibson & Fedorenko 2013: 88f, 98f). Apart from linguists, we decided to exclude language teachers and all students of languages since they can demonstrate rather prescriptive attitudes and may rely heavily

---

<sup>44</sup>It must be said that contrary to the abovementioned reasons against using linguists as participants, some argue that professional linguists’ expert knowledge may increase their reliability and perhaps also their sensitivity, since they are able to detect fine-grained distinctions which inexperienced participants simply cannot perceive (see Newmeyer 1983: 61, 66, Newmeyer 2007: 397, Fanselow 2007: 354, Devitt 2006: 497–500, Devitt 2010: 860f). There are several contradictory studies regarding this issue. On the one hand Spencer (1973), Gordon & Hendrick (1997), and Dąbrowska (2010) point out that there are differences in ratings between linguist and non-linguist populations, while on the other hand Sprouse & Almeida (2012) and Sprouse et al. (2013) found strong agreement in ratings by linguists and non-linguists. For more information on this problem see Section 3.1.



Table 15.9: Higher education institutions attended by our participants

Higher education institutions	<i>N</i>	Higher education institutions	<i>N</i>
Faculty of Agriculture Faculty of Forestry	55	Academy of Music	24
Faculty of Chemical Engineering and Technology Faculty of Food Technology and Biotechnology Faculty of Science Department of Biology Department of Physics Department of Ecology, Agriculture and Aquaculture School of Medicine	53	Faculty of Kinesiology	23
Faculty of Electrical Engineering Faculty of Electrical Engineering and Computing Faculty of Mechanical Engineering and Naval Architecture	50	Faculty of Economics and Business Faculty of Economics Zagreb School of Economics and Management	21
Faculty of Transport and Traffic Sci- ences	33	Faculty of Humanities and Social Sciences	18
Polytechnic Lavoslav Ružička Vuko- var Polytechnic of Požega Polytechnic of Šibenik Polytechnic of Zagreb	6	Faculty of Law Faculty of Political Science	2
Catholic Faculty of Theology	50	Faculty of Tourism and Hospitality Management	1

on the notion of a narrowly defined standard language usage (cf. Krug & Sell 2013: 78). Moreover, we wanted to control for dialect, since CLs behave differently in the Čakavian and Kajkavian dialects. Therefore we chose only speakers of Neo-Štokavian dialects (cf. Cowart 1997: 45).<sup>45</sup>

The experiment was conducted in three university cities: Zagreb, Split, and Osijek. Although the dialect of native speakers of Croatian from Zagreb is not

<sup>45</sup>For more information on CLs in Štokavian dialects see Chapter 7.

## 15 *Experimental study on clitic climbing out of infinitive complements*

Neo-Štokavian, the research was conducted in Zagreb because it is the city with the biggest and oldest university in Croatia. As such it is attractive to students from other cities and regions. Therefore, many Neo-Štokavian speakers can be found in Zagreb. Unlike Zagreb, Split and Osijek are cities where Neo-Štokavian is spoken and they were chosen precisely for that reason.

The average age of our speakers is 21.5 years. We recruited participants from 30 different higher education institutions located in seven different cities in Croatia.<sup>46</sup> For details see Table 15.9 below.

### 15.3.5 Recruiting participants

Since relying on volunteers only turned out to be inefficient (in particular in terms of time), participants were rewarded with cinema coupons for participating in the experiment. Information about the experiment (time, place, mode of procedure) was distributed among students by university teachers, on Facebook study groups, online learning platforms, official faculty web sites, flyers and official faculty email addresses.<sup>47</sup> Participants were able to schedule their appointment and to book their place via Google spreadsheets in which they signed up for the experiment with a pseudonym.

The yes/no task requires forty participants to reach 80% coverage (statistical power) (Schütze & Sprouse 2013: 40). This was found to be the minimum number of participants needed to achieve the given statistical power, assuming that each participant provides only one response per condition. In our case, participants provided us with multiple responses per condition. However, in order to stay on the safe side, we kept the minimum recommended sample size.

### 15.3.6 Procedure

As recommended, we followed basic ethical guidelines. We provided participants with general information on the purpose of the study.<sup>48</sup> We also obtained writ-

---

<sup>46</sup> Although the experiment was conducted in three university cities, the participants attended higher education institutions located in seven different cities in Croatia. Some participants were recruited while visiting their friends in one of the university cities. Others were recruited during the Christmas holidays in their home villages Rokovci and Andrijaševci where one of the authors also stayed in winter 2017.

<sup>47</sup> This would not be possible without the help of many enthusiastic university teachers and administrators, who are all listed in the Acknowledgments.

<sup>48</sup> This does not mean that we told participants that we are investigating CLs and word order. We simply informed them that we wanted to investigate certain structures in Croatian and that their native speaker intuition was an invaluable tool for us when distinguishing between acceptable and unacceptable structures.

ten consent to using their data, i.e. we were assured that their participation was voluntary. We guaranteed our participants complete anonymity and gave them our contact details. This is so that participants can have access to the research findings (cf. Krug & Sell 2013: 71).

Prior to conducting the experiment we explicitly informed the participants that the data would be used for scientific purposes only. We emphasized that it was important for the data to be reliable and that the experiment required a great deal of concentration.

Since most of the participants had never taken part in acceptability judgment tasks, they had to be familiarised with the method. Therefore, before collecting real data, the participants had to complete a training session in which they rated 24 sentences in order to prepare them for the task in the experiment.<sup>49</sup> The instruction as to how to complete the task were provided in writing and orally, and repeated twice: before showing of the training set and of the experimental set. Below are the written instructions in Croatian.<sup>50</sup>

Na zaslonu će se pojaviti niz riječi u obliku rečenice. Vaš je zadatak da procijenite je li dani niz riječi, tj. dana rečenica prihvatljiva kao rečenica hrvatskoga jezika.

---

<sup>49</sup>In the training session participants rated six target sentences, six syntactically and morphologically ill-formed target-like sentences, six filler sentences, and six syntactically and morphologically ill-formed filler-like sentences. As indicated in Section 15.4.1, all sentences from the practise session were removed from the analysis.

<sup>50</sup>A string of words in the form of a sentence will appear on the screen. Your task is to determine whether the given string of words, i.e. the given sentence, is acceptable as a Croatian-language sentence.

It is important for you to know that this is not a formal test of Croatian language knowledge. It is exceptionally important to us that you answer in accordance with your personal sense of language.

We would ask you to read each sentence carefully, but not to spend too much time thinking about it, instead you should answer by following your first impulse.

If you consider the sentence to be acceptable, press the left mouse button with your pointer finger. If you consider the sentence to be unacceptable in your language, press the right mouse button with your middle finger. Therefore, hold the mouse as you usually do when working on a computer.

Try to answer as quickly and as accurately as possible.

At the beginning you will take part in a short exercise. If you have any questions you may ask them after the exercise, i.e. before the experiment.

Press any key to start the exercise.

\*\*\*\*\*

This was the exercise. Press any key to start the experiment.

## 15 *Experimental study on clitic climbing out of infinitive complements*

Važno je da znate da ovo nije formalni test znanja hrvatskoga jezika. Nama je iznimno važno da odgovorite u skladu s vlastitim jezičnim osjećajem.

Molimo Vas da svaku rečenicu pažljivo pročitate, ali da ne provedete previše vremena razmišljajući o njoj, nego da odgovorite prateći svoj prvi impuls.

Smatrate li da je rečenica prihvatljiva, kažiprstom pritisnite lijevu tipku miša. Smatrate li da dana rečenica nije prihvatljiva u Vašemu jeziku, srednjim prstom pritisnite desnu tipku miša. Miš, dakle, držite onako kako ga inače držite kad radite na računalu.

Pokušajte odgovarati što brže i što točnije.

Na početku ćete imati kratku vježbu. Budete li imali pitanja, možete ih postaviti poslije vježbe, odnosno prije eksperimenta.

Za početak vježbe pritisnite bilo koju tipku.

\*\*\*\*\*

Ovo je bila vježba. Za početak eksperimenta pritisnite bilo koju tipku.

Since there was a danger that the participants' answers would be a compromise between actual language usage and the socially desired answers, we decided to make it explicit in the instruction that this was not a formal test of Croatian language knowledge.<sup>51</sup> That means that there were no desirable answers and no wrong answers per se (cf. Krug & Sell 2013: 75, Hoffmann 2013: 103). Additionally, since at this point we were not interested in possible diaphasic variation, we instructed orally participants to rate whether each sentence presented could be said or written by a native Croatian speaker, i.e., considering the Croatian language as a whole (not comparing the sentences in the experiment with a specific Croatian variety).<sup>52</sup>

Each trial began with the presentation of the fixation cross in the centre of the screen for 2000 ms in order to draw the participant's eyes to neutral position.

<sup>51</sup>The danger that participants' answers would be a compromise between actual language usage and the socially desired answers is often the case when vernacular, non-standard forms and usages are stigmatised as a hallmark of uneducated people.

<sup>52</sup>Participants were told this orally (in addition to reading it in the instructions themselves) to help them familiarize themselves with the task and to make sure that they were fully aware what the task entailed. The basic idea was that the participants should understand that they would be making judgments based on their own experience of the language (which arises from their contact, oral and written, with other native speakers), not the formal knowledge of grammar taught in schools. It was crucial to emphasise this part since sometimes people use constructions not approved in the norm. We wanted them to know that we would not stigmatise any language use.

Next, a sentence was presented in the centre of the screen. The presentation time for each sentence was until the participant's response or time-out, which was set to 8000 ms. The participants had to give their answer by pressing the left mouse key with their index finger to judge a sentence as acceptable or the right mouse key with their middle finger to judge a sentence as unacceptable.<sup>53</sup> The response and response time were recorded automatically. If a participant did not make a response within 8000 ms, the trial was aborted. As already stated, each participant received 24 practice items before the experimental session started. In the experimental session each participant rated 296 sentences.

The participants rated the sentences in a quiet room (classroom or office) at respective faculties on four laptops provided by a member of the study team, under the supervision of that member. Some faculties provided us with additional laptops so that we could conduct the experiment faster (test more than 4 participants at the same time) and some even allowed us to install OpenSesame and experimental files on their computers and to use their computer labs. Each acceptability judgment experiment session took about 30 minutes per participant.<sup>54</sup>

## 15.4 Data analysis

### 15.4.1 Data preprocessing

This stage involved two steps: identification of participants who did not perform the task correctly and identification of extreme outliers in the responses.

In the first step, we removed all the data collected during the practice sessions and all the filler sentences from our dataset. Next, we focused on obviously syntactically and morphologically ill-formed sentences (unrelated to our study design) and we analysed by-participant accuracy for these sentences.<sup>55</sup> Participants who accepted obviously grammatically ill-formed stimuli, and did so multiple times, were most likely not paying attention to the task. Therefore, participants who accepted more than 25% of the sentences from the subset of clearly unacceptable sentences were excluded from further analyses. This resulted in the exclusion of 18 participants. After this, the subset of clearly unacceptable sentences

---

<sup>53</sup>This was done to ensure that the dominant motor action is mapped to the yes response, to reduce the variation in response time that can be attributed to the execution of the motoric component of the response action.

<sup>54</sup>Although it seems rather long, Rosenbach (2013: 283) claims that 30–45 minutes is the ideal time span for holding a subject's full concentration without tiring them out.

<sup>55</sup>We were not able to rely on total accuracy, as the acceptability of our target sentences was itself the subject of our inquiry. Therefore, it was impossible to use their acceptability as the criterion for engagement in the task.

## 15 Experimental study on clitic climbing out of infinitive complements

was removed from the set, and all the analyses were conducted on the main set of target sentences described in Section 15.3.3.

We also inspected the distribution of timeout data, i.e., situations in which participants failed to respond by either “yes” or “no” to the sentence in question. We did so to make sure that the sentences which were not considered acceptable were not predominantly those that went unrated, i.e. those that were timed out instead of receiving a “no” response. The data revealed that this was not the case – only around 1% of responses were timeouts and they were evenly distributed across conditions. We also conducted parallel statistical analyses with and without timeout data and found no substantial differences between them. Therefore, we decided to keep the timeout data and treat them as lack of acceptance.

In the next step, we inspected the reaction time (RT) distribution. Before analysing RT, we removed five data points that were below 700ms and clearly outliers. Finally, the RT data were log-transformed to approach normality (as suggested by Baayen & Milin 2010).

### 15.4.2 Statistical analysis

The data were analysed in the R statistical environment (R Core Team 2017) using the *LMER4* (Bates, Mächler, et al. 2015), *LMERTTEST* (Kuznetsova et al. 2017) and *LSMEANS* (Lenth 2016) packages. We applied mixed-effects regression with participants and sentence endings as random variables.<sup>56</sup> This statistical method has become the golden standard in psycholinguistic research (Baayen 2012, Baayen et al. 2008, Jaeger 2008).

In this analysis the random effects of participants and items are taken into account simultaneously. Unlike the fixed effects which are tested deliberately by the researcher, participants and items are considered to be the source of random effects. These are the differences that originate in factors that are beyond control in the current experiment, i.e. overall processing speed among participants, differences among language items that are as yet unknown and so on.

By applying mixed effects analysis, the fixed effects of the variables which were manipulated (or introduced in some other way) by the researcher can be generalised beyond the set of stimuli presented in the experiment and beyond the speakers who participated. In other words, these effects can be generalised to other language stimuli of the same kind and to other speakers who belong to the same population, i.e. healthy speakers of the given language.

---

<sup>56</sup>These included the matrix verb followed by the rest of the sentence (see positions *Matrix.PRS*<sub>1</sub>, *Infinitive*<sub>2</sub>, *CL*<sub>2</sub>, *PP Complement/Adjunct* in Table 15.8 examples (I15.6)–(I15.12), which were kept constant while the *CL* type and *CL* position were manipulated.

Wherever possible, we included random slopes of the variables that were tested as fixed effects, as suggested by Barr et al. (2013). However, when convergence of the model was not achievable, we included random slopes one by one as suggested by Barr (2013) and tested whether their inclusion in the model was justified by the data (cf. Bates, Kliegl, Vasishth & Baayen 2015, Matuschek et al. 2017).

In each analysis we fitted two parallel models: one for acceptance and one for reaction time. For the models that were fitted to the binomial variable of acceptance (1 = accept; 0 = reject), we used binomial distribution as the underlying functional form and we fitted the model by using generalised linear mixed effects regression (GLMER). For the models that were fitted to reaction time, we used Gaussian distribution as the appropriate underlying functional form and we fitted the model using the LMER function.

## 15.5 Results: Regression model for acceptance rates

In Table 15.10–15.14 and Figure 15.1–15.3 we report the results of the acceptance rate analysis. The pattern of results reveals a three-way interaction of:

- CC (CC, noCC),
- matrix verb type (raising, simple subject control, reflexive subject control with the REFL<sub>LEX</sub> CL *se*, object control with the pronominal CL controller in the dative, object control with the pronominal CL controller in the accusative, object control with the REFL<sub>2ND</sub> CL *si* controller, and object control with the REFL<sub>2ND</sub> CL *se* controller),
- and infinitive CL type (pronominal, REFL<sub>2ND</sub>, REFL<sub>LEX</sub>).

Detailed data from the generalised mixed effects regression model are presented in Tables 15.10–15.14.<sup>57</sup> Those interested in the tables who cannot easily follow their content can use the explanations in this section as a guide. Those who are not interested in the details of the generalised mixed effects regression model can skip the table and go directly to following subsections, in which we present the results according to the order of the research questions presented in Section 15.2.

---

<sup>57</sup>The model formula is: Acceptance ~ CC \* Matrix verb \* Infinitive CL type + (1|Participant) + (1|Item); for explanation of statistical measures in Tables 15.10–15.14 see Appendix B.

15 *Experimental study on clitic climbing out of infinitive complements*

Table 15.10: Random effects from generalised mixed effects regression model fitted to acceptance data (1 = acceptable; 0 = unacceptable).

	Var.	SD
participants (intercept adjustments)	1.410	1.187
items (intercept adjustments)	0.631	0.794

We describe the observed results in more detail by referring to the fixed effects in Tables 15.11–15.14. The first row presents the intercept of the model, and in this case it corresponds to noCC sentences with raising predicates whose infinitive complement governs pronominal CLs (leftmost bar in Figure 15.1). In comparison, there is a statistically significant rise in acceptance of the same sentences with CC (the estimated coefficient is above zero and statistically significant; see row 2 in Table 15.11). In the next six rows (row 3–row 8) the noCC sentences with raising predicates whose infinitive complement governs a pronominal CL (intercept) are compared to the same sentences (pronominal CL, noCC) with the six remaining matrix verb types. Based on these rows it can be inferred which of the matrix verb types make the noCC sentences whose infinitive complement governs a pronominal CL more/less acceptable compared to the same sentences with raising predicates (whose infinitive complement governs a pronominal CL, noCC). Compared to the intercept, i.e. raising predicates, the acceptance rate is higher for the same sentence variants in which the matrix verb is a simple subject control verb, reflexive subject control verb with the REFL<sub>LEX</sub> CL *se*, object control verb with a REFL<sub>2ND</sub> CL *si* governor, and object control verb with a REFL<sub>2ND</sub> CL *se* governor. In contrast, the acceptance rate for the sentence variants with an object control matrix verb whose governor is a pronominal CL in the dative or with an object control matrix verb whose governor is a pronominal CL in the accusative is the same as for the sentences with raising-type matrix verbs.

In rows 9 and 10 noCC sentences with raising matrix verbs whose infinitive complement governs a pronominal CL are compared to noCC sentences with the same matrix verb type, but a different critical CL type: the REFL<sub>LEX</sub> CL *se* (row 9) and the REFL<sub>2ND</sub> CL *si/se* (row 10). None of the two differences is significant (see also the three black bars in Figure 15.3A).

Rows 11–16 (Table 15.12) inform us whether the relationship between acceptance rates of noCC and CC sentence variants which was observed for sentences with raising-type predicates whose infinitive complement governs a pronominal CL is the same for the remaining six types of matrix verbs whose infinitive



15.5 Results: Regression model for acceptance rates

Table 15.11: Rows 1–10 of fixed effects from generalised mixed effects regression model fitted to acceptance data (1 = acceptable; 0 = unacceptable).

	Est.	SE	$z$	Pr(>   $z$  )
1 intercept (CC – no; matrix verb – raising; infinitive CL type – pronominal CL)	0.063	0.294	0.214	0.830
2 CC – yes	2.593	0.36	7.194	<0.0001 ***
3 matrix verb – simple subject control	1.096	0.421	2.605	0.009 **
4 matrix verb – reflexive subject control	2.454	0.438	5.608	<0.0001 ***
5 matrix verb – object control with pronominal CL controller in dative	0.450	0.415	1.083	0.279
6 matrix verb – object control with pronominal CL controller in accusative	0.443	0.414	1.069	0.285
7 matrix verb – object control with REFL <sub>2ND</sub> CL <i>si</i> CONTROLLER	0.985	0.417	2.363	0.018 *
8 matrix verb – object control with REFL <sub>2ND</sub> CL <i>se</i> controller	1.738	0.423	4.111	<0.0001 ***
9 infinitive CL type – REFL <sub>LEX</sub> CL <i>se</i>	-0.141	0.330	-0.427	0.669
10 infinitive CL type – REFL <sub>2ND</sub> CL <i>si/se</i>	-0.210	0.331	-0.636	0.525

## 15 Experimental study on clitic climbing out of infinitive complements

complement governs a pronominal CL. Statistical significance associated with a coefficient presented in one of these rows indicates that this relationship differs. Based on the results observed, we can infer that the described ratio is the same for simple subject verb predicates, whereas for all of the remaining matrix verb types, the relationship is different – sentences with CC are not favoured over the corresponding noCC variants (compare leftmost pairs of bars across the panels in Figures 15.1–15.3).

The numbers in rows 17–18 (Table 15.12) allow similar comparisons: here, the relationship between CC and noCC variants of the sentences with raising predicates whose infinitive complements govern a pronominal CL is compared with the same relationship for sentences with raising matrix verbs whose infinitive complements govern REFL<sub>2ND</sub> CLs *se* and *si* and the REFL<sub>LEX</sub> CL *se*. The fact that none of the two coefficients are significant tells us that the type of the complementing CL does not affect the relationship between acceptance of CC and noCC variants in the case of sentences with raising predicates.

Rows 19–24 (Table 15.13) inform us whether the finding which relates noCC sentences with raising predicates whose infinitive governs a pronominal CL and the same sentence type whose infinitive governs the REFL<sub>LEX</sub> CL *se* is affected by the change in predicate type. The results indicate that the absence of difference between noCC raising sentences whose infinitives govern a pronominal CL and those whose infinitives govern the REFL<sub>LEX</sub> CL *se* which we found (i.e. the acceptance rates for the two sentence groups are identical, as indicated in row 9), can be generalised to sentences in which the REFL<sub>LEX</sub> CL *se* is governed by infinitive complements of simple subject control predicates, sentences with object control predicates with a REFL<sub>2ND</sub> CL *si* controller, and object control predicates whose infinitives govern the REFL<sub>2ND</sub> CL *se*. However, this does not hold for the remaining predicate types. In the case of noCC sentences with reflexive subject control matrix predicates, the acceptance rates are lower for sentences with the REFL<sub>LEX</sub> CL *se* as the infinitive complement than for those containing a pronominal CL in the same position. In contrast, in the case of noCC sentences with object control matrix predicates and whose controller is a pronominal CL in the dative or the accusative, the acceptance rates are higher for sentences whose infinitive complement contains the REFL<sub>LEX</sub> CL *se* compared to those that contain a pronominal CL as an infinitive complement.

Similarly, rows 25–30 (Table 15.13) inform us whether the finding which relates noCC sentences with a pronominal CL as the infinitive complement with those whose infinitive governs the REFL<sub>2ND</sub> CL is affected by the change in predicate type. The results show that the acceptance rates for the noCC sentence variants

15.5 Results: Regression model for acceptance rates

Table 15.12: Rows 11–18 of fixed effects from generalised mixed effects regression model fitted to acceptance data (1 = acceptable; 0 = unacceptable).

	Est.	SE	<i>z</i>	Pr(>   <i>z</i>  )
11 CC – yes : matrix verb – simple subject control	–0.549	0.520	–1.055	0.291
12 CC – yes : matrix verb – reflexive subject control	–3.066	0.518	–5.924	<0.0001 ***
13 CC – yes : matrix verb – object control with pronominal CL controller in dative	–4.984	0.500	–9.964	<0.0001 ***
14 CC – yes : matrix verb – object control with pronominal CL controller in accusative	–6.365	0.532	–11.962	<0.0001 ***
15 CC – yes : matrix verb – object control with REFL <sub>2ND</sub> CL <i>si</i> CONTROLLER	–5.684	0.502	–11.333	<0.0001 ***
16 CC – yes : matrix verb – object control with REFL <sub>2ND</sub> CL <i>se</i> controller	–4.160	0.496	–8.385	<0.0001 ***
17 CC – yes : infinitive CL type – REFL <sub>LEX</sub> CL <i>se</i>	0.699	0.521	1.342	0.180
18 CC – yes : infinitive CL type – REFL <sub>2ND</sub> CL <i>si/se</i>	0.417	0.511	0.816	0.414

whose infinitives govern a pronominal CL on the one hand and whose infinitives govern the REFL<sub>2ND</sub> CL on the other hand differ only within sentences with reflexive subject control matrix predicates. Namely, sentences whose infinitives govern the REFL<sub>2ND</sub> CL *si/se* are less acceptable than those with a pronominal CL in the same position.

The remaining rows make analogous comparisons for sentences with CC. Rows 31–36 (Table 15.14) show whether matrix predicate type modifies the relationship between CC sentences whose infinitives govern a pronominal CL on the one hand and those whose infinitives contain the REFL<sub>LEX</sub> CL *se* on the other hand. As observed for sentences with raising matrix predicates (row 17), the two sentence types are equally acceptable in the case of CC sentences with simple subject control matrix predicates and object control predicates whose controller

Table 15.13: Rows 19–30 of fixed effects from generalised mixed effects regression model fitted to acceptance data (1 = acceptable; 0 = unacceptable).

	Est.	SE	z	Pr(>  z )
19 matrix verb – simple subject control : infinitive CL type – REF <sub>LEX</sub> CL <i>se</i>	0.203	0.475	0.427	0.670
20 matrix verb – reflexive subject control : infinitive CL type – REF <sub>LEX</sub> CL <i>se</i>	-1.323	0.487	-2.716	0.007 **
21 matrix verb – object control with pronominal CL controller in dative : infinitive CL type – REF <sub>LEX</sub> CL <i>se</i>	0.849	0.470	1.805	0.071 .
22 matrix verb – object control with pronominal CL controller in accusative : infinitive CL type – REF <sub>LEX</sub> CL <i>se</i>	0.922	0.469	1.966	0.049 *
23 matrix verb – object control with REF <sub>2ND</sub> CL <i>si</i> controller : infinitive CL type – REF <sub>LEX</sub> CL <i>se</i>	0.593	0.475	1.248	0.212
24 matrix verb – object control with REF <sub>2ND</sub> CL <i>se</i> controller : infinitive CL type – REF <sub>LEX</sub> CL <i>se</i>	-0.673	0.475	-1.418	0.156
25 matrix verb – simple subject control : infinitive CL type – REF <sub>2ND</sub> CL <i>si/se</i>	0.257	0.475	0.541	0.588
26 matrix verb – reflexive subject control : infinitive CL type – REF <sub>2ND</sub> CL <i>si/se</i>	-1.079	0.488	-2.209	0.027 *
27 matrix verb – Object control with pronominal CL controller in dative : Infinitive CL type – REF <sub>2ND</sub> CL <i>si/se</i>	0.090	0.468	0.193	0.847
28 matrix verb – object control with pronominal CL controller in accusative : infinitive CL type – REF <sub>2ND</sub> CL <i>si/se</i>	0.079	0.465	0.171	0.864
29 matrix verb – object control with REF <sub>2ND</sub> CL <i>si</i> controller : infinitive CL type – REF <sub>2ND</sub> CL <i>si/se</i>	-0.161	0.470	-0.342	0.732
30 matrix verb – object control with REF <sub>2ND</sub> CL <i>se</i> controller : infinitive CL type – REF <sub>2ND</sub> CL <i>si/se</i>	-0.701	0.475	-1.476	0.140

Table 15.14: Rows 31–42 of fixed effects from generalised mixed effects regression model fitted to acceptance data (1 = acceptable; 0 = unacceptable).

	Est.	SE	z	Pr(>  z )
31 CC – yes : matrix verb – simple subject control : infinitive CL type – REF <sub>LEX</sub> CL <i>se</i>	-0.445	0.747	-0.595	0.552
32 CC – yes : matrix verb – reflexive subject control : Infinitive CL type – REF <sub>LEX</sub> CL <i>se</i>	-3.271	0.728	-4.492	<0.0001 ***
33 CC – yes : matrix verb – object control with pronominal CL controller in dative : infinitive CL type – REF <sub>LEX</sub> CL <i>se</i>	-1.188	0.712	-1.668	0.095 .
34 CC – yes : matrix verb – object control with pronominal CL controller in accusative : infinitive CL type – REF <sub>LEX</sub> CL <i>se</i>	-0.205	0.738	-0.279	0.781
35 CC – yes : matrix verb – object control with REF <sub>2ND</sub> CL <i>si</i> controller : Infinitive CL type-- REF <sub>LEX</sub> CL <i>se</i>	-2.541	0.742	-3.423	0.001 **
36 CC – yes : matrix verb – object control with REF <sub>2ND</sub> CL <i>se</i> controller : Infinitive CL type-- REF <sub>LEX</sub> CL <i>se</i>	-2.334	0.713	-3.271	0.001 **
37 CC – yes : matrix verb – simple subject control : infinitive CL type – REF <sub>2ND</sub> CL <i>si/se</i>	-0.166	0.739	-0.224	0.823
38 CC – yes : matrix verb – reflexive subject control : infinitive CL type – REF <sub>2ND</sub> CL <i>si/se</i>	-3.566	0.727	-4.905	<0.0001 ***
39 CC – yes : matrix verb – object control with pronominal CL controller in dative : infinitive CL type – REF <sub>2ND</sub> CL <i>si/se</i>	-0.778	0.710	-1.096	0.273
40 CC – yes : matrix verb – object control with pronominal CL controller in accusative : infinitive CL type – REF <sub>2ND</sub> CL <i>si/se</i>	0.961	0.730	1.317	0.188
41 CC – yes : matrix verb – object control with REF <sub>2ND</sub> CL <i>si</i> controller : infinitive CL type-- REF <sub>2ND</sub> CL <i>si/se</i>	-0.536	0.711	-0.754	0.451
42 CC – yes : matrix verb – object control with REF <sub>2ND</sub> CL <i>se</i> controller : infinitive CL type-- REF <sub>2ND</sub> CL <i>si/se</i>	-2.273	0.710	-3.200	0.001 **

is a pronominal CL in the accusative. For all the remaining predicate types, CC sentences whose infinitives contain the REFL<sub>LEX</sub> CL *se* are less acceptable than the same sentence type whose infinitives govern a pronominal CL in the same position.

Finally, rows 37–42 (Table 15.14) present similar comparisons between CC sentences whose infinitives govern a pronominal CL on the one hand and the REFL<sub>2ND</sub> CL *si/se* on the other. The results show that the two are different with respect to acceptability only in the case of CC sentences with reflexive subject control matrix predicates and CC sentences with object control matrix predicate whose controller is the REFL<sub>2ND</sub> CL *se*. Within each of the two CC sentence types, those whose infinitives govern the REFL<sub>2ND</sub> CL *si/se* are less acceptable than those whose infinitive governs a pronominal CL in the same position.

Given the complex structure of the results, we organise the presentation according to the research questions presented in Section 15.2.

## 15.6 Discussion

### 15.6.1 Raising, subject and object control predicates

The first research question was formulated at the highest level of generality with the aim of comparing CC and noCC stimuli which contain one of three types of predicates: raising-type predicates, subject control predicates and object control predicates. As can be seen in Figures 15.1–15.3, there are striking differences in the profiles of acceptance for CC and noCC variants for the three predicate types. These differences are particularly pronounced for CC sentences. Sentences with raising matrix predicates show a convincing preference for the acceptance of the CC variants (Figure 15.1: compare examples (4a) and (4b)). Object control predicate sentences, meanwhile, show the opposite pattern – there is a clear preference for acceptance of the noCC sentence variants (Figure 15.3: compare examples (5a) and (5b)).<sup>58</sup>

---

<sup>58</sup>The type of sentence presented in (5b) was accepted in 20% of cases: see first grey column in Panel C1 of Figure 15.3. Therefore we do not mark the example with \*, which is usually used to indicate that a sentence is unacceptable and/or grammatically incorrect. Moreover, in this chapter we avoid using \* for our stimuli since all sentence types were acceptable at least to some degree. Instead, we speak of graded acceptability.

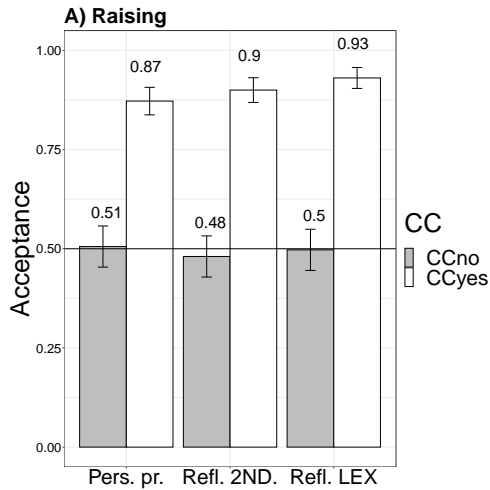


Figure 15.1: Sentences with raising-type predicates

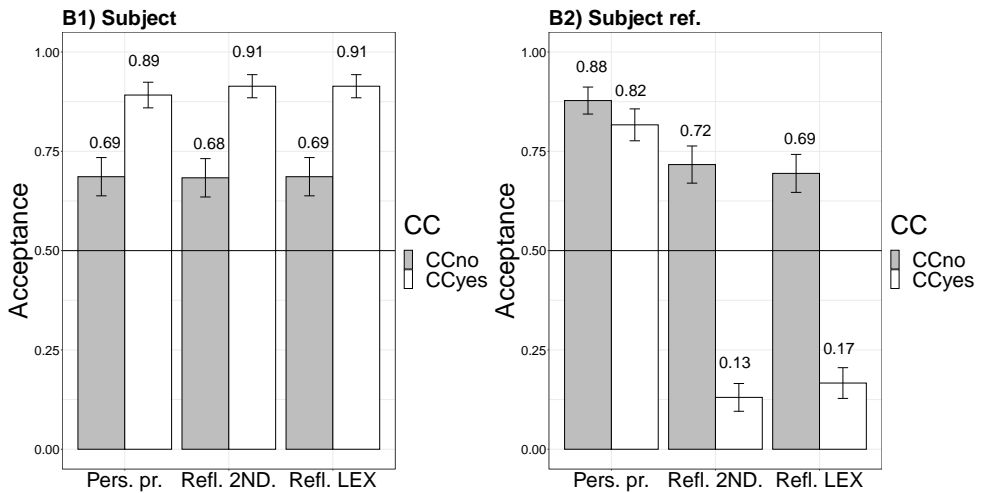


Figure 15.2: Sentences with subject control predicates:  
 Panel B1 (left) – simple, Panel B2 (right) – reflexive

15 Experimental study on clitic climbing out of infinitive complements

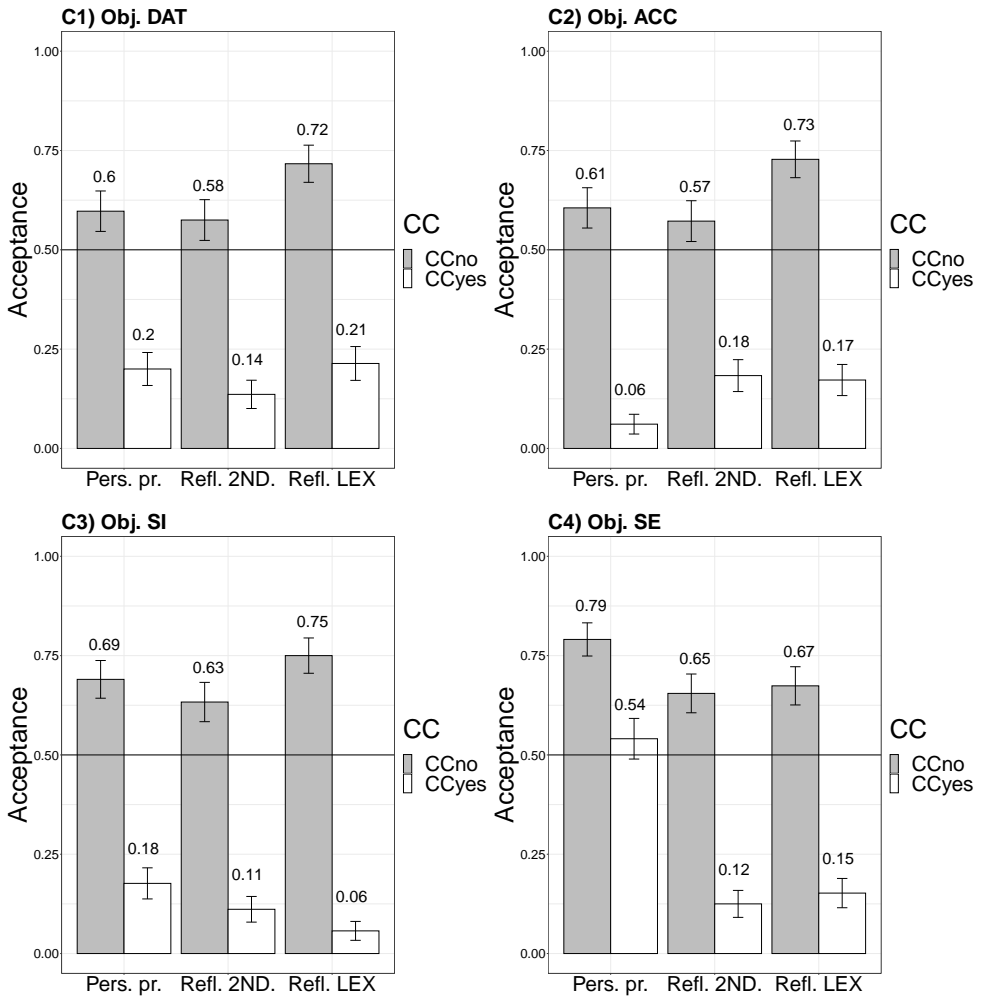


Figure 15.3: Sentences with object control predicates: Panels C1 and C2 (upper row) – pronominal CL controller, Panels C3 and C4 (bottom row) –  $REFL_{2ND}$  CL controller; Panels C1 and C3 (left) – dative, Panels C2 and C4 (right) – accusative

Acceptance rates (proportion of “yes” responses) of corresponding CC and noCC stimuli for different matrix predicates and different types of critical CLs: Pers. pr. – pronominal CLs; Refl. 2ND. –  $REFL_{2ND}$  CLs *se* and *si*; Refl. LEX. –  $REFL_{LEX}$  CL *se*. The horizontal black line marks a 50% acceptance rate.



- (4) a. Polako počinju<sub>1</sub> konkurirati<sub>2</sub> joj<sub>2</sub> na drugim područjima.  
slowly start.3PRS compete.INF her.DAT on other areas
- b. Polako joj<sub>2</sub> počinju<sub>1</sub> konkurirati<sub>2</sub> na drugim područjima.  
slowly her.DAT start.3PRS compete.INF on other areas  
'They are slowly starting to compete with her in other areas.'
- (5) a. Vjerojatno im<sub>1</sub> omogućuju<sub>1</sub> konkurirati<sub>2</sub> joj<sub>2</sub> na drugim  
probably them.DAT enable.3PRS compete.INF her.DAT on other  
područjima.  
areas
- b. ? Vjerojatno im<sub>1</sub> joj<sub>2</sub> omogućuju<sub>1</sub> konkurirati<sub>2</sub> na drugim  
probably them.DAT her.DAT enable.3PRS compete.INF on other  
područjima.  
areas  
'They are probably making it possible for those ones to compete with her  
in other areas.'

In the case of raising predicates, the preference of CC stimuli over the noCC variants was expected and is in line with the previous theoretical studies on CC (cf. Aljović 2005, Rezac 1999, 2005, Dotlačil 2004, Hana 2007) and with our corpus studies on CC out of *da*<sub>2</sub>-complements, infinitives and stacked infinitives presented in Chapters 13, 14, and in Hansen, Kolaković & Jurkiewicz-Rohrbacher (2018).

In contrast to sentences with raising and object control matrix predicates, sentences with subject control predicates could be seen as the middle ground between the two extremes. It must not be forgotten that these sentences can appear in two conditions: as simple subject control predicates (of the *planirati* 'plan' type) and as reflexive subject control predicates (of the *bojati se* 'be afraid' type). Once we account for this difference, as illustrated in Figure 15.2, it becomes obvious that the two differ drastically, thus revealing a much more complex pattern of results. This issue, which as far as we know has not been tackled before, is the subject of RQ 2.

Finally, including the type of both matrix predicate and infinitive CL in the analysis reveals an additional complexity in the results, as demonstrated by the three-way interaction in the statistical model presented in Tables 15.10–15.14. This complexity is addressed separately and explored through the relevant research questions that we presented in Section 15.2.

### 15.6.2 Simple subject control predicates and subject control predicates with REFL<sub>LEX</sub> CL *se*

The second research question addresses the relationship between CC and noCC variants for sentences containing simple subject control predicates and reflexive subject control predicates. When compared, the two variants of subject control predicates reveal different patterns of CC: compare Panels B1 and B2 from Figure 15.2.

In the case of simple subject control predicates (of the *planirati* ‘plan’ type), speakers show a clear preference for the CC stimuli, as compared to the corresponding noCC stimuli (compare examples (6a) and (6b)). This preference is independent of the infinitive CL type, i.e. the observed advantage is strikingly similar for pronominal CLs (69% noCC vs 89% CC), REFL<sub>2ND</sub> CLs *se* and *si* (68% noCC vs 91% CC), and the REFL<sub>LEX</sub> CL *se* (69% noCC vs 91% CC).

- (6) a. Polako nastojite<sub>1</sub> konkurirati<sub>2</sub> joj<sub>2</sub> na drugim područjima.  
           slowly try.2PRS compete.INF her.DAT on other areas
- b. Polako joj<sub>2</sub> nastojite<sub>1</sub> konkurirati<sub>2</sub> na drugim područjima.  
           slowly her.DAT try.2PRS compete.INF on other areas  
       ‘You are slowly trying to compete with her in other areas.’

However, the pattern of results changes drastically for sentences with reflexive subject control predicates (of the *bojati se* ‘be afraid’ type). Here, the acceptability of the CC stimuli highly depends on the critical CL type. If an infinitive governs a pronominal CL, the CC stimulus is acceptable (82%), but no longer favoured over its noCC variant (88%). In fact, we can say that both versions (CC and noCC) of sentences with reflexive subject control predicates whose infinitives govern a pronominal CL are, statistically speaking, equally acceptable (compare examples (7a) and (7b)).

- (7) a. Ipak se<sub>1</sub> trudim<sub>1</sub> konkurirati<sub>2</sub> joj<sub>2</sub> na drugim područjima.  
           still REFL try.1PRS compete.INF her.DAT on other areas
- b. Ipak joj<sub>2</sub> se<sub>1</sub> trudim<sub>1</sub> konkurirati<sub>2</sub> na drugim područjima.  
           still her.DAT REFL try.1PRS compete.INF on other areas  
       ‘Still, I am trying to compete with her in other areas.’

However, if the critical CL is REFL<sub>2ND</sub> or REFL<sub>LEX</sub>, then the CC stimuli are evidently unacceptable to the respondents (13% and 17% compared to 72% and 69% for their noCC versions). Compare examples (8a) and (8b) with the critical CL REFL<sub>LEX</sub> *se*.

- (8) a. Silno  $se_1$  boje $_1$  očitovati $_2$   $se_2$  o iznesenim  
immensely REFL be.afraid.3PRS declare.INF REFL about presented  
prijedlozima.  
suggestions
- b. ? Silno  $se_1$   $se_2$  boje $_1$  očitovati $_2$  o iznesenim  
immensely REFL REFL be.afraid.3PRS declare.INF about presented  
prijedlozima.  
suggestions
- ‘They are immensely afraid to voice their opinion on the presented sugges-  
tions.’

As we already reported in Section 11.3, both Rezac (2005) and Hana (2007) clearly state that in Czech there are no restrictions on CC out of infinitive complements not only of raising but also of subject control predicates. Therefore it should not come as a surprise that our experiment participants favoured CC versions of sentences with simple subject control predicates. Moreover, the results of the acceptability judgment experiment do not diverge from the corpus linguistic data on CC out of  $da_2$ -complements, and infinitives presented in Chapters 13 and 14 – in the case of simple subject control predicates, CC sentences are more frequent than noCC sentences.

However, as our results show, for CC it does not only matter whether CTPs are of the raising or subject control type, it also matters whether the latter are simple or reflexive. Namely, as Figure 15.3 suggests, in the case of simple subject control CTPs there are no constraints on CC. However, if CTPs are of the reflexive subject control type, only pronominal CLs can climb out of the infinitive complement, whereas the  $REFL_{2ND}$  CL  $se$ ,  $REFL_{2ND}$  CL  $si$  and  $REFL_{LEX}$  CL  $se$  do not climb. These results confirm Junghanns’ (2002: 79) pseudo-twins constraint.<sup>59</sup> However, the mentioned phenomenon has not hitherto been linked to the control phenomenon.

Further, Hana (2007) recognises only the object control reflexive constraint, while in his opinion subject control CTPs do not prevent (reflexive) CLs from climbing. However, according to our findings, the list of constraints on CC should clearly be updated, at least with respect to BCS. There is one more type of control constraint: the subject control reflexive constraint. Namely, if the matrix predicate is of the reflexive subject control type, reflexive infinitive CLs cannot climb.

<sup>59</sup>For more information on the pseudo-twins constraint see Section 11.4.1.

### 15.6.3 Object control predicates with a pronominal CL controller in the dative and with a pronominal CL controller in the accusative

The third research question aims to compare CC and noCC stimuli which contain object control matrix predicates whose controllers are a pronominal CL in the dative on the one hand and a pronominal CL in the accusative on the other. Sentences with object control predicates which have a pronominal CL controller in the dative (of the *dopuštati* ‘allow’ type) show an identical pattern of results regarding the acceptability rate of CC as those with a pronominal CL controller in the accusative (of the *prisiljavati* ‘force’ type) – compare Panels C1 and C2 from Figure 15.3.

The CC sentences are consistently rated as unacceptable, regardless of the controller (dative, accusative) and type of climbing CL (pronominal CL, REFL<sub>2ND</sub>, REFL<sub>LEX</sub>). NoCC sentences ((5a) and (9)) were clearly favoured over their permuted CC variants.

- (5a) Vjerojatno *im*<sub>1</sub> omoguđuju<sub>1</sub> konkurirati<sub>2</sub> *joj*<sub>2</sub> na drugim  
 probably them.DAT enable.3PRS compete.INF her.DAT on other  
 područjima.  
 areas  
 ‘They are probably making it possible for those ones to compete with her in other areas.’
- (9) Zakonski *te*<sub>1</sub> primoravamo<sub>1</sub> konkurirati<sub>2</sub> *joj*<sub>2</sub> na drugim  
 legally you.ACC compel.1PRS compete.INF her.DAT on other  
 područjima.  
 areas  
 ‘We are legally compelling you to compete with her in other areas.’

Following the discussion from Section 11.3 of strong restrictions on CC out of object-controlled infinitive complements, the experimental data indicate that we can generalise this rule to CC in Croatian too. Moreover, the case of the pronominal CL which is the controller has no impact on acceptability. These results are in line with our corpus linguistic study on CC out of *da*<sub>2</sub>-complements presented in Chapter 13. Both production data on CC out of *da*<sub>2</sub>-complements from srWaC and acceptability judgments, i.e., comprehension data on CC out of infinitive complements made by native speakers of Croatian, reveal the same pattern. Namely, CC out of object-controlled infinitive complements with either a dative or an accusative pronominal CL controller is highly restricted.

To sum up: we cannot say that one of the abovementioned object control predicate types is more likely to trigger CC. However, the theoretical discussion of CC in Czech indicates that the constraints on CC in the case of object control do not depend only on object control or on the case of the controller itself. Constraints most probably involve a combination of several features. Therefore, the results for Croatian presented in this section are supplemented in Section 15.6.9 with more detailed analyses, where not only the case of the matrix CL, but also the case of the critical CL are considered. This allows us to test the special object control person-case constraint described in Section 11.3.3. Finally, when the critical CL is of the  $REFL_{LEX}$  type (see (10) and (11) below), the contrast between acceptance rates of the noCC and CC stimuli seems to be even bigger than for pronominal or  $REFL_{2ND}$  critical CLs: compare the third bar pair with the first and second in Panels C1 and C2 from Figure 15.3.

- (10) Čak  $nam_1$  dopuštaš<sub>1</sub> očitovati<sub>2</sub>  $se_2$  o iznesenim prijedlozima.  
 even us.DAT allow.2PRS declare.INF REFL about presented suggestions  
 ‘You even allow us to voice our opinion on the presented suggestions.’
- (11) Vidno  $je_1$  požurujem<sub>1</sub> očitovati<sub>2</sub>  $se_2$  o iznesenim  
 visibly her.ACC hurry.1PRS declare.INF REFL about presented  
 prijedlozima.  
 suggestions  
 ‘I visibly hurry her to voice her opinion on the presented suggestions.’

In other words, whereas we observe no effect of the matrix CL type (i.e. controller), we do observe an effect of the infinitive CL type. Although this result is only marginally significant, it nicely fits Hana (2007: 129f) object control reflexive constraint. Even though Hana (2007) does not distinguish between different kinds of reflexives in Czech, our data show that the constraint seems to be slightly more salient in the case of  $REFL_{LEX}$  than in the case of  $REFL_{2ND}$ .

#### 15.6.4 Object control predicates with a $REFL_{2ND}$ CL *si* controller and with a $REFL_{2ND}$ CL *se* controller

The fourth research question addresses the relationship between CC and noCC stimuli for sentences with matrix object control predicates which have the  $REFL_{2ND}$  CL controller *si* (of the *dozvoljavati si* ‘allow yourself’ type) and the  $REFL_{2ND}$  CL controller *se* (of the *prisiljavati se* ‘force yourself’ type). Contrasting these sentences reveals an overall similarity with respect to CC acceptance with one exception: object control sentences with a  $REFL_{2ND}$  CL *se* controller whose infinitives govern a pronominal CL: compare Panels C3 and C4 from Figure 15.3.

Regardless of the case of the reflexive controller, noCC sentences are clearly favoured over the same CC sentences. See examples of noCC (12a) and CC (12b) sentences with the reflexive dative controller *si* presented below.

(12) a. Vjerojatno *si*<sub>1</sub> dozvoljavate<sub>1</sub> konkurirati<sub>2</sub> *joj*<sub>2</sub> na drugim  
probably REFL allow.2PRS compete.INF her.DAT on other  
područjima.  
areas

b. ? Vjerojatno *joj*<sub>2</sub> *si*<sub>1</sub> dozvoljavate<sub>1</sub> konkurirati<sub>2</sub> na drugim  
probably her.DAT REFL allow.2PRS compete.INF on other  
područjima.  
areas

‘You are probably allowing yourselves to compete with her in other areas.’

In their CC version, sentences with an object control matrix predicate whose controller is a REFL<sub>2ND</sub> CL are highly unacceptable in all conditions but one: when the REFL<sub>2ND</sub> CL *se* is the controller and the infinitive governs a pronominal CL: see example (13b) presented below. This is clearly visible in Panel C4 in Figure 15.3: compare the first bar pair with the second and the third.

(13) a. Sada *se*<sub>1</sub> prisiljavate<sub>1</sub> zahvaliti<sub>2</sub> *im*<sub>2</sub> na nesebičnoj pomoći.  
now REFL force.2PRS thank.INF them.DAT on unselfish help

b. Sada *im*<sub>2</sub> *se*<sub>1</sub> prisiljavate<sub>1</sub> zahvaliti<sub>2</sub> na nesebičnoj pomoći.  
now them.DAT REFL force.2PRS thank.INF on unselfish help

‘Now you are forcing yourselves to thank them for their unselfish help.’

CC sentences like (13b) are more acceptable than any other CC sentences containing an object control matrix predicate with either a REFL<sub>2ND</sub> CL *si* or a REFL<sub>2ND</sub> CL *se* controller. However, we must point out that even in this condition, CC sentences similar to (13b) are less favoured than their noCC variants (similar to (13a)), and are acceptable to only slightly over 50% of the participants.

Summing up, with our data we extend the existing discussion on the object control constraint to matrix predicates with REFL<sub>2ND</sub> CL *si* and REFL<sub>2ND</sub> CL *se* controllers. The empirical analysis shows that in the case of object control matrix predicates, the type of controller (pronominal CL or REFL<sub>2ND</sub>) apparently does not play an important role in CC. However, as already stated above, there is one exception. Namely, the climbing of pronominal CLs in the context of a REFL<sub>2ND</sub> CL *se* controller (sentences like (13b)) is somewhat acceptable to our participants. Nevertheless, even in this case the acceptance rate barely reached the threshold of 50%.<sup>60</sup> This finding is interesting for several reasons. First, we must emphasise

<sup>60</sup>For more information on acceptability thresholds see Section 3.3.3.4.

that this setup is the only case in which the climbing of a pronominal CL out of an object-controlled infinitive seems to be possible. Second, in respect of CC, the REFL<sub>2ND</sub> CL *se* as controller behaves more like its phonetically closer REFL<sub>LEX</sub> CL *se* than its morphologically and syntactically closer REFL<sub>2ND</sub> CL *si*: compare Panel B2 from Figure 15.2 and Panels C4 with C3 from Figure 15.3.

One more finding merits attention in this section. The very low acceptance rates of CC sentences with the strings *se se* and *si se* (see second and third CCyes bars in both Panels C1 and C3 from Figure 15.3) confirm Junghanns (2002: 79) pseudo-twins constraint (see Section 11.4.1). Since in corpora we do not find evidence for such CC structures (mixed clusters with pseudo-twins) and the acceptability of noCC stimuli (pseudodiaclisis of pseudo-twins) lies well below 80%, we assume there should exist another, more acceptable construction.<sup>61,62</sup> This could be either haplogy/haplogy of unlikes or the *da*<sub>2</sub>-construction.<sup>63</sup> However, an additional acceptability judgment test would be needed to establish the most acceptable variant.

### 15.6.5 Object control predicates with a pronominal CL controller in the dative and with a REFL<sub>2ND</sub> CL *si* controller

The fifth research question was formulated to compare CC and noCC sentences containing object control matrix predicates with a pronominal CL controller in the dative (5a) and analogous sentences containing object control matrix predicates with a REFL<sub>2ND</sub> CL *si* controller (12a). Our results revealed a similar pattern of acceptance with respect to CC: compare Panels C1 and C3 from Figure 15.3.

- (5a) Vjerojatno *im*<sub>1</sub> omogućuju<sub>1</sub> konkurirati<sub>2</sub> *joj*<sub>2</sub> na drugim  
probably them.DAT enable.3PRS compete.INF her.DAT on other  
područjima.  
areas  
'They are probably making it possible for those ones to compete with her  
in other areas.'
- (12a) Vjerojatno *si*<sub>1</sub> dozvoljavate<sub>1</sub> konkurirati<sub>2</sub> *joj*<sub>2</sub> na drugim  
probably REFL allow.2PRS compete.INF her.DAT on other  
područjima.  
areas  
'You are probably allowing yourselves to compete with her in other areas.'

<sup>61</sup>For basic information on mixed clusters see Section 2.4.2.1.

<sup>62</sup>For more information on pseudodiaclisis see Section 2.4.5.

<sup>63</sup>For more information on haplogy of unlikes see Section 2.4.2.2.

Participants clearly favoured the noCC sentence variants over the CC variants, and perceived the CC stimuli as highly unacceptable. Therefore, we can conclude on empirical grounds that CLs cannot climb into an object-control matrix clause whose controller is in the dative case, regardless of the controller type (pronominal vs REFL<sub>2ND</sub> CL).

### 15.6.6 Object control predicates with a pronominal CL controller in the accusative and with a REFL<sub>2ND</sub> CL *se* controller

The sixth research question was intended to capture the difference between noCC and CC sentences with object control predicates whose controller is a pronominal CL in the accusative (14) and object control sentences whose controller is the REFL<sub>2ND</sub> CL *se* (15). The latter type is discussed above (Section 15.6.4, Panels C1 and C3 in Figure 15.3.). Our comparison revealed comparable preference of noCC sentences (similar to (14a) and (15a)) and very low acceptability of CC sentences (similar to (14b) and (15b)) in all conditions but one: compare Panels C2 and C4 from Figure 15.3.

- (14) a. Javno *ih*<sub>1</sub> obvezujem<sub>1</sub> pozivati<sub>2</sub> *ga*<sub>2</sub> na mjesečne  
publicly them.ACC oblige.1PRS invite.INF him.ACC on monthly  
sastanke.  
meetings
- b. ? Javno *ih*<sub>1</sub> *ga*<sub>2</sub> obvezujem<sub>1</sub> pozivati<sub>2</sub> na mjesečne  
publicly them.ACC him.ACC oblige.1PRS invite.INF on monthly  
sastanke.  
meetings
- ‘I publicly oblige them to invite him to the monthly meetings.’
- (15) a. Nevoljko *se*<sub>1</sub> prisiljavamo<sub>1</sub> pozivati<sub>2</sub> *ga*<sub>2</sub> na mjesečne  
unwillingly REFL force.1PRS invite.INF him.ACC on monthly  
sastanke.  
meetings
- b. Nevoljko *ga*<sub>2</sub> *se*<sub>1</sub> prisiljavamo<sub>1</sub> pozivati<sub>2</sub> na mjesečne  
unwillingly him.ACC REFL force.1PRS invite.INF on monthly  
sastanke.  
meetings
- ‘We begrudgingly force ourselves to invite him to the monthly meetings.’

As already shown in Section 15.6.4, object control sentences with a REFL<sub>2ND</sub> CL *se* controller and a pronominal infinitive CL are the only category of object control sentences which are somewhat acceptable in their CC version. The acceptance



rate for such sentences (see example presented in (15b) above) is above 50%. However, as already pointed out in Section 15.6.4, even for this sentence type the noCC variant is still preferred over the CC variant.

Our study is the first attempt to include the difference in controller type (pronominal vs reflexive) in the discussion of constraints on CC in the context of object control matrix predicates. As we showed in the previous section, CC is blocked no matter if the controller is a dative pronominal CL or the REFL<sub>2ND</sub> CL *si* in the dative. In contrast, when the controller is in the accusative, CC is blocked only if the controller is a pronominal CL, whereas in the case of the REFL<sub>2ND</sub> CL *se* controller, only the climbing of REFL<sub>LEX</sub> and REFL<sub>2ND</sub> CLs out of the infinitive is truly blocked. Pronominal CLs can climb out of an infinitive which is controlled by the REFL<sub>2ND</sub> CL *se*. Nevertheless, it must be pointed out that the CC stimuli in the mentioned setup barely reached the threshold of acceptability.

### 15.6.7 Reflexive subject control predicates and object control predicates with a REFL<sub>2ND</sub> CL *se* controller

The seventh research question was formulated to establish the relationship between CC and noCC variants for sentences containing reflexive subject control matrix predicates and object control matrix predicates with the REFL<sub>2ND</sub> CL *se* controller, which have already been discussed in Sections 15.6.2, 15.6.4, and 15.6.6. As we can see from Panel B2 in Figure 15.2 and C4 in Figure 15.3, the ratings of these sentences give strikingly similar patterns.

If an infinitive governs the REFL<sub>2ND</sub> CLs *se* and *si* or the REFL<sub>LEX</sub> CL *se*, the two contrasted sentence categories showed identical patterns of acceptance: CC sentence variants were accepted very rarely (13%, 17%, 12%, 15%) and were outranked by noCC sentence variants (72%, 69%, 65%, 67% respectively). For examples of the latter see the sentences in (8a) and (16) presented below.

- (8a) Silno *se*<sub>1</sub> boje<sub>1</sub> očitovati<sub>2</sub> *se*<sub>2</sub> o iznesenim  
immensely REFL be.afraid.3PRS declare.INF REFL about presented  
prijedlozima.  
suggestions  
'They are immensely afraid to voice their opinion on the presented  
suggestions.'
- (16) Redovito *se*<sub>1</sub> ovlašćujem<sub>1</sub> očitovati<sub>2</sub> *se*<sub>2</sub> o iznesenim  
regularly REFL authorise.1PRS declare.INF REFL about presented  
prijedlozima.  
suggestions  
'I regularly authorise myself to voice my opinion on the presented  
suggestions.'

## 15 Experimental study on clitic climbing out of infinitive complements

As argued in Sections 15.6.2 and 15.6.4, the acceptability of noCC stimuli (pseudodiachisis), which is well below 80%, indicates that either haplogy (deletion of one *se* CL) or the *da*<sub>2</sub>-construction might be more acceptable. Nevertheless, without an additional acceptability judgment test we can only speculate which construction is the most acceptable variant.

Furthermore, in Section 15.6.4 and this section we deliver empirical proof supporting Rosen's claim that the pseudo-twins constraint is "blind" to different types of reflexives (cf. Rosen 2014: 104). As may be seen from Panel B2 of Figure 15.2 and Panels C3 and C4 of 15.3, neither the REFL<sub>2ND</sub> CLs *se* and *si* nor the REFL<sub>LEX</sub> CL *se* can climb out of the infinitive if there is already a reflexive in the matrix clause.<sup>64</sup> If they are part of the matrix clause, both tested types of reflexives, REFL<sub>LEX</sub> CL *se* and REFL<sub>2ND</sub> CL *si* and *se*, block the climbing of other infinitive reflexive CLs. In other words, sentences with mixed reflexive clusters have very low acceptance rates. Since this constraint appears with both subject and object control matrix predicates, we can conclude that it does not depend on the difference between the mentioned predicates.

In contrast, in the case of reflexive subject control sentences with the REFL<sub>LEX</sub> CL *se* and object control sentences with the REFL<sub>2ND</sub> CL *se* controller, CC sentence variants in which infinitives govern a pronominal CL tend to be acceptable: see examples of such sentences in (17) and (18).

(17) Barem *ih*<sub>2</sub>        *se*<sub>1</sub>    *usuđujem*<sub>1</sub> *zaposliti*<sub>2</sub> *na neodređeno*.  
 at.least them.ACC REFL dare.1PRS hire.INF on indefinitely  
 'At least I dare to employ them indefinitely.'

(18) *Doslovno ih*<sub>2</sub>        *se*<sub>1</sub>    *primoravam*<sub>1</sub> *zaposliti*<sub>2</sub> *na neodređeno*.  
 literally them.ACC REFL compel.1PRS hire.INF on indefinitely  
 'I literally compel myself to employ them on a permanent basis.'

The only difference between the two is that acceptance of CC sentences with a reflexive subject control matrix predicate and an infinitive which governs a pronominal CL is significantly higher (82%; see example (17) provided above) than acceptance of CC sentences with an object control matrix predicate with a REFL<sub>2ND</sub> CL *se* controller and an infinitive which governs a pronominal CL (54%; see example (18) provided above). Although Panel B2 in Figure 15.2 and Panel C4 in Figure 15.3 suggest that there are some differences among the acceptance rates

<sup>64</sup>The reader should bear in mind that we did not test sentences with haplogy. This was not possible due to our experiment design. In order to test if CC is possible in such structures separate experiments should be designed. For more information on CC in the context of reflexives and haplogy see Section 11.4.1.

for noCC sentences, these differences were not statistically significant (88% and 79%).

Thus, as we already discussed in the previous section, in the case of object control predicates the REFL<sub>2ND</sub> CL *se* controller appears to be more similar to the REFL<sub>LEX</sub> CL *se* than to the REFL<sub>2ND</sub> CL *si* controller when it comes to CC.

### 15.6.8 Pronominal and reflexive (REFL<sub>LEX</sub> CL *se* and REFL<sub>2ND</sub> CLs *se* and *si*) infinitive CLs

The eighth research question addressed the importance of infinitive CL type for CC. Although (due to the observed three-way interaction) we have already discussed the role of infinitive CL type in the previous sections, we will briefly summarise the findings here. We can distinguish three effect types: sentences containing only one CL, reflexive subject control sentences and object control sentences with a REFL<sub>2ND</sub> CL *se* controller, and lastly object control sentences with a REFL<sub>2ND</sub> CL *si* or a pronominal CL controller.

First, for sentences containing only one CL, i.e. sentences with raising (examples 4b, 19, 20) and simple subject control (examples 6b, 21, 22) matrix predicates, the type of infinitive CL plays no role. In these two sentence categories, the effect of CC is universal, i.e. consistent across all types of infinitive CLs (pronominal CLs, REFL<sub>2ND</sub> CLs *se* and *si*, REFL<sub>LEX</sub> CL *se*).

- (19) Zato  $se_2$  moram<sub>1</sub> braniti<sub>2</sub> od apsurdnih optužbi.  
for.that.reason REFL must.1PRS defend.INF from absurd accusations  
'For that reason, I have to defend myself from absurd accusations.'
- (20) Aktivno  $se_2$  nastavljaju<sub>1</sub> baviti<sub>2</sub> tim pitanjem.  
actively REFL continue.3PRS deal.INF that question  
'They are actively continuing to deal with that question.'
- (21) Zato  $se_2$  odlučuje<sub>1</sub> braniti<sub>2</sub> od apsurdnih optužbi.  
for.that.reason REFL decide.3PRS defend.INF from absurd accusations  
'For that reason, s/he decides to defend herself/himself from absurd accusations.'
- (22) Aktivno  $se_2$  pokušavaju<sub>1</sub> baviti<sub>2</sub> tim pitanjem.  
actively REFL try.3PRS deal.INF that question  
'They are actively trying to deal with that question.'

Second, as already discussed in the previous section, for reflexive subject control sentences (8b, 23, 24) and object control sentences with a REFL<sub>2ND</sub> CL *se* controller (25, 26, 27), the effect of CC is identical for the infinitive CLs: REFL<sub>LEX</sub> CL *se* (8b,

15 *Experimental study on clitic climbing out of infinitive complements*

25) and REFL<sub>2ND</sub> CL *se* (23, 26) and *si* (24, 27). No reflexive-type CLs can climb out of the infinitive into the matrix clause. The strings *se se* and *si se* are ruled out.

- (23) ? Pogotovno *se*<sub>2</sub> *se*<sub>2</sub> boji<sub>1</sub>            zaštititi<sub>2</sub> od smrtonosnih  
 especially REFL REFL be.afraid.3PRS protect.INF from lethal  
 bolesti.  
 diseases  
 ‘He/she is especially afraid to protect himself/herself from lethal  
 diseases.’
- (24) ? Pogotovno *si*<sub>2</sub> *se*<sub>2</sub> boji<sub>1</sub>            nauditi<sub>2</sub> takvom odlukom.  
 especially REFL REFL be.afraid.3PRS protect.INF that.kind decision  
 ‘He/she is especially afraid of harming himself/herself with such  
 decision.’
- (25) ? Potpuno *se*<sub>1</sub> *se*<sub>2</sub> prisiljavaš<sub>1</sub> oslanjati<sub>2</sub> na njegovu riječ.  
 even REFL REFL force.2PRS rely.INF on his word  
 ‘You completely force yourself to rely on his word.’
- (26) ? Čak *se*<sub>1</sub> *se*<sub>2</sub> prisiljava<sub>1</sub> zaštititi<sub>2</sub> od smrtonosnih bolesti.  
 even REFL REFL force.3PRS protect.INF lethal diseases  
 ‘He/she even forces himself/herself to protect himself/herself from  
 lethal diseases.’
- (27) ? Nesumnjio *si*<sub>2</sub> *se*<sub>2</sub> prisiljavamo<sub>1</sub> štedjeti<sub>2</sub> za sigurnu budućnost.  
 Undoubtedly REFL REFL force.1PRS save.INF for secure future  
 ‘Undoubtedly we force ourselves to save for a secure future.’

However, the contrast between CC and noCC stimuli is smaller for infinitive pronominal CLs. In other words, pronominal CLs can climb out of the infinitive in the case of reflexive subject control matrix predicates (17) and object control matrix predicates with the REFL<sub>2ND</sub> controller *se* (18).

- (17) Barem *ih*<sub>2</sub>            *se*<sub>1</sub> usuđujem<sub>1</sub> zaposliti<sub>2</sub> na neodređeno.  
 at.least them.ACC REFL dare.1PRS hire.INF on indefinitely  
 ‘At least I dare to employ them indefinitely.’
- (18) Doslovno *ih*<sub>2</sub>            *se*<sub>1</sub> primoravam<sub>1</sub> zaposliti<sub>2</sub> na neodređeno.  
 literally them.ACC REFL compel.1PRS hire.INF on indefinitely  
 ‘I literally compel myself to employ them indefinitely.’

Third, for object control sentences with a pronominal CL controller in the dative (examples 28 and 29), the accusative (examples 30 and 31), or a REFL<sub>2ND</sub> CL *si*

controller (examples 32 and 33), the effect of CC is identical for pronominal and REFL<sub>2ND</sub> *si* and *se* infinitive CLs: they cannot climb out of the infinitive into the matrix clause.

- (28) ? Opet *im*<sub>1</sub>        *mu*<sub>2</sub>    dopuštate<sub>1</sub> prigovarati<sub>2</sub> zbog lošeg  
 again them.DAT him.DAT allow.2PRS complain.INF because bad  
 društva.  
 company  
 ‘You are allowing them to complain about the bad company he keeps  
 again.’
- (29) ? Strogo *mi*<sub>1</sub>        *se*<sub>2</sub>    preporučuješ<sub>1</sub> odijevati<sub>2</sub> po najnovijoj modi.  
 strictly me.DAT REFL suggest.2PRS dress.INF on latest fashion  
 ‘You are strictly recommending me to dress according to the latest  
 fashion.’
- (30) ? Opet *ga*<sub>1</sub>        *mu*<sub>2</sub>    prisiljavaš<sub>1</sub> prigovarati<sub>2</sub> zbog lošeg  
 again him.ACC him.DAT force.2PRS complain.INF because bad  
 društva.  
 company  
 ‘You are forcing him to complain about the bad company that one keeps  
 again.’
- (31) ? Zbilja *ga*<sub>1</sub>        *se*<sub>2</sub>    tjeraš<sub>1</sub>        odijevati<sub>2</sub> po najnovijoj modi.  
 truly him.ACC REFL force.2PRS dress.INF on latest fashion  
 ‘You are truly forcing him to dress according to the latest fashion.’
- (32) ? Opet *mu*<sub>2</sub>        *si*<sub>1</sub>    dopuštaš<sub>1</sub> prigovarati<sub>2</sub> zbog lošeg društva.  
 again him.DAT REFL allow.2PRS complain.INF because bad company  
 ‘You are allowing yourself to complain about the bad company he keeps  
 again.’
- (33) ? Ponekad *si*<sub>1</sub>        *se*<sub>2</sub>    priuštīm<sub>1</sub>    odjenuti<sub>2</sub> po najnovijoj modi.  
 sometimes REFL REFL afford.1PRS dress.INF on latest fashion  
 ‘Sometimes I allow myself to dress according to the latest fashion.’

The above empirical findings can be summarised as follows. In the case of raising and simple subject control predicates, the type of the infinitive CL does not influence CC at all. In the case of reflexive subject control matrix predicates and object control matrix predicates with the REFL<sub>2ND</sub> controller *se*, we observe an effect of CL type on CC: pronominal infinitive complements can climb whereas the remaining two reflexive CL types cannot.

## 15.6.9 Dative and accusative infinitive CL complements

The ninth research question tackles the relationship between the case of the infinitive CL and CC. To answer it, we performed additional regression analyses with acceptance as dependent variable and CC (CC, noCC), infinitive CL type (pronominal CL, REFL<sub>2ND</sub> CL), and infinitive CL case (dative, accusative) as independent variables. Sentences containing the REFL<sub>LEX</sub> CL *se* as critical CL were excluded from the analyses. To avoid a possible four-way interaction which we would not be able to interpret, we conducted the analyses separately for each predicate type.

Table 15.15: Generalised mixed effects regression model fitted to acceptance data (1 = acceptable; 0 = unacceptable) for sentences containing an object control matrix predicate whose controller is a pronominal CL in the dative.

Random Effects		Var	SD	
participants (intercept adjustments)		0.770	0.877	
items (intercept adjustments)		0.463	0.680	
Fixed effects	Est.	SE	z	Pr(>  z )
intercept (CC - no; infinitive complement case - accusative)	1.461	0.237	6.167	< 0.0001 ***
CC - yes	-2.715	0.245	-11.098	< 0.0001 ***
infinitive complement case - dative	-1.964	0.238	-8.244	< 0.0001 ***

Regression models reveal that the effect of the case of the critical CL is significant only for object control matrix predicates with pronominal CL controllers.<sup>65</sup> Therefore, we now report only these results. They are shown in Figure 15.4 and Tables 15.15 (dative controller) and 15.16 (accusative controller).

As presented in Table 15.15 and Panel D1 from Figure 15.4, if an object control matrix predicate with a pronominal CL controller in the dative is followed by an infinitive which governs either a pronominal CL in the accusative or the REFL<sub>2ND</sub> CL *se*, the CC sentence (such as those in (34) and (35)) is more acceptable than

<sup>65</sup>Figure 15.4 presents data from models presented in Table 15.15 and 15.16. As we already emphasised at the beginning of this section, we conducted analyses separately for each predicate type to avoid a possible four-way interaction.

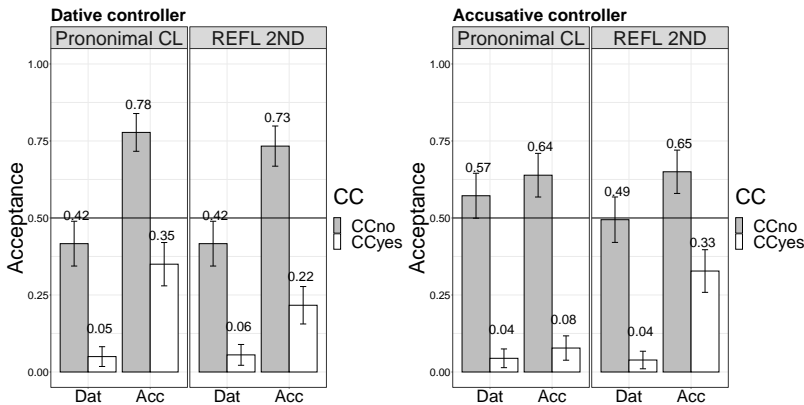


Figure 15.4: Effect of infinitive CL case on CC acceptance rate: Panel D1 (left) – sentences with object control matrix verb and pronominal dative CL controller, Panel D2 (right) – sentences with object control matrix verb and pronominal accusative CL controller.

the CC sentence in which the infinitive governs a pronominal CL in the dative or the REFL<sub>2ND</sub> CL *si* (such as those in (28) and (36)).<sup>66</sup>

- (34) ? *Uporno mi<sub>1</sub> ga<sub>2</sub> naređuju<sub>1</sub> pozivati<sub>2</sub> na mjesečne*  
 persistently me.DAT him.ACC order.3PRS invite.INF on monthly  
*sastanke.*  
 meetings  
 ‘They are persistently ordering me to invite him to the monthly  
 meetings.’
- (35) ? *Rado nam<sub>1</sub> se<sub>2</sub> pomažete<sub>1</sub> postaviti<sub>2</sub> prema takvom ponašanju.*  
 gladly us.DAT REFL help.2PRS set.up.INF towards such behaviour  
 ‘You gladly help us stand up to such behaviour.’
- (28) ? *Opet im<sub>1</sub> mu<sub>2</sub> dopuštate<sub>1</sub> prigovarati<sub>2</sub> zbog lošeg*  
 again them.DAT him.DAT allow.2PRS complain.INF because bad  
*društva.*  
 company  
 ‘You are allowing them to complain about the bad company he keeps  
 again.’

<sup>66</sup>The formula of the reported model is: Acceptance ~ CC + Infinitive complement case + (1|Participant) + (1|Item).

15 *Experimental study on clitic climbing out of infinitive complements*

- (36) ? Odnedavno  $mu_1$   $si_2$  omogućuju<sub>1</sub> posvetiti<sub>2</sub> dovoljno vremena.  
 recently him.DAT REFL enable.3PRS devote.INF enough time  
 ‘Recently, they have been making it possible for him to devote enough  
 time to himself.’

A slightly different pattern of results is observed for sentences containing an object control predicate with a pronominal CL controller in the accusative followed by an infinitive whose complement is a pronominal CL in the accusative or the REFL<sub>2ND</sub> CL *se*. As presented in Table 15.16 and Panel D2 from Figure 15.4, in this case we observe an interaction of CC and the case of the infinitive CL. Again, CC variants are less acceptable than the same sentences without CC. Sentences with infinitive CLs in the accusative are more acceptable than sentences with infinitive CLs in the dative, but only in CC sentence variants (in noCC sentence variants infinitive CL case has no effect).

- (31) ? Zbilja  $ga_1$   $se_2$  tjeraš<sub>1</sub> odijevati<sub>2</sub> po najnovijoj modi.  
 truly him.ACC REFL force.2PRS dress.INF on latest fashion  
 ‘You are truly forcing him to dress according to the latest fashion.’

In other words, stimuli like the one presented in (31) above are more acceptable than sentences like those presented below. It means that sentences with an object control matrix predicate which have a dative (5b) or an accusative (37) pronominal CL controller are, just like those with the REFL<sub>2ND</sub> CL *si* controller (12b), less acceptable in their CC version if the infinitive CL is in the dative.

- (5b) ? Vjerojatno  $im_1$   $joj_2$  omogućuju<sub>1</sub> konkurirati<sub>2</sub> na drugim  
 probably them.DAT her.DAT enable.3PRS compete.INF on other  
 područjima.  
 areas  
 ‘They are probably making it possible for those ones to compete with  
 her in other areas.’
- (37) ? Zakonski  $joj_2$   $te_1$  primoravamo<sub>1</sub> konkurirati<sub>2</sub> na drugim  
 legally her.DAT you.ACC compel.1PRS compete.INF on other  
 područjima.  
 areas  
 ‘We are legally compelling you to compete with her in other areas.’
- (12b) ? Vjerojatno  $joj_2$   $si_1$  dozvoljavate<sub>1</sub> konkurirati<sub>2</sub> na drugim  
 probably her.DAT REFL allow.2PRS compete.INF on other



područjima.

areas

‘You are probably allowing yourselves to compete with her in other areas.’

Furthermore, object control sentences with an accusative pronominal CL controller are less acceptable in their CC version if the infinitive CL is a pronoun in the accusative, like in the example (38) presented below.

- (38) ? Uistinu *me*<sub>1</sub> *ih*<sub>2</sub> *potičeš*<sub>1</sub> *tužiti*<sub>2</sub> za medijsku klevetu.  
 truly me.ACC them.ACC encourage.2PRS sue.INF for media slander  
 ‘You are truly encouraging me to sue them for media slander.’

Table 15.16: Generalised mixed effects regression model fitted to acceptance data (1 = acceptable; 0 = unacceptable) for sentences containing an object control matrix predicate whose governor is a pronominal CL in the accusative.

Random Effects		Var.	SD	
participants (intercept adjustments)		0.508	0.713	
items (intercept adjustments)		0.669	0.818	
Fixed Effects	Est.	SE	z	Pr(>  z )
intercept (CC - no; infinitive complement case - accusative)	0.718	0.261	2.747	0.006 **
CC - yes	-2.486	0.359	-6.930	<0.0001 ***
infinitive CL case - dative	-0.5623	0.335	-1.678	0.093
CC - yes: infinitive CL case - dative	-1.309	0.550	-2.379	0.017 *

However, although the accusative case of the infinitive CL is a statistically relevant factor for CC with respect to object control matrix predicates with a pronominal CL controller, we still cannot claim that there is no constraint on the climbing of pronominal accusative CLs in such a context (pace Dotlačil 2004, Rezac 2005). Namely, as can be seen in Figure 15.4, even if the climbing CL is in the accusative, object control sentences with dative and accusative pronominal CL controllers still do not reach the acceptability threshold of 50% in their CC versions.

## 15.7 Reaction time analysis

One reason for choosing the speeded yes-no acceptability judgment task was the possibility of obtaining a second, control measure – reaction time (see Section 3.3.3.1). In the present section we analyse the reaction time for accepted sentences.

The acceptance rates for almost all two-CL sentences are very low. In other words, the number of observations available for this part of the analysis is very small and does not allow the impact of all previously discussed predictors on reaction time to be checked.

Therefore, we pooled the data for all stimuli with two CLs, i.e. all sentences containing a reflexive subject control matrix predicate and all sentences with object control matrix predicates. Although the acceptance rates for stimuli with one CL were high, in order to make the conditions comparable, we also pooled the data for all stimuli with one CL, i.e. sentences with raising and simple subject control matrix predicates. We thus compared processing latencies for one-CL and two-CL CC and noCC sentences.

As presented in Figure 15.5, stimuli with only one CL took less time to be accepted than sentences with two CLs. Additionally, within the one-CL sentences, stimuli with CC were accepted faster than the corresponding noCC stimuli. In other words, CC sentences with raising and simple subject control matrix predicates were accepted faster than the noCC version of these sentences with the same matrix predicates. Two-CL sentences (i.e. sentences with reflexive subject control and object control matrix predicates) show no such advantage, as both variants are equally demanding in terms of processing.

The results of regression analysis summarised in Table 15.17 confirm these results.<sup>67</sup> Number of CLs and CC for stimuli with one CL are significant factors, and so is their interaction. There seems to be a numerical trend suggesting that two-CL CC sentences are processed even longer than their noCC counterparts, but this difference is not statistically significant.

Our findings clearly indicate a processing difference with respect to CC in two types of sentences. The reaction time analysis results for accepted sentences are in accordance with acceptance rate analysis results. One-CL sentences, i.e. sentences with raising and simple subject control matrix predicates, are more acceptable and processed faster in their CC version than in their noCC version. Similar patterns cannot be observed for sentences with two CLs.

---

<sup>67</sup>The formula of the model is:  $RT \sim \text{Number of CLs} * \text{CC} + \text{Infinitive CL type} + (1|\text{Participant}) + (1|\text{Item})$

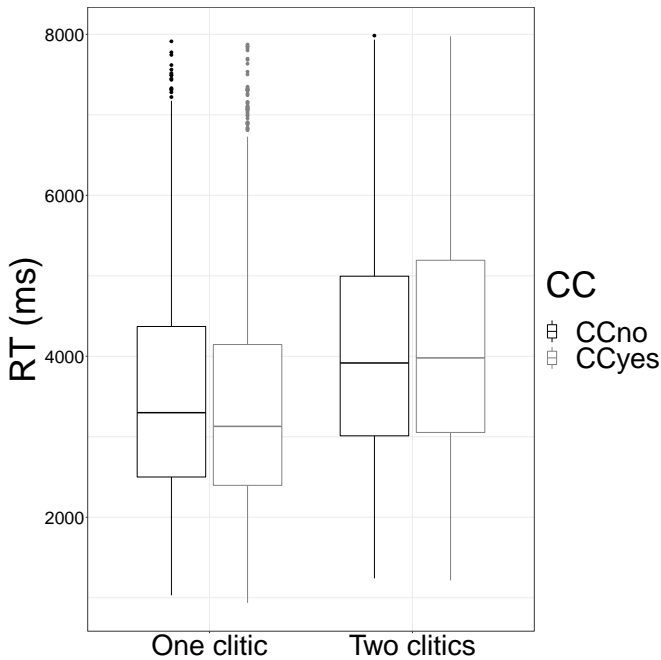


Figure 15.5: The effect of CC and number of CLs on processing latencies in accepted sentences. Sentences with one CL – bars on the left; sentences with two CLs – bars on the right. NoCC sentences in black; CC sentences in grey.

## 15.8 Conclusions

Hitherto studies of CC in Slavonic languages have had a weak empirical basis. Our study is the first psycholinguistic study which delivers experimental evidence for this syntactic phenomenon. We believe that our results contribute to the discussion on obligatoriness of CC in BCS and the constraints on it. The summarised results of our experimental study on CC in Croatian, in which native speakers rated the acceptability of CC and noCC sentences with raising, subject control and object control matrix predicates are presented comprehensively in Table 15.18. In Chapter 16 we compare the results of this study with the results of our corpus studies presented in Chapters 13 and 14.

We have solid reasons to believe that our data are reliable. As we already pointed out in Section 15.3.3, besides 48 target sentences we had 48 target-like syntactically and morphologically ill-formed sentences, which served us as control items. Through them we could establish whether our participants' responses had the necessary quality. The majority of our 336 participants rejected these

15 *Experimental study on clitic climbing out of infinitive complements*

Table 15.17: Mixed-effect regression model fitting CC and number of CLs to processing latencies for accepted sentences

Random Effects		Var.	SD	
participants (intercept adjustments)		0.050	0.224	
items (intercept adjustments)		0.009	0.094	
Fixed Effects	Est.	SE	$z$	Pr(>   $z$  )
intercept (number of CLs – two; CC – no; infinitive CL type – pronominal CL)	8.296	0.018	455.547	<0.0001 ***
number of CLs – one	-0.126	0.032	-3.968	<0.0001 ***
CC – yes	0.019	0.014	1.382	0.167
infinitive CL type – REFL <sub>LEX</sub> CL <i>se</i>	-0.033	0.012	-2.684	0.008 **
infinitive CL type – REFL <sub>2ND</sub>	-0.016	0.012	-1.286	0.199
number of CLs – one : CC – yes	-0.109	0.022	-4.940	<0.0001 ***

straightforwardly ill-formed control items. As we already indicated in Section 15.4.1, participants who accepted these clearly ill-formed sentences (18 in total) were excluded from further analysis.

As can be seen in Figure 15.3, results are internally consistent, i.e. parallel structures obtained similar scores in each mode. Moreover, the data collected in the second phase of the field research, which was conducted in March 2018 in Osijek, were not significantly different from the data collected in Zagreb and Split in December 2017 in the first phase of the field research.

In the following we will briefly discuss how our results contribute to the discussion on obligatoriness of CC in BCS. It has become clear that CC is not a unified phenomenon. To start with, Aljović (2005) for instance claims that in BCS CC is obligatory with restructuring predicates. In contrast, we have shown that CC is not obligatory in the context of raising and subject control predicates since the acceptability rate for noCC versions of sentences with the mentioned matrix predicates was around 50%.<sup>68</sup> However, we must point out that our participants clearly did favour CC over noCC versions of sentences with the mentioned matrix predicate types – see the summary of results in Table 15.18. Moreover, CC

<sup>68</sup> Although we are aware 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.

Table 15.18: Summary of the observed effects: stars denote significant effect of the given factor.

Num. CLs	Predicate type	CC	infinitive CL		Description
			Type	Case	
one	raising	*			CC preferred
	simple subject control	*			CC preferred
two	reflexive subject control	*	*		noCC preferred in general; CC somewhat more preferred for sentences with pronominal infinitive CLs than for sentences with reflexive infinitive CLs
	object control with pronominal controller in the dative	*	*	*	noCC preferred; contrast marginally higher if infinitive CL is REF <sub>L,LEX</sub> CL <i>se</i> ; pronominal CL and REF <sub>L,2ND</sub> CL more likely to climb if in accusative
	object control with pronominal controller in accusative	*	*	*	noCC preferred; contrast marginally higher if infinitive CL is REF <sub>L,LEX</sub> CL <i>se</i> ; REF <sub>L,2ND</sub> infinitive CL more likely to climb if in accusative
	object control with REF <sub>L,2ND</sub> CL <i>si</i> controller	*	*	*	noCC preferred; contrast marginally higher if infinitive CL is REF <sub>L,LEX</sub> CL <i>se</i>
	object control predicate with REF <sub>L,2ND</sub> CL <i>se</i> controller	*	*	*	noCC preferred; CC somewhat more preferred for sentences with pronominal infinitive CL than for sentences with reflexive infinitive CL

versions of sentences with raising and subject control predicates were processed faster than their noCC counterparts.

The results for object control matrix predicates are also partially in accordance with claims made for CC in Czech. As we already pointed out in Chapter 11, the object control constraint has been controversially discussed in the literature. With some exceptions, which we will briefly comment on, our findings are generally in line with what Thorpe (1991) and Junghanns (2002) claimed: CLs cannot climb out of object-controlled infinitives. As our empirical study shows, this claim made for Czech also holds for Croatian. In contrast to Thorpe (1991) and Junghanns (2002), Dotlačil (2004) and Rezac (2005) claim that in Czech CC out of object-controlled infinitives is possible in some special cases. For instance, in the theoretical syntactic literature it is claimed that in Czech, the climbing of accusative CLs is not blocked if the controller of the object control matrix predicate is in the dative (cf. Rezac 2005). In our CC experiment the ninth research question addressed this issue in Croatian. Our results revealed that the effect of the case of the infinitive CL was indeed significant for CC out of object-controlled infinitives. However, as can be seen in Figure 15.4, even if the controller is in the dative and the infinitive CL is in the accusative, CC is far from absolutely acceptable. Even in that configuration, the acceptability of CC sentences does not reach 50%. Although we did not conduct a systematic corpus experiment which would cover CC in the context of the mentioned matrix predicate types, as we pointed out in Chapter 11, some examples for the climbing of accusative CLs out of object-controlled infinitives with a dative controller can be found in hrWaC v2.2. Conversely, we did not come across examples of accusative CLs climbing out of object-controlled infinitives with an accusative controller. As can be seen in Figure 15.4, acceptability of the latter CC sentences was even lower than acceptability of CC sentences in which the accusative CL climbed out of an object controlled infinitive with a dative controller.

Furthermore, we can submit empirical evidence for an object control reflexive constraint in Croatian, which was reported for Czech by Hana (2007). As shown in Table 15.18 above, for object control matrix predicates the contrast between the acceptability of CC and noCC versions of sentences was higher if the infinitive CL was REFL<sub>LEX</sub>.

The novelty of our study lies also in the range of matrix predicates included. Namely, as far as we know, the authors who studied CC in Czech included neither reflexive subject control matrix nor object control matrix predicates with the REFL<sub>2ND</sub> *se* and REFL<sub>2ND</sub> *si* controllers in their discussion of CC. Our study revealed interesting insights into how the mentioned three matrix predicate types influence CC.

First, we can say that in respect of CC, object control matrix predicates with a  $\text{REFL}_{\text{LEX}}$  CL *si* controller do not behave differently than object control matrix predicates with dative and accusative pronominal CL controllers. In other words, all of them block CC out of the infinitive. Second, our empirical study showed that reflexive subject control matrix predicates behave the same in respect of CC as object control matrix predicates with the  $\text{REFL}_{2\text{ND}}$  CL *se* controller. In the case of both matrix predicates, the climbing of reflexive CLs, both the  $\text{REFL}_{2\text{ND}}$  CLs *se* and *si* and the  $\text{REFL}_{\text{LEX}}$  CL *se*, is blocked. These results are in line with the observation of Junghanns (2002) on the pseudo-twins constraint reported for Czech. While climbing of reflexive CLs is blocked, the mentioned matrix predicates with  $\text{REFL}_{\text{LEX}}$  CL *se* and  $\text{REFL}_{2\text{ND}}$  CL *se* allow pronominal CLs to climb out of infinitives. However, we must point out that noCC sentences were preferred by our participants. At first glance, the Panel B2 in Figure 15.2 and C4 in Figure 15.3 might suggest that there are differences in the acceptance rates of CC sentences between those two predicate types since the acceptance rate for the climbing of pronominal CLs is somewhat higher in the case of reflexive subject control predicates than in the case of object control predicates with the  $\text{REFL}_{2\text{ND}}$  *se* controller. However, those differences were not statistically significant.





# 16 On the heterogeneous nature of constraints on clitic climbing: Complexity effects

## 16.1 Introduction

In this chapter we summarise our main findings concerning constraints on CC. We also offer an explanation of this phenomenon in more general terms. We draw our conclusions from the theoretical literature and informal judgments described in Chapter 11, as well as the empirical studies in Chapters 13–15. Note that our empirical studies focus exclusively on CLs in the context of matrix embedding structures. Triangulation of methods, in which we compare observations from other studies with empirical results from our corpus studies and from a psycholinguistic experiment, gives us a very interesting picture of language production (real language usage) and language comprehension (what is judged acceptable).<sup>1</sup> As we explained in Section 3.3.3.5, the psycholinguistic experiment is necessary since it allows rare phenomena to be investigated and negative data to be obtained (cf. Hoffmann 2013: 100). We are also interested in the relationship between frequencies obtained in the corpus studies and the acceptance rates from the psycholinguistic experiment, as the connection between low frequency and acceptability is a contentious issue in linguistics (see Bermel & Knittl 2012, Divjak 2017).

As we discussed in Chapter 10 scholars working within formal-theoretical frameworks disagree as to optionality (Progovac 1993b, 1996, Čavar & Wilder 1994, Stjepanović 2004) and obligatoriness (Aljović 2005) of CC in BCS upon restructuring. In our empirical studies, we identified a certain variation which lets us conclude that even upon restructuring contexts CC is not obligatory. We observe that the significant factors influencing the probability of CC are very heterogeneous in their nature. Therefore, we go away from established methods of analysing the mechanism of clitic climbing towards the new paths which probabilistic syntax (Manning 2002) offers. In our view, constraints on CC are best

---

<sup>1</sup>For more information on triangulation of methods see Section 3.2.1.

interpreted in the context of complexity, which allows us to include also non-systemic sources of variation. It also enables constructing a series of hierarchies where the factors relevant for predicting clitic climbing interact with each other.

This chapter is structured as follows. In Section 16.2 we describe the concept of complexity. In the next three sections we discuss in detail the results of our study in the context of different types of complexity. In Section 16.3 we focus on systemic constraints related to embedding type. In Section 16.4 we proceed to systemic constraints related to the interaction of matrix and embedding. Section 16.5 explores the non-systemic factors constraining CC. In Section 16.6 we summarise our findings and draft a model of constraints based on the notion of complexity.

## 16.2 Complexity

### 16.2.1 The complexity of a system

We put forward the hypothesis that the mechanism behind CC and its constraints could best be explained as a certain type of complexity effect in the sense that the complexity of the sentence structures involved is the driving force for CC or blocking of CC.

Before we present our thoughts on these complexity effects, we have to discuss the somewhat elusive term of “complexity” as such. As Pallotti (2015: 118) shows, in the linguistic literature complexity has many competing meanings which tend to be confused. For instance, the formal properties of a construction are frequently identified with issues of difficulty or costs for the language user or learner. To avoid this polysemy, we follow Karlsson & Sinnemäki (2008: vii), who use the philosophical approach by Rescher (1998) in order to disentangle the bewildering range of studies on linguistic complexity. The point of departure is Rescher’s definition of the COMPLEXITY of a system:

A system’s complexity is a matter of the quantity and variety of its constituent elements and of the interrelational elaborateness of their organizational and operational make-up. Rescher (1998: 1)

Karlsson & Sinnemäki (2008: vii) deplore that this approach has not come to the fore in linguistic debates on complexity. We agree with the authors that it is well suited to application to syntactic structures (and beyond). Therefore, we will try to apply this precise and comprehensive approach to linguistic data, especially to clause structures relevant to the constraints on CC. Further, Rescher

(1998) makes a primary distinction between the complexity of systems themselves (ontological modes of complexity) and the complexity of how knowledge about systems can be presented (epistemic modes of complexity).

### 16.2.2 Ontological modes of complexity

In the current discussion we focus mainly on the ontological modes of complexity. Three perspectives may be offered on a system's complexity: compositional, structural and functional. Compositional complexity refers to parts of a system, structural complexity applies to the way the parts of a system can be combined, while functional complexity measures the variety of roles and contexts a system can be applied to. Below, we list the subtypes of complexity relevant to the following discussion as defined by Rescher (1998: 9) together with the modes of complexity they belong to:

1. **CONSTITUTIONAL COMPLEXITY** defined as “the number of constituent elements or components” (a subtype of compositional complexity);
2. **ORGANISATIONAL COMPLEXITY** defined as “variety of different possible ways of arranging components in different modes of interrelationship” (a subtype of structural complexity);
3. **OPERATIONAL COMPLEXITY** understood as “variety of modes of operation or types of functioning” (a subtype of functional complexity).

From the above, it is clear that there is no one absolute mode of complexity, but instead different modes of complexity interact with each other. For example, we can say that the constitutional complexity of pronominal CLs is higher than that of reflexive CLs. This is because they encode case and number, and for the third person singular also gender. This leads to a great number of unique items, and thus constitutional complexity decreases operational complexity of pronominal CLs, since each pronominal CL form has a narrow scope of reference. In contrast, the operational complexity of reflexive CLs is higher than that of pronominal CLs, as the number of functional contexts in which they can appear is greater (see Section 2.5.4).

We use complexity-related terms exclusively in relation to clause structures. We do not say anything about comprehension difficulty or about “overall complexity” ascribed to a language as a whole. With respect to structural complexity we can state that structures where both CC and noCC are possible are characterised by a higher grade of organisational complexity in comparison to those

where only CC or only noCC is allowed. In the following, we claim that certain types of complexity in a construction decrease the organisational complexity related to CL positioning. Next, we would like to discuss how complexity effects can explain differences in positioning of CLs belonging to embedded structures, i.e., CC vs noCC structures.

As outlined in Section 2.3, we distinguish systemic and non-systemic microvariation. In the particular case of CC, the former is defined as variation between the dependent variable CL position (CC vs noCC) and independent variables encoded in the linguistic context. The latter is defined as variation between the dependent variable CL position and independent variables related to space (diatopic variation) or to the modes of language use in different situations (e.g. oral vs written, diaphasic variation). Accordingly, we identify systemic and non-systemic constraints.

Following these lines of thought we propose two different types of systemic constraints. First, constraints related to the syntactic environment of the embedding. Second, constraints related to the matrix predicate which can potentially open a slot for a climbing CL. The constraints related to the matrix predicate are further subdivided into constraints related to predicate type with respect to the raising–control dichotomy and constraints related to the slot in the CL cluster in the matrix.

Rescher (1998: 11) points out that “[a] complex system that embodies subsystems can be organized either hierarchically through their subordination relations among its elements or coordinatively through their reciprocal interrelationships”. Here, we argue that some constraints on CC form hierarchical relationships: that is, they nest other constraints. Others operate as coordinative structures: that is, they interact with each other. The former applies to the type of embedding discussed in the next section. The latter applies to the constraints where predicate type is involved. These constraints, however, should not be seen as ultimate laws blocking CC, but rather as factors which each have a certain impact on the likelihood that CC will occur. Based on the findings from previous chapters we now discuss the individual types of constraints.

## 16.3 Systemic constraints related to the embedding type

### 16.3.1 Islands

In Chapter 11, following Franks & King (2000: 245) we used the concept of island for phrases showing a specific locality constraint on CC. In further chapters we observed that the spectrum of syntactic variation between embeddings as to



- (2) a. Nisam<sub>1</sub> imao<sub>1</sub> izbora<sub>1</sub> nego prodati<sub>2</sub> ga<sub>2</sub>.  
 NEG.be.1SG have.PTCP.SG.M choice than sell.INF him.ACC  
 b. \*Nisam<sub>1</sub> ga<sub>2</sub> imao<sub>1</sub> izbora<sub>1</sub> nego prodati<sub>2</sub>.  
 NEG.be.1SG him.ACC have.PTCP.SG.M choice than sell.INF  
 ‘I had no choice but to sell him (a football player).’ [bsWaC v1.2]

2. embedded wh-infinitives

- (3) a. Mila je<sub>1</sub> odlučila<sub>1</sub> kome ga<sub>2</sub> preporučiti<sub>2</sub>.  
 Mila be.3SG decide.PTCP.SG.F who him.ACC recommend.INF  
 b. \*Mila ga<sub>2</sub> je<sub>1</sub> odlučila<sub>1</sub> kome preporučiti<sub>2</sub>.  
 Mila him.ACC be.3SG decide.PTCP.SG.F who recommend.INF  
 ‘Mila decided to whom to recommend him.’ (BCS; Aljović 2005: 8)

Permutations of both structures show that CC leads to unacceptable sentences. Neither climbing of the accusative CL *ga* ‘him’ generated in the *nego* infinitive (2b) nor climbing of the same CL generated in the embedded infinitive headed by the wh-word *kome* ‘whom’ is possible (3b).

3. *da*<sub>1</sub>-complements

Although we do not discuss it in any detail, we assume that *da*<sub>1</sub>-complements, which unlike *da*<sub>2</sub>-complements are also inflected for tense, function as an additional true island.<sup>3,4</sup> The REFL<sub>LEX</sub> CL *se* cannot climb out of a future-tense *da*<sub>1</sub>-complement and form a mixed cluster with the matrix CLs *mi* ‘me’ and *je* ‘is’ – see example presented in (4a) and its permutation (4b).

- (4) a. On mi<sub>1</sub> je<sub>1</sub> obećao<sub>1</sub> da će<sub>2</sub> se<sub>2</sub>  
 he me.DAT be.3SG promise.PTCP.SG.M that FUT.3SG REFL  
 vratiti<sub>2</sub> u Kragujevac [...].  
 return.INF in Kragujevac  
 b. \*On mi<sub>1</sub> se<sub>2</sub> je<sub>1</sub> obećao<sub>1</sub> da će<sub>2</sub>  
 he me.DAT REFL be.3SG promise.PTCP.SG.M that FUT.3SG  
 vratiti<sub>2</sub> u Kragujevac [...].  
 return.INF in Kragujevac  
 ‘He promised me that he would come back to Kragujevac [...].’  
 [srWaC v1.2]

<sup>3</sup>For more information on *da*<sub>1</sub>-complements see Section 2.5.3.

<sup>4</sup>The assumption that *da*<sub>1</sub>-complements function as a true island is based on the study conducted by Hansen, Kolaković & Jurkiewicz-Rohrbacher (2016).

### 16.3.4 Complexity effects in embeddings

We argue that in the case of true islands 1, 2 and 3 we are dealing with phrases which show a higher degree of constitutional complexity than simple infinitive complements in matrix complement structures. Namely, if we look closer at the examples in (2a) and (3a), we see that the number of constituent elements or components is higher than in simple infinitive complements. In example (2a) the infinitive phrase is headed by the comparative marker *nego*, and in (3a) the infinitive phrase is headed by a wh-element. Furthermore, we can see structural similarities between the tied island described in Section 16.3.2 and the three true islands: all four of them contain phrases headed with an element in initial position:

- *da* + semifinite verb ( $da_2$ -complement),
- *nego* + infinitive,
- wh-element + infinitive,
- *da* + finite verb ( $da_1$ -complement).

Further, we would argue that constitutional complexity also explains the general differences in the behaviour of the different types of complements in respect to CC. Infinitive complements are less complex than  $da_2$ -complements because they do not contain agreement marking. In other words, unlike in  $da_2$ -complements the infinitive is not marked for number and person. There are thus two grammatical markers less. This is a difference in constitutional complexity. In a  $da_1$ -complement such as the one in (6a) the verb is even more complex as it additionally contains a tense marker.<sup>5</sup> As an island it totally blocks CC.

The fact that  $da_2$ -complements do allow CC to a certain degree seems to contradict our claim concerning constitutional complexity. At second glance, however, it can be explained by the assumption that in contrast to the *da* in  $da_1$ -complements, the wh-element and *nego*, the *da* in  $da_2$ -complements has lost its status as a complementiser. It has become a modal particle followed by a subjunctive, like the Albanian *të*, Bulgarian *da*, Greek *na*, and Romanian *să* (see Turano 2017). As shown by Joseph (1983: passim) the replacement of the infinitive construction with a subjunctive introduced by a modal particle started spreading north from Middle Greek. In BCS it first developed in the eastern varieties, but

---

<sup>5</sup>Our analysis correlates with works which assume that CLs climb from domains that are “functionally poor” like Aljović (2005).

later spread to other areas as well. Marković (1955: 33f), for instance, observed a drastic increase in the frequency of the structure on Bosnian language territory in the last century. The development of the complementiser into the subjunctive particle can be explained as a process of contact-induced grammaticalisation; for more details and further studies on this change see Wiemer & Hansen (2012: 80–83).

This assumption is supported by the fact that although both islands 1 and 2 are discussed above in the context of infinitives, they can also appear in the context of *da*<sub>2</sub>-complements, as shown in (5a) and (6a) and in their respective permutations:

1. comparative sentences with *nego*

- (5) a. [...] i nije<sub>1</sub> joj<sub>1</sub> ostalo<sub>1</sub> drugo nego  
 and NEG.be.3SG she.DAT remain.PTCP.SG.M else than  
 da mu<sub>2</sub> dozvoli<sub>2</sub> da ide<sub>3</sub> na Olimp.  
 that he.DAT allow.3PRS that go.3PRS on Olimp
- b. \* [...] i nije<sub>1</sub> joj<sub>1</sub> mu<sub>2</sub> ostalo<sub>1</sub> drugo  
 and NEG.be.3SG she.DAT he.DAT remain.PTCP.SG.M else  
 nego da dozvoli<sub>2</sub> da ide<sub>3</sub> na Olimp  
 than that allow.3PRS that go.3PRS on Olimp  
 ‘[...] and she had no choice but to allow him to climb Olympus.’  
 [srWaC v1.2]

2. embedded wh-*da*<sub>2</sub>-complements

- (6) a. [...] i razmišljam<sub>1</sub> kome da ga<sub>2</sub> poklonim<sub>2</sub>.  
 and think.1PRS who.DAT that him.ACC donate.1PRS
- b. \* [...] i razmišljam<sub>1</sub> ga<sub>2</sub> kome da poklonim<sub>2</sub>.  
 and think.1PRS him.ACC who.DAT that donate.1PRS  
 ‘[...] and I am thinking about whom I should give it to.’  
 [srWaC v1.2]

## 16.4 Systemic constraints related to the interaction of matrix and embedding

### 16.4.1 Constraints in the light of empirical evidence

Our discussion on systemic constraints related to the interaction of the matrix and the embedding is based on empirical evidence. We examined the behaviour



#### 16.4 Systemic constraints related to the interaction of matrix and embedding

of infinitive complement CLs of raising, simple subject control, reflexive subject control and object control predicates in different contexts and from various perspectives. Below we recapitulate the results of the corpus-based study (Figure 16.1) and of the psycholinguistic experiment (Figure 16.2), which are our primary points of reference in this section.

Figure 16.1 prepared on the basis of corpora shows the predicted probability of CC in the context of different matrix verbs. The result is driven by frequency of usage, and therefore, the figure models production. Figure 16.2 prepared on the basis of the experiment shows the predicted probability of a sentence with CC (right) or without CC (left) being accepted by a native speaker; hence, it models comprehension. It accounts for CTP type and CL type. The factor missing from both figures is CL case, because its impact is either insignificant for the model or was tested separately. In the following subsections we discuss each of the factors mentioned independently.

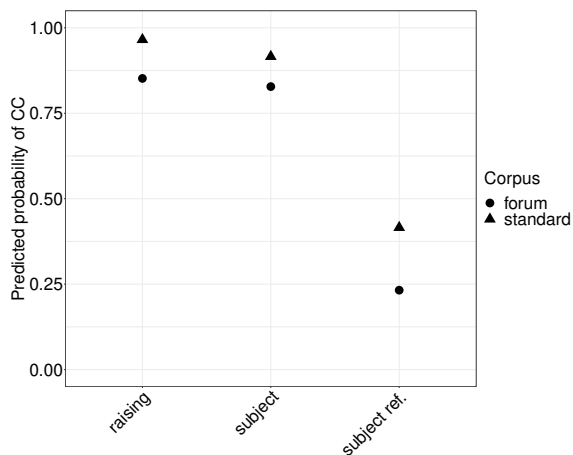


Figure 16.1: Results of the corpus-based study from Chapter 14

In addition to the two studies on constructions with infinitive complements, we draw on further evidence on systemic constraints related to the matrix from the study on *da*<sub>2</sub>-constructions and the raising–control distinction (Chapter 13). Since the data obtained in this study are more modest, they cannot be analysed with comparable quantitative methods. Therefore, in the argumentation which follows we use them only as a supplementary source.

We will now discuss each of the relevant factors (predicate type, CL type, CL case, mixed cluster effects).

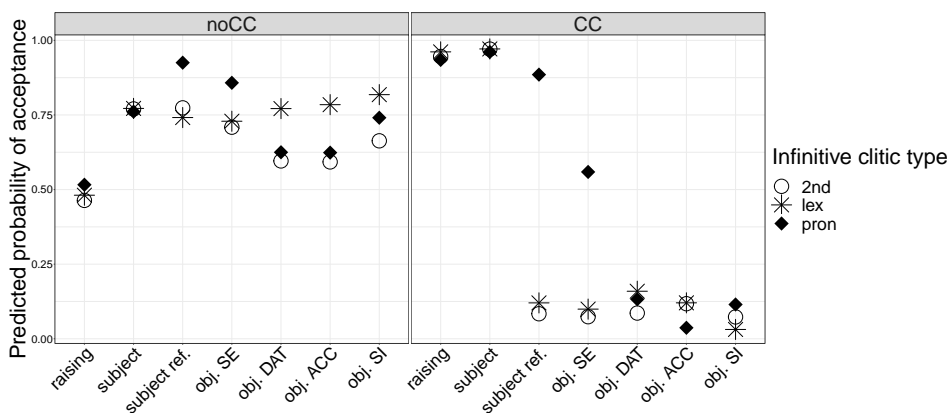


Figure 16.2: Main results of the psycholinguistic experiment from Chapter 15

### 16.4.2 Predicate type (CTP)

The models summarised in Figures 16.1 and 16.2 allow us to conclude that raising and simple subject control predicates do not differ drastically with respect to CC. CC is slightly more frequent and also more acceptable with raising than with simple subject control predicates. However, we observe a significant difference between reflexive subject and object control predicates on the one hand, and the other two types on the other. In the model built on experimental data, we additionally see an interaction with the CL type discussed in more detail in the next subsection. As our corpus studies and experimental data clearly show, neither reflexive subject control nor object control predicates allow reflexive CLs to climb.

### 16.4.3 CL type

In the corpus data for raising and simple subject control verbs, the type of the infinitive CL plays no role in CC. In other words, climbing of all types of CLs is similarly frequent for both types of predicates, and the differences in acceptability of such structures are insignificant. However, in the context of reflexive subject control matrix predicates, climbing of reflexive CLs seems completely impossible: see the example in (7a) without CC.

- (7) a. [...] koji se<sub>1</sub> boje<sub>1</sub> odreći<sub>2</sub> se<sub>2</sub> grijeha [...].  
 which REFL be.afraid.3PRS give.up.INF REFL sin

16.4 Systemic constraints related to the interaction of matrix and embedding

- b. \* [...] koji  $se_1$   $se_2$  boje<sub>1</sub> odreći<sub>2</sub> grijeha [...].  
 which REFL REFL be.afraid.3PRS give.up.INF sin  
 ‘[...] who are afraid to renounce their sins [...]’ [hrWaC v2.2]

We have not found evidence for examples similar to (7b) in corpora, nor were they accepted by native speakers in the psycholinguistic experiment. Furthermore, for sentences with reflexive subject control predicates even noCC versions with reflexive infinitive CLs (as in example (7a)) are extremely rare in corpora. However, the probability of speakers accepting such a construction is 0.75. This appears to be a major difference between production and comprehension.

We have found some evidence for climbing of pronominal CLs occurring with reflexive subject control CTPs in corpora, but its distribution is different than for raising and simple subject control predicates, which appear with CC in the majority of sentences. In corpora, sentences with reflexive subject control predicates are more frequent without CC, whereas in the experiment both versions ((8a) and (8b)) are equally acceptable.

- (8) a. Ipak  $se_1$  trudim<sub>1</sub> konkurirati<sub>2</sub> joj<sub>2</sub> na drugim područjima.  
 still REFL try.1PRS compete.INF her.DAT on other areas  
 b. Ipak joj<sub>2</sub>  $se_1$  trudim<sub>1</sub> konkurirati<sub>2</sub> na drugim područjima.  
 still her.DAT REFL try.1PRS compete.INF on other areas  
 ‘Still, I am trying to compete with her in other areas.’

Sentences with object control predicates were not retrieved from the corpus, as this turned out to be an extremely hard and costly task.<sup>6</sup> Like sentences with reflexive subject control predicates, sentences with CC and any kind of object control predicates are not acceptable, and we observe a further drop in acceptability for pronominal CLs. The CC version is somewhat likely to be accepted only in object control sentences with the reflexive controller *se*, like (9a), with a probability slightly over 0.5. However, sentences without climbing, like (9b), are accepted with a probability of about 0.8.

<sup>6</sup>Note, however, that this does not mean that such sentences are not retrievable from corpora. In Section 11.3.2 we give a number of examples with object control verbs and CC such as (i), although the experiment reveals that such sentences are not usually accepted:

- (i) [...] koji  $mi_1$   $ga_2$  pomažu<sub>1</sub> nositi<sub>2</sub>.  
 which me.DAT him.ACC help.3PRS carry.INF  
 ‘[...] which help me to carry it.’ [hrWaC v2.2]

- (9) a. Sada *se*<sub>1</sub> prisiljavate<sub>1</sub> zahvaliti<sub>2</sub> *im*<sub>2</sub> na nesebičnoj pomoći.  
 now REFL force.2PRS thank.INF them.DAT on unselfish help
- b. Sada *im*<sub>2</sub> *se*<sub>1</sub> prisiljavate<sub>1</sub> zahvaliti<sub>2</sub> na nesebičnoj pomoći.  
 now them.DAT REFL force.2PRS thank.INF on unselfish help  
 ‘Now you are forcing yourselves to thank them for their unselfish help.’

We see that the CL type factor operates in combination with the predicate type factor. Namely, whereas it does influence CC in sentences with reflexive subject control matrix predicates, it is not significant at all for raising and simple subject control matrix predicates. As may be seen in Figure 16.2, the probability that a sentence without CC will be accepted is always above 0.5 but mostly below 0.8.

When analysing the CL type factor, we observe some correspondence between the results for CC in the corpora and in the experiment. The drop in frequency of CC for individual CL types in the context of reflexive subject predicates in corpora corresponds to the decrease in acceptability of such sentences in the experiment. However, this is not the case for sentences without CC. In standard corpora we rarely find variants with pseudodiaclisis for reflexive subject control matrices and the infinitive reflexive CL (similar to the structure presented in (7a)).<sup>7</sup> Additionally, we have little corpus data for both variants with reflexive subject control predicates in general. Conversely, such sentences are acceptable to the participants of the experiment, but not at a level of 90–100% but rather a level of 50–80%. We can speculate that this is due to the form of the complement: even Croatian object control predicates seem to demand *da*<sub>2</sub>-complements and not the infinitive, and something similar might be true also in the case of reflexive subject control predicates. Notice also that we did not examine haplogy, which might be another highly acceptable structure, either in the experiment or in the corpus.

#### 16.4.4 CL case

Neither in the corpus studies nor in the experiment did we find evidence that the case of the infinitive CL might play a significant role for CC in the context of raising and subject control predicates. As it is very hard to search for examples of CC with object control predicates in corpora, we have only experimental data for this predicate type.<sup>8</sup> It turns out that the case of the infinitive CL is significant only for object control predicates with pronominal CL controllers in the

<sup>7</sup>For more information on pseudodiaclisis see Section 2.4.5.

<sup>8</sup>For an explanation why it is very hard to search for examples of CC with object control predicates in corpora see Section 14.4.



matrix predicates whose accusative CL climbs are accepted with a probability below 0.5.

We found some further evidence for the mixed cluster effect with the tied-island  $da_2$ -complement mentioned above. First, if two CLs are generated in a  $da_2$ -complement and occur in pseudodiacclisis, it is the pronominal that climbs and the reflexive that stays in the  $da_2$ -complement. Second, it seems that the reflexive CL *se* does not climb with the pronominal CL if there is a verbal CL in the matrix clause: see example (11).

- (11) [...] *i počelo<sub>1</sub> mi<sub>2</sub> je<sub>1</sub> da se<sub>2</sub> vrti<sub>2</sub> u glavi.*  
 and start.PTCP.SG.N me.DAT be.3SG that REFL spin.3PRS in head  
 ‘[...] and I started to feel dizzy.’ [srWaC v1.2]

#### 16.4.6 Complexity effects related to the interaction of matrix and embedding

Summing up these findings on constraints related to the interaction of the matrix and the embedding, which were best observed in the context of mixed clusters, pseudodiacclisis and haplology, we see a strong interaction of the individual factors. We put forward the hypothesis that they can be described by the following types of ontological complexity:

1. Constitutional complexity measured by the number of arguments of the CTP.
2. Organisational complexity measured by the number and types of slots available in the CL cluster. These slots can potentially be already taken by CLs belonging to the CTP and consequently not be available for climbing CLs.
3. Constitutional complexity related to each CL appearing in a structure, measured by the number of grammatical categories it encodes.
4. Operational complexity related to each CL appearing in a structure, measured by the number of functions that CL serves.

We apply these measures to our studies. Taking the typology of CTPs and the structure of the CL cluster, we see that the different predicate types may introduce their own CLs that fill the slots in the CL cluster sequence. As shown in Section 2.4.2.1, we assume the following slots and their relative order in the CL cluster:

#### 16.4 Systemic constraints related to the interaction of matrix and embedding

$li > \text{VERBAL} > \text{PRON}_{\text{DAT}} > \text{PRON}_{\text{ACC}} > \text{PRON}_{\text{GEN}} > \text{REFL} > je$

A cluster may contain the polar marker *li*, but it is not generated by a predicate. In the studies conducted, *li* does not show any variation in its behavior and regularly appears in 2P. Cases where *li* does not form clusters with other CLs are extremely rare. We thus have no grounds to assume that its presence is a large constraint on CC. CTPs differ as to the number and type of slots they can potentially cause to be occupied, as shown in Table 16.1.

Table 16.1: Complexity related to CTPs. Non-obligatory CL types are given in brackets.

CTP type	No. of arguments of CTP	No. of slots in cluster related to CTP	Type of slot in cluster related to CTP
raising	1	1	(V)
simple subject control	2	1	(V)
reflexive subject control	2	2	(V)+REFL
object control	3	2	(V)+(REFL/PRON)

Accordingly, raising predicates are the least complex of all CTPs as they have only one semantic argument into the complement whereas simple control predicates additionally have a semantic subject argument. As may be seen in Table 16.1, these predicate types potentially have only one, verbal position in the CL cluster to fill – when the matrix verb is in the past and future tense or in the conditional. Since the slots in the cluster reserved for pronominal and reflexive CLs remain free, CC is possible and very likely to take place.<sup>10</sup>

Like simple subject control predicates, reflexive subject control predicates have two semantic arguments and they can potentially fill one position in the CL cluster with a verbal CL. However, they also fill the position of the reflexive CL in the cluster. Since the slot reserved for a reflexive CL in the CL cluster is already occupied, CC is restricted. Namely, only pronominal CLs can climb and fill the free positions reserved for them in the CL cluster, whereas reflexive CLs must either stay in situ or haploglise. Therefore, the constitutional complexity of CTPs restricts organisational complexity with respect to the position of infinitive CLs.

<sup>10</sup>We are aware that there are subject control predicates CTPs denoting commissive speech acts (e.g. *obećati* ‘promise’) which can potentially fill not only the verbal slot but also the pronominal slot. Such verbs, however, were not surveyed in our studies. More information on those verbs and an example can be found in Section 2.5.2.

In terms of constitutional complexity, object control predicates are the most complex predicates as they have three semantic arguments. Moreover, they potentially fill two slots in the CL cluster with verbal and pronominal or reflexive CLs. Although a controller can be expressed as an NP, note that we have only studied structures in which it was expressed as a CL. In the studied structures, object control predicates always filled one position: either of the pronominal or of the reflexive CL.

Additionally, as we already pointed out in Section 16.2.2, pronominal CLs increase the constitutional complexity of a structure as a whole since they encode case and number (and for the third person singular also gender), while reflexive CLs increase its operational complexity due to their polyfunctionality.

As regards object control predicates with reflexive CL controllers, reflexive CLs may increase both constitutional complexity (CL *si*) and operational complexity. In general, they are unlikely to climb. The constructions with object control predicates that have reflexive CL controllers show some similarity to those with reflexive subject control predicates, as climbing of pronominal CLs is quite probable, though to a lesser degree than in the case of reflexive subject control CTPs. This can be considered an argument supporting the claim that the polyfunctionality of a CL (operational complexity) is an important factor. A CL belonging to a reflexive subject control CTP is “poorer” in that respect compared to other types of reflexive CLs. Further, in the light of constitutional complexity the REFL<sub>LEX</sub> CL *se* belonging to the reflexive subject control CTP is also less complex than the REFL<sub>2ND</sub> CL *se* belonging to the reflexive object control CTP since the latter encodes case.<sup>11</sup>

Finally, in Section 16.4.4 we noted that climbing of pronominal CLs in the context of object control predicates with a pronominal dative controller is less likely to be accepted than climbing of other CLs. This result can be explained by the fact that structures with mixed clusters containing two pronominal CLs represent the highest level of constitutional complexity, and compared to other cases the dative increases operational complexity, as it is the most polyfunctional case (cf. Silić & Pranjković 2007: 219f, 223).

---

<sup>11</sup>This difference in constitutional complexity and, accordingly, in CC between the two studied reflexive types REFL<sub>LEX</sub> and REFL<sub>2ND</sub> becomes even more apparent in the case of the REFL<sub>2ND</sub> CL *si*. The difference in constitutional complexity between the REFL<sub>LEX</sub> *se* and REFL<sub>2ND</sub> *se* is not immediately apparent since they coincide phonologically (which also explains why they behave in a similar but not the same way). In contrast, the difference in constitutional complexity between the REFL<sub>LEX</sub> *se* and REFL<sub>2ND</sub> *si* is more pronounced since their morphological differences have additional support in their different phonological realisation.



The combinations of complexity measures nicely explain the differences in frequency and acceptance of all four types of CTPs, including the less pronounced difference between raising and simple subject control predicates.

## 16.5 Non-systemic factors related to the diaphasic dimension

We did not test all the factors and their interactions as regards diatopic microvariation in all three BCS varieties. Nevertheless, in our earlier study on stacked infinitives in web corpora we did not find statistically relevant differences in CC between Croatian, Bosnian and Serbian (cf. Hansen, Kolaković & Jurkiewicz-Rohrbacher 2018).<sup>12</sup> Therefore, based on what we know, we have no reason to expect diatopic variation to be involved as a constraint on CC out of stacked infinitives. However, this does not exclude the possibility that diatopic microvariation does affect CC in the case of other factors. To establish this empirically further studies are needed.<sup>13</sup>

Apart from the systemic factors triggering microvariation in the domain of CC, we detected that the non-systemic diaphasic factor has an impact on CC at least for Croatian. In other words, there is a higher frequency of CC in the standard Croatian variety than in informal Croatian as presented in web fora: see Chapter 14. Namely, CC is used significantly more frequently in the standard than in informal language, in particular in the case of raising verbs. However, we have to point out that this is not a universal tendency as Spanish and European Portuguese, which have a pronominal CL system with CC phenomena, show the reverse tendency. In these Romance languages, CC is less frequent in written than in spoken texts.

In terms of complexity, we might argue that diaphasic variation is related to operational complexity: formal language is more codified, following rules, so operational complexity understood as a variety of modes should be lower. Low operational complexity means little variety, hence little flexibility in the way linguistic units can be combined. Necessarily, organisational complexity decreases too. In terms of CC, this results in the availability of only one position for CLs. As the study in Chapter 14 suggests, while in Croatian the prescribed variant is CC, in Romance it is noCC.

---

<sup>12</sup>Although there are differences in the frequency of stacked infinitives we have not found language-specific differences in the distributions of constructions with and without CC.

<sup>13</sup>For instance, we do not have empirical data on the difference in CC out of *da*<sub>2</sub>-complements between Bosnian and Serbian.

## 16.6 Interaction, optionality and preferences

Our findings for BCS fully corroborate the claim of Rosen (2001: 205) that word order properties of (Czech) CLs defy straightforward explanation due to the following two facts:

1. “several factors interact to determine their position [...]”,
2. “[o]nly some generalizations concerning their ordering behaviour can be expressed by strict rules, while other properties have to be stated as mere preferences”.

We have seen that most factors interact with each other. The factors CL type and CL case interact with the factor of matrix predicate type, but they are not active on their own. As to preferences, we saw certain patterns of CC which show graded acceptability. A nice case of peripheral usage is CC out of *da*<sub>2</sub>-constructions, which is rare as such, albeit possible in certain contexts (i.e. certain combinations of factors). This is why we proposed the new term *tied island*.

In relation to the question raised by some scholars of whether CC is obligatory, we have solid empirical grounds for regarding CC as optional. Namely, the acceptability rates for noCC versions of sentences with raising and simple subject control predicates are around 50%; that is, they reach the threshold of acceptability.

If we recall the theoretical accounts of CC briefly discussed in Chapter 10, we come to the conclusion that our findings are compatible with Junghanns’ (2002: 85f) claim that CC does not take place if the CL cannot reach the corresponding landing site in the matrix. We would argue that the effects responsible for blocking these landing sites can best be explained in terms of ontological complexity.

We can now show the changes in probabilities of CC based on the different types of complexity described above, starting from the most pervasive island constraint, as shown in Table 16.2. In the case of no islands and tied islands the

Table 16.2: The probability of CC with regard to the type of embedding (constitutional complexity)

Embedding	No island	>	Tied Island	>	True island
Probability of CC	Very high	>	Very low	>	0

Table 16.3: The probability of CC with regard to the type of CTP (constitutional complexity)

CTP type	Raising	Simple subject control	Reflexive subject control	Object control with <i>se</i> controller	Object control with pronominal and reflexive CL controller clearly marked for case
Probability of CC	High	> High	> Middle-low	> Middle-low	> Low
Coordination with other systemic factors	NA	NA	Infinitive CL type	Infinitive CL type and mixed cluster	Infinitive CL type and case, mixed cluster

Table 16.4: The probability of CC with regard to the mixed cluster effects (organisational complexity)

Slot occupied	Yes	No
Probability of CC	0	Regulated by other factors, in particular complexity related to other CLs in cluster

probability of CC varies depending on the systemic subordinate types of complexities (Tables 16.3 and 16.4) and on non-systemic factors.

According to our studies, the factors CL type and CL case are relevant only in the context of mixed clusters. Here, the operational complexity arising from the polyfunctionality of CLs appears to play a certain role. Since the exact cognitive processes behind the formation of mixed clusters are unknown, we refrain from drawing any strong conclusions.

We conclude that CC is not ruled by a single strict constraint. As a sort of perspective we put forward the hypothesis that the functioning of CC and its constraints could best be explained as a type of complexity effect in the sense that the constitutional (or operational) complexity of the involved sentence structures is the driving force for CC or for blocking it.



## **Part IV**

# **Final remarks**



# 17 Overall summary

## 17.1 Scope of the book

At the end of this data-oriented, empirical in-depth study of the BCS CLs we would like to repeat that CLs remain an extremely interesting subject for both data- and theory-driven syntactic research. This is because “the study of clitics can shed light on the interfaces between syntactic, morphological, and phonological linguistic representations” (Franks et al. 2004: 12). Yet, as we have pointed out, nearly all theoretical models assume a more or less stable and homogeneous CL system in BCS. However, it is no exaggeration to say that in many works data quality leaves much to be desired, which leads to contradictory statements about the acceptability of certain structures. Through our focus on empirical study of microvariation within the CL system in BCS we show that CLs are also an ideal test case for variationist approaches and the theory of linguistic complexity.

The book is divided into three main parts and a part with final remarks. Part I functions as an introduction containing definitions of our terms and concepts in the context of the briefly presented main assumptions of a selected number of theoretical approaches to CLs in BCS (Chapter 2). However, we tried to avoid premature theoretical commitments and as far as possible we used definitions as mere descriptive labels. As explained in Chapter 2, we identified three main parameters of variation:

1. inventory,
2. internal organisation of the CL cluster (CL ordering and morphonological processes within the cluster),
3. position of the CL or CL cluster (i.e. second position, delayed placement and phrase splitting).

Moreover, Part I contains the main tenets of our methodology (Chapters 3 and 4).

Part II targets systemic microvariation and selected cases of microvariation in the diatopic and the diaphasic dimensions. In it, we gave an empirical account of

variation in all three standard varieties: Bosnian, Croatian, and Serbian (Chapter 6), and in Štokavian dialects spoken on the territory of the former Yugoslavia and in some neighbouring countries (Chapter 7). Furthermore, we investigated the CL system in a spoken variety (Chapter 8). Due to the lack of comparable resources for all three varieties we restricted ourselves only to spoken Bosnian. We detected some global patterns of microvariation in the three above-mentioned parameters.

Our in-depth analysis of the existing literature on Czech showed CC to be a major source of variation which causes considerable disagreement among scholars (Chapter 11). As we are convinced that it is a highly important feature affected by systemic microvariation, we dedicated all of Part III to CC (Chapters 10–16).

In Chapter 10.1 we also offered our definition of CC as a phenomenon whereby a CL is not realised in a position contiguous to the elements of the embedding to which it belongs, but in a position contiguous to elements of the matrix. Finally, we tried to explain the constraints on CC in terms of complexity (Chapter 16).

## 17.2 Empirical approach

As to the empirical approach chosen, the current work is language-use oriented and is based on the triangulation of methods: intuition/theory – observation – experiment. The first step always involved a thorough analysis of the whole body of existing research literature, independently of the respective theoretical framework. We put theory-driven studies on an equal footing with normative and descriptive work. We thus strove to overcome the deplorable lack of exchange of thought between formal syntacticians and descriptive linguists, who tend to ignore each other. For the features showing some degree of microvariation we used empirical data collected in the years 2015–2018 from large web corpora {bs,hr,sr}WaC – our first source of observations. A selection of hypotheses concerning factors determining variation in the usage of CLs, formulated on the basis of corpus material, were further tested in acceptability judgment experiment where the level of control could be adjusted for individual factors. The whole book was actually inspired by Diesing et al. (2009: 60) who emphasised the need for empirical data in research on CLs. In their own words, the “[c]urrent research has [...] relied heavily on native speaker judgments that have been culled primarily from previously published work, or from interrogating native speaker linguists. While these are not uncommon methods in theoretical linguistics, it is well worth augmenting the database with other sources” (Diesing et al. 2009: 60).

Here we would like to underline that we do not entirely reject introspection as a linguistic method, but we argue in favour of a thorough documentation. The



reader should bear in mind that the first step of our research (intuition/theory) would not be possible without an informal judgment taks. However, we do believe that this as the only method of obtaining data might not be robust enough to permit generalisation, especially in a study dealing with variation, since the judgments of one or a few people cannot truly account for it. Since we are aware of the risks related to data gathering via intuition, we chose to design our studies in a way that avoids the traps mentioned in the methodological literature. We are well aware that “[s]cientific research is collaborative and incremental in nature, with researchers building on and extending each other’s work” (Stefanowitsch 2020: 133f). Therefore we have tried hard to be as transparent as possible concerning our data and methods. Namely, we have done our best to describe our research designs, research materials, and procedures associated with them explicitly and in sufficient detail. This should allow other researchers to retrace and check the correctness of each step of our analyses (cf. Stefanowitsch 2020: 133f). Such a practice, with some exceptions like Diesing et al. (2009), Zec & Filipović-Đurđević (2017), and Diesing & Zec (2017), is rarely applied in the study of BCS CLs.

### 17.3 The role of prescriptivism in the BCS clitic systems

Presumably like many scholars working on BCS CLs, we are also profoundly interested in the role played by prescriptivism in CL placement. The only way we were able to address this, at least indirectly, besides examining CL placement in BCS standard varieties (Chapter 6), was to observe CL placement in non-standard varieties, i.e. dialects (Chapter 7), spoken Bosnian (Chapter 8), and colloquial Croatian (Chapter 14). However, we are aware that it is not the same to compare CL placement in standard varieties (Chapter 6), whose descriptions are mainly based on written language, with CL placement in non-standard spoken varieties (Chapters 7 and 8).

In our search for variation, we tried our best not to depart from the standard assumption that prescriptive linguistic norms are much more rigorous in Croatia than in Serbia. While often cited in scholarly literature, we believe that this assumption is something of an unsubstantiated misconception.<sup>1</sup> The fact that the rules of standard Serbian differ from the Croatian norm does not make them any

---

<sup>1</sup>We would like to point out that no grammar book of Croatian language contains the word “normative” or “standard” in its title. On the other hand, a closer look at *Normativna gramatika srpskog jezika* (Piper & Klajn 2014) will reveal to the careful reader that not only Croatian but also Serbian has strict normative rules

less normative. In Chapter 6 we show how even in the period of the so-called Serbo-Croatian language Serbian and Croatian linguists recommended different CL placement. The Serbian linguist Pešikan (1958: 308) claimed that it is better to place CLs after a two-word phrase than to use DP of CLs. Moreover, the same linguist openly argued that the Croatian tendency to insert CLs after the first stressed word and to split phrases is, in his term, an “exaggeration” (cf. Pešikan 1958: 309).

The fact that standard Serbian chooses a different variant from the possible variants than Croatian does not make it less normative. According to dialectological literature (Chapter 7), Serbian speakers do split different kinds of phrases, some of which did not find their way into the Serbian norm, such as first name and surname, and coordinative phrases. Similarly, in the same chapter, we show that the reflexive *si* and the pronominal CL *ju* do occur in some Serbian dialects and that the latter is not always necessarily restricted to the context of the morphological process of suppletion. These features actually vanished from standard Serbian due to prescriptivism. In contrast, in standard Croatian they are recognized as legitimate variants available in the language system.

We can, of course, speculate that depending on teachers and their beliefs and attitudes to the norm, speakers of Croatian, just like speakers of Serbian, might have been instructed in school to place CLs according to the rules described in their standard grammar books. Nevertheless, normative instructions are one thing, and the internalisation of such rules is another. While obeying the rules is more likely in written language, it is very hard to control for in spoken language: “[l]evel of formality (style) may be easier to manipulate in performing for the linguist than pronunciation, which is easier to manipulate than morphological or syntactic behavior” (Stefanowitsch 2020: 26). As our descriptive and empirical studies on standard and non-standard varieties in Chapters 6, 7, 8, and 14 indicate, normative instructions and their internalisation do not always coincide.

To control for and to avoid potential prescriptive attitudes towards CL placement as a confounding variable in our psycholinguistic experimental study (Chapter 15), we deliberately excluded linguists and students of language studies as participants. It is well known that precisely these groups of participants can demonstrate rather prescriptive attitudes and may rely heavily on the notion of a narrowly defined standard language usage (cf. Krug & Sell 2013). Indeed, if we had not made this decision, it would have been much simpler to find participants for our experiment. However, as in many other respects, we deliberately chose to avoid the easier way in favour of one which would guarantee us better data and consequently give us better insights into the CL system in BCS.

## 17.4 Results: Parameters of variation

### 17.4.1 Inventory

We now briefly present our main results. First, it is interesting to note that all varieties of BCS seem to have a CL inventory which comprises four structural types: the polar question marker *li*, pronominal, reflexive and verbal CLs. No other types of CLs have been found. We detected only minor diatopic variation, mainly involving the reflexive CL *si* and the third person singular feminine accusative CL *ju*.

In the standard varieties only Croatian grammarians accept the reflexive CL *si*. The analysis of the dialectological literature, however, clearly shows that this form is found not only on Croatian, but also on Bosnian and Serbian language territory.

Further, Croatian and Serbian authors differ in their recommendations for the usage of the third person singular feminine accusative CL *ju*. Some Croatian authors treat *ju* as a separate unit of the inventory and not as a case of morphological suppletion (repeated morph constraint). Moreover, in the corpus of spoken Bosnian the CL *ju* is not attested at all. However, dialects give a very varied picture in this respect. Many idioms belonging to Old and Neo-Štokavian dialects have both the forms *ju* and *je* for the feminine singular accusative. Nevertheless, we also encounter dialects which use *je* exclusively or *ju* exclusively.

Finally, some dialects exhibit forms of pronominal CLs which have not found their way into any of the three standard norms: of them the *Prizrensko-južnomoravski* dialect spoken in Southern Serbia and Kosovo shows the greatest number of forms.

With respect to variation in the usage of verbal CLs we would like to emphasise that the conditional auxiliary form *bi* used for all persons (without inflection) is a case of allomorphy and not a true difference in the inventory.

### 17.4.2 Internal organisation of the clitic cluster

We find more variation in the parameter of internal organisation of the CL cluster. As to CL ordering in the cluster, the potential co-occurrence of the reflexive CL *se* and the verbal CL *je* is a clear case of microvariation. Whereas in standard Bosnian and Croatian both haplogy and the cluster *se je* are allowed, the Serbian normative grammar book Piper & Klajn (2014) accepts only haplogy. Further, we find ample evidence for the reversed CL order *je se*, which is attested in the central BCS territory, e.g. in the *Šumadijsko-vojvođanski*, *Zapadni*, *Slavonski*, *Srednjobosanski* and *Istočnohercegovački* dialects. In the data from the corpus

of spoken Bosnian the *je se* cluster sequence is four times more frequent than *se je* prescribed in standard Bosnian and Croatian.

The verbal CL *je* is affected by similar type of variation: according to our data, in some Serbian dialects it can precede pronominal CLs. Hence, it looks as if in these idioms the CL cluster was simpler than in the standard languages as it contains a single slot for all verbal CLs. To confirm this observation a separate study, involving bigger and better quality data, would be necessary.

Furthermore, in a very small number of Štokavian dialects, contrary to the CL ordering in the standard BCS varieties, pronominal CLs can stand in front of verbal CLs (the present tense and conditional forms of *biti*).

### 17.4.3 Morphological processes within the cluster

Another major source of variation is related to morphological processes, once again involving the CLs *se* and *je*. As we show in Chapter 6, haplology of unlikes is a normatively regulated feature, and even within standard varieties it is treated differently by different authors. The uneven distribution is clearly visible in our empirical data from the corpus of spoken Bosnian where haplology of unlikes is far more common than the co-occurrence of these two CLs (Chapter 8). Namely, haplology (with only the CL *se* occurring) is found in 68.8% of cases; a CL cluster (the CLs co-occur; the sequence *je se* is more frequent than *se je*), in 25.4% of cases; and in 5.6% of the cases the reflexive CL *se* and the verbal CL *je* appear in (pseudo)diacsis. Since non-omission is well attested in non-standard varieties (Chapters 7 and 8, we come to the conclusion that the processes in place to avoid the repetition of morphemes are rather unstable. For BCS we can conclude that the repeated morph constraint is a question of preference. Finally, there are some clues that the constraint on the string *se je* is not exclusively morphological in nature. As we note in Chapter 6, Ridjanović (2012: 564) argues that the deletion does not affect the verbal CL *je*, which is a copula, and the genitive CL *je* as an argument in the string *je se* is not affected either. Hence, in standard varieties only the auxiliary *je* seems susceptible to this process.

### 17.4.4 Position of clitics and clitic clusters

Summing up our discussion of the parameter position of clitics and clitic clusters, we present interesting findings on three levels. First, absolute 1P is ruled out in all of the three standard varieties. Accordingly, absolute 1P is not found in the corpus of spoken Bosnian either. We have only come across sentences in which CLs follow insertions, DSEs and retrospectives. However, the dialectal

survey shows that absolute 1P is attested in idioms which have been in intensive language contact with Romanian (*Šumadijsko-vojvođanski*) or with Macedonian (*Kosovsko-resavski*, *Prizrensko-južnomoravski* and *Timočko-lužnički*). These findings strongly suggest that absolute 1P is more probable in Štokavian contact varieties.

One further observation concerns CLs after the conjunctions *a* and *i*. Unlike standard BCS varieties, some dialects spoken in Croatia and Serbia allow the positioning of CLs directly after these coordinative conjunctions. Second, in standard Croatian and standard Bosnian the second position rule is understood as 2W, whereas in the literature on standard Serbian it is emphasised that 2P is normally understood as the position posterior to the first phrase (which may or may not be compound). Croatian and Bosnian standards allow the insertion of CLs into far more types of phrases than the Serbian standard does. However, both dialectal data and the data from the corpus of spoken Bosnian show ample evidence for splitting of conjoined NPs and quantificational phrases, which is not only widespread in Bosnian and Croatian territory, but can also be found in Serbian language territory. Contrary to some claims from the theoretical literature, in the spoken Bosnian variety not only subject phrases but also prepositional phrases can be split. Furthermore, we show that in some dialects and in the spoken Bosnian variety more than one CL can be inserted into a phrase. These data contradict Progovac (1996) and Radanović-Kocić (1988, 1996), who claim that clusters are not used as splitting elements.

Third, we are the first to empirically measure the heaviness of the initial constituent in spoken Bosnian. This feature is claimed to be a factor responsible for DP. Although many authors do mention heaviness, most of them do not provide any information on how to distinguish initial constituents which are heavy and cause DP from those which are not heavy and allow 2P. Inspired by the Czech linguists Kosek et al. (2018) we conducted an empirical study on spoken Bosnian. In the results we see a strong tendency towards 2W. The typical CL position in the clause is after the first word (77% of all observations), which is most frequently two graphemes long. The most frequent initial constituent in DP is three graphemes long, but in general its length is not limited, while the most frequent host in DP is four graphemes long; thus, it is longer than the initial constituent. This seems to suggest that DP depends not only on the heaviness of the first constituent, but might also be related to its phonological properties.

Although this was not in the focus of our study, in the spoken Bosnian data we find evidence that the polar question marker *li* differs from all the other CL types as in 100% of the cases it is placed in 2P after one short word (only 2 to 4

graphemes long). Its positioning is thus far more uniform than the positioning of verbal, pronominal and reflexive CLs.

#### 17.4.5 Constraints on clitic climbing

Part III is dedicated to CC and its constraints in BCS, a hitherto underresearched topic. CC deserves a separate part because the data from the grammar handbooks, dialectological sources and available corpora of spoken language are too scarce to allow for any sound conclusions. Since CC in Czech has received much more attention than in BCS, as a starting point we can draw on the body of research on CC in Czech. Because CC shows considerable similarity in both languages, we can start from Junghanns' (2002) findings. Our use of findings on CC in Czech as a point of departure does not mean that we assume that the constraints on CC in Czech can be automatically postulated and carried over to BCS. Instead, we transparently work bottom-up: we generate all possible hypotheses to check whether something from Czech might hold up for BCS. In other words, observations made for Czech serve as a starting point for hypothesis generation and not as assumptions that the very same mechanisms are present in BCS. The amount of empirical data in this book effectively proves that we have not tried to carry assumptions over from one language to another.

For the study of constraints on CC we combine all three types of methods, starting with intuition/theory based on the existing literature (mainly on Czech), through observation based on empirical corpus studies, and ending with a large psycholinguistic experiment involving acceptability judgments. In the corpus study we look into the highly controversial question whether CLs can climb out of *da*<sub>2</sub>-complements. Whereas some authors treat such sentences as completely normal, others reject them outright. We show that CC is indeed marginally possible, but exclusively with raising and simple subject control predicates. This study and a further study conducted by Hansen, Kolaković & Jurkiewicz-Rohrbacher (2018) on stacked infinitives, i.e. matrix predicates with multiple embeddings, serve as a basis for the test set-up of the psycholinguistic experiment. The test, which comprised 296 sentences, was carried out on 336 participants from various university institutions in Croatia. In the experiment we were able to test the impact of the following factors:

1. type of matrix verb (including the reflexivity of the predicate),
2. number of CLs in a sentence (one, two),

3. CL type, CL case,
4. position of infinitive CL (CC vs noCC).

We analysed the data, applying mixed-effects regression with participants and sentence endings as random variables, a statistical method which has become the golden standard in psycholinguistic research.

In addition to the systemic and non-systemic diatopic factors triggering microvariation, at least for Croatian we detected a higher frequency of CC in the standard than in the informal variety presented in web fora.

The findings presented in Chapter 16 show that the constraints belong to various levels of syntax. First, we fully agree with Rosen (2001: 205) that “several factors interact to determine their [CLs] position” and “only some generalizations concerning their ordering behaviour can be expressed by strict rules, while other properties have to be stated as mere preferences”. Most factors interact with some other factor. The factors CL type and CL case interact with the factor matrix predicate type and are not active in themselves. As to preferences, we saw certain patterns of CC which show graded acceptability.

Therefore, in our view, a single syntactic mechanism like restructuring cannot account for all of constraints to CC. Instead, we offer a solution alternative to the already existing theoretical approaches on CC in BCS, which enables us to account the whole broad spectrum of variation in empirical data we identify. We argue that the heterogeneous nature of CC can be best accounted for by complexity effects, which in the domain of syntax belong primarily to the constitutional and organisational subtypes of ontological complexity. Building on the approach by Rescher (1998) who gives a consistent typology of modes of complexity, we construct a series of hierarchies for the probability of CC based on the interaction between three factors: island > CTP type > mixed cluster effect. We propose a division of syntactic islands, a sort of locality constraint, into two subtypes: true islands and tied islands. The latter marginally allow CC (*da*<sub>2</sub>-complements).

Next to the structural constraints, we identify the diaphasic factor regulating CC in BCS. Diaphasic variation represents operational subtype of ontological complexity.

It goes without saying that we have not covered all the problems involving microvariation in the CL system of BCS. Especially in the vast field of CC, we leave to future research the similarity constraint and the haplology of pseudo-twins, for example. What we do plan to study in more detail, however, is haplology of unlikes and the constraints on CC related to the infinitive complement, in particular, stacked infinitives.





# Appendix A: Stimuli design

In the following we explain in more detail the design of stimuli from Chapter 15. The number of an example refers to the number of the experimental list. Each experimental list consisted of stimuli with only one of the seven matrix predicate types, i.e. raising, simple subject control, etc. as explained in Section 15.3.1. Similarly, the Latin letter assigned to each example indicates the infinitive CL type. The letter *a* is assigned to examples with third person pronominal CLs in the dative, while *b* stands for examples with third person pronominal CLs in the accusative. Examples with the REFL<sub>2ND</sub> CLs *si* and *se* are marked with *c* and *d*, respectively, while the letter *e* stands for examples with the REFL<sub>LEX</sub> CL *se*.

Examples of noCC stimuli sentences for each matrix predicate type for infinitives with pronominal dative CLs are presented in (A.1a)–(A.7a).<sup>1</sup>

- (A.1a) Posve      prestajemo prigovarati *mu* zbog lošeg društva.  
(A.2a) Ubrzo      odlučujete prigovarati *mu* zbog lošeg društva.  
(A.3a) Opet *se* stidiš      prigovarati *mu* zbog lošeg društva.  
(A.4a) Opet *im* dopuštate prigovarati *mu* zbog lošeg društva.  
(A.5a) Opet *ga* prisiljavaš prigovarati *mu* zbog lošeg društva.  
(A.6a) Opet *si* dopuštaš prigovarati *mu* zbog lošeg društva.  
(A.7a) Opet *se* spremaju prigovarati *mu* zbog lošeg društva.

---

<sup>1</sup> (A.1a) ‘We are entirely stopping complaining about the bad company he keeps.’

(A.2a) ‘You are presently deciding to complain about the bad company he keeps.’

(A.3a) ‘You are once again ashamed to complain about the bad company he keeps.’

(A.4a) ‘You are once again allowing them to complain about the bad company he keeps.’

(A.5a) ‘You are once again forcing him to complain about the bad company that one keeps.’

(A.6a) ‘You are once again allowing yourself to complain about the bad company he keeps.’

(A.7a) ‘They are once again preparing themselves to complain about the bad company he keeps.’

## A Stimuli design

Examples of noCC stimuli for each matrix predicate type for infinitives with pronominal accusative CL are presented in (A.1b)–(A.7b).<sup>2</sup>

- (A.1b) Stoga krećem pozivati *ga* na mjesečne sastanke.  
(A.2b) Čak nastojim pozivati *ga* na mjesečne sastanke.  
(A.3b) Nekako *se* ustručavamo pozivati *ga* na mjesečne sastanke.  
(A.4b) Uporno *mi* naređuju pozivati *ga* na mjesečne sastanke.  
(A.5b) Javno *ih* obvezujem pozivati *ga* na mjesečne sastanke.  
(A.6b) Ujedno *si* dopuštam pozivati *ga* na mjesečne sastanke.  
(A.7b) Nevoljko *se* prisiljavamo pozivati *ga* na mjesečne sastanke.

Examples of noCC stimuli for each matrix predicate type for infinitives with REFL<sub>2ND</sub> CL *si* are presented in (A.1c)–(A.7c).<sup>3</sup>

- (A.1c) Zaista mora ugadati *si* u svakom pogledu.  
(A.2c) Zaista nastojimo ugadati *si* u svakom pogledu.  
(A.3c) Zaista *se* ustručavam ugadati *si* u svakom pogledu.  
(A.4c) Zaista *ti* dozvoljavamo ugadati *si* u svakom pogledu.  
(A.5c) Zaista *je* potičemo ugadati *si* u svakom pogledu.  
(A.6c) Zbilja *si* naređujem ugadati *si* u svakom pogledu.  
(A.7c) Zaista *se* ohrabruje ugadati *si* u svakom pogledu.

---

<sup>2</sup> (A.1b) ‘Therefore, I am starting to invite him to the monthly meetings.’

(A.2b) ‘I am even trying to invite him to the monthly meetings.’

(A.3b) ‘We kind of hesitate to invite him to the monthly meetings.’

(A.4b) ‘They are persistently ordering me to invite him to the monthly meetings.’

(A.5b) ‘I publicly oblige them to invite him to the monthly meetings.’

(A.6b) ‘At the same time I am allowing myself to invite him to the monthly meetings.’

(A.7b) ‘We begrudgingly force ourselves to invite him to the monthly meetings.’

<sup>3</sup> (A.1c) ‘He really has to please himself in every way.’

(A.2c) ‘We are really trying to please ourselves in every way.’

(A.3c) ‘I truly hesitate to please myself in every way.’

(A.4c) ‘We really allow you to please yourself in every way.’

(A.5c) ‘We are truly encouraging her to please herself in every way.’

(A.6c) ‘I am really ordering myself to please myself in every way.’

(A.7c) ‘He is really encouraging himself to please himself in every way.’

Examples of noCC stimuli for each matrix predicate type for infinitives with REFL<sub>2ND</sub> CL *se* are presented in (A.1d)–(A.7d).<sup>4</sup>

- (A.1d) Svjesno            prestajete    skrivati *se* od znatiželjnih pogleda.  
 (A.2d) Svjesno            pokušavaš   skrivati *se* od znatiželjnih pogleda.  
 (A.3d) Čak            *se* usuđuju    skrivati *se* od znatiželjnih pogleda.  
 (A.4d) Oduvijek *mu*    dozvoljavaš   skrivati *se* od znatiželjnih pogleda.  
 (A.5d) Oduvijek *te*    puštam        skrivati *se* od znatiželjnih pogleda.  
 (A.6d) Uvijek        *si* dozvoljavaš   skrivati *se* od znatiželjnih pogleda.  
 (A.7d) Oduvijek *se*    primoravaju   skrivati *se* od znatiželjnih pogleda.

Examples of noCC stimuli for each matrix predicate type for infinitives with REFL<sub>LEX</sub> CL *se* are presented in (A.1e)–(A.7e).<sup>5</sup>

- (A.1e) Službeno            počinjemo    očitovati *se* o iznesenim prijedlozima.  
 (A.2e) Jasno            znaš            očitovati *se* o iznesenim prijedlozima.  
 (A.3e) Silno            *se* boje            očitovati *se* o iznesenim prijedlozima.  
 (A.4e) Čak            *nam* dopuštaš    očitovati *se* o iznesenim prijedlozima.  
 (A.5e) Vidno        *je* požurujem    očitovati *se* o iznesenim prijedlozima.  
 (A.6e) Uredno        *si* dozvoljavam   očitovati *se* o iznesenim prijedlozima.  
 (A.7e) Redovito *se*    ovlašćujem    očitovati *se* o iznesenim prijedlozima.

---

<sup>4</sup> (A.1d) ‘You consciously stop hiding from curious glances.’

(A.2d) ‘You are consciously trying to hide from curious glances.’

(A.3d) ‘They even dare to hide from curious glances.’

(A.4d) ‘You have been allowing him to hide from curious glances since always.’

(A.5d) ‘Since always I have been letting you hide from curious glances.’

(A.6d) ‘You always allow yourself to hide from curious glances.’

(A.7d) ‘Since always they have been forcing themselves to hide from curious glances.’

<sup>5</sup> (A.1e) ‘We are officially starting to voice your opinion on the presented suggestions.’

(A.2e) ‘You clearly know to voice your opinion on the presented suggestions.’

(A.3e) ‘They are immensely afraid to voice their opinion on the presented suggestions.’

(A.4e) ‘You even allow us to voice our opinion on the presented suggestions.’

(A.5e) ‘I visibly hurry her to voice her opinion on the presented suggestions.’

(A.6e) ‘I regularly allow myself to voice my opinion on the presented suggestions.’

(A.7e) ‘I regularly authorise myself to voice my opinion on the presented suggestions.’



# Appendix B: Explanation of statistical measures

In the following, we provide the explanation of statistical measures and significance codes used in Tables 14.2 and 15.11–15.17

Var.: Estimated variance of the random effects.

SD: Standard deviation corresponding to the estimated variance of the random effects.

Est.: Estimated value of the coefficient for the fixed effects.

SE: Standard error of the coefficient for the fixed effects.

z: The value of the z statistic corresponding to the estimated value of the fixed-effect coefficient.

Pr(> |z|): Probability of obtaining the given value of the z statistic or a value greater than the one which was obtained, assuming that the null hypothesis is true; statistical significance is marked as follows:  $p < 0.001$  (\*\*\*),  $p < 0.01$  (\*\*),  $p < 0.05$  (\*),  $p < 0.1$ .



## References

- Abbuhl, Rebekha, Susan Gass & Alison Mackey. 2013. Experimental research design. In Robert J. Podesva & Devyani Sharma (eds.), *Research methods in linguistics*, 116–134. Cambridge: Cambridge University Press.
- Adamovičová, Ana & Martin Vavřín. 2020. *Korpus InterCorp – srbština, verze 13 z 1. 11. 2020*. Praha: Ústav Českého národního korpusu, FF UK. <http://www.korpus.cz>.
- Aladrović, Katarina. 2011. Govorna raznolikost štokavskog narječja u Požeškoj kotlini i njezin utjecaj na usvajanje standardnoga jezika i razvoj komunikacijske kompetencije. In Đuro Blažeka (ed.), *1. Međimurski filološki dani. Zbornik radova*, 159–170. Zagreb: Učiteljski fakultet Sveučilišta u Zagrebu/Grafički zavod Hrvatske.
- Alexander, Ronelle. 2008. Rhythmic structure constituents and clitic placement in Bosnian, Croatian, Serbian. In Christina Y. Bethin (ed.), *American contributions to the 14th International Congress of Slavists*, vol. 1 Linguistics, 1–19. Bloomington, IN: Slavica Publishers.
- Alexander, Ronelle. 2009. What is second position, anyway? BCS clitics revisited. In Steven Franks, Vrinda Chidambaram & Brian Joseph (eds.), *A linguist's linguist: Studies in South Slavic linguistics in honor of E. Wayles Browne*, 43–67. Bloomington, IN: Slavica Publishers.
- Aljović, Nadira. 2005. On clitic climbing in Bosnian/Croatian/Serbian. In Nedžad Leko (ed.), *Lingvistički vidici*, vol. 34, 58–84. Sarajevo: Forum Bosnae.
- Aljović, Nadira. 2004. Cliticization domains: Clitic climbing in Romance and in Serbo-Croatian. In Olivier Crouzet, Hamida Demirdache & Sophie Wauquier-Gravelines (eds.), *Proceedings of JEL'2004 Domain(e)s*, 169–175. Université de Nantes.
- Angouri, Jo. 2010. Quantitative, qualitative or both? Combining methods in linguistic research. In Lia Litosseliti (ed.), *Research methods in linguistics*, 29–45. London, New York: Continuum.
- Avgustinova, Tania & Karel Oliva. 1995. The position of sentential clitics in the Czech clause. *CLAUS-Report 68* 68(12). 1–43. <https://ufal.mff.cuni.cz/~hana/rep/cl/avgustinova-oliva-1995-The%20Position%20of%20Sentential%20Clitics%20in%20the%20Czech%20Clause.pdf>.

## References

- Baayen, R. Harald. 2012. Mixed-effects models. In Abigail C. Cohn, Cécile Fougeron & Marie K. Huffman (eds.), *Handbook of laboratory phonology*, 668–677. Oxford: Oxford University Press.
- Baayen, R. Harald, Debra J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412. DOI: 10.1016/j.jml.2007.12.005.
- Baayen, R. Harald & Petar Milin. 2010. Analyzing reaction times. *International Journal of Psychological Research* 3(2). 12–28. DOI: doi.org/10.21500/20112084.807.
- Baayen, R. Harald, Petar Milin & Michael Ramscar. 2016. Frequency in lexical processing. *Aphasiology* 30(11). 1174–1220. DOI: 10.1080/02687038.2016.1147767.
- Babić, Stjepan. 1963. Enklitika između imena i prezimena. *Jezik: časopis za kulturu hrvatskoga književnog jezika* 11(2). 63–64. <https://hrcak.srce.hr/226150>.
- Babić, Stjepan. 1964. Lingvistika i logika Ivana Brabeca. *Jezik: časopis za kulturu hrvatskoga književnog jezika* 12(5). 151–159. <https://hrcak.srce.hr/226212>.
- Bader, Markus & Jana Häussler. 2010. Toward a model of grammaticality judgments. *Journal of Linguistics* 46(2). 273–330. <https://www.jstor.org/stable/40731815>.
- Balvet, Antonio, Dejan Stošić & Aleksandra Miletić. 2014. TALC-sef a manually-revised POS-tagged literary corpus in Serbian, English and French. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the ninth international Conference on Language Resources and Evaluation (LREC'14)*, 4105–4110. Reykjavik: ELRA.
- Baotić, Josip. 1985. Govori sjeverne i sjeveroistočne Bosne: Iz sintakse. In Asim Peco (ed.), *Bosanskohercegovački dijalektološki zbornik* 5, 337–372. Sarajevo: RO Minerva.
- Barac-Grum, Vida, Dragica Malić, Slavko Pavešić & Zlatko Vince. 1971. *Jezični savjetnik s gramatikom*. Zagreb: Matica hrvatska.
- Barić, Eugenija, Mijo Lončarić, Dragica Malić, Slavko Pavešić, Mirko Peti, Vesna Zečević & Marija Znika. 1997. *Hrvatska gramatika*. Zagreb: Školska knjiga.
- Barić, Eugenija, Lana Hudeček, Milica Mihaljević & Luka Vukojević. 1999. *Hrvatski jezični savjetnik*. Zagreb: Institut za hrvatski jezik i jezikoslovlje.
- Barjaktarević, Danilo. 1966. Novopazarsko-sjenički govori. In Mihailo Stevanović (ed.), *Srpski dijalektološki zbornik* 16, 1–177. Belgrade: Naučno delo.
- Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. The WaCky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 3(43). 209–226. DOI: 10.1007/s10579-009-9081-4.



- Barr, Dale J. 2013. Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology* 4(328). 1–2. DOI: 10.3389/fpsyg.2013.00328.
- Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278. DOI: <http://dx.doi.org/10.1016/j.jml.2012.11.001>.
- Bates, Douglas, Reinhold Kliegl, Shravan Vasishth & R. Harald Baayen. 2015. *Par-simonious mixed models*. <https://arxiv.org/abs/1506.04967>.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI: 10.18637/jss.v067.i01.
- Benko, Vladimír. 2017. Are web corpora inferior? The case of Czech and Slovak. In Marc Kupietz, Andreas Witt, Piotr Bański, Dan Tufiş, Dan Cristea & Tamás Váradi (eds.), *Proceedings of the workshop on challenges in the management of large corpora and big data and natural language processing (CMLC-5+BigNLP 2017) including the papers from the Web-as-Corpus (WAC-XI) guest section*. Birmingham, 43–48. Mannheim: Institut für Deutsche Sprache.
- Bermel, Neil & Luděk Knittl. 2012. Corpus frequency and acceptability judgments: A study of morphosyntactic variants in Czech. *Corpus Linguistics and Linguistic Theory* 8(2). 241–275. DOI: 10.1515/cllt-2012-0010.
- Biber, Douglas. 2005. Representativeness in corpus design. In Geoffrey Sampson & Diana McCharly (eds.), *Corpus linguistics: Readings in a widening discipline*, 174–197. London: Continuum.
- Biber, Douglas. 2015. Corpus-based and corpus-driven analyses of language variation and use. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 193–225. Oxford: Oxford University Press.
- Biber, Douglas & Jesse Egbert. 2016. Register variation on the searchable web: A multidimensional analysis. *Journal of English Linguistics* 44(2). 95–137. DOI: 10.1177/0075424216628955.
- Bierwisch, Manfred. 1986. On the nature of semantic form in natural language. In Friedhart Klix & Herbert Hagedorf (eds.), *Human memory and cognitive capabilities: Mechanisms and performances*, 765–784. Amsterdam: North-Holland.
- Birtić, Matea, Ivana Brač & Siniša Runjaić. 2017. The main features of the e-Glava online valency dictionary. In Iztok Kosem, Carole Tiberius, Miloš Jakubiček, Jelena Kallas, Simon Krek & Vít Baisa (eds.), *Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference. Leiden, the Netherlands, 19–21 September 2017*, 43–62. Brno, Czech Republic: Lexical Computing CZ s.r.o.

## References

- Birzer, Sandra. 2015. Generally speaking, connectivity and conversation management combined: The functions of Russian *voobščę govorja* and Polish *ogólnie mówiąc*. *Russian Linguistics* 39(1). 81–115. DOI: 10.1007/s11185-014-9143-0.
- Blaikie, Norman. 2003. *Analyzing qualitative data*. London: Sage Publications.
- Bonet, Eulàlia. 1991. *Morphology after syntax: Pronominal clitics in Romance*. Department of Linguistics & Philosophy, Massachusetts Institute of Technology. (Doctoral dissertation).
- Bonet, Eulàlia. 1983. The person-case constraint: A morphological approach. In Heidi Harley & Colin Phillips (eds.), *The morphology-syntax connection*. MIT working papers in linguistics 22, 33–52. Cambridge: Cambridge MA.
- Bošković, Željko. 1995. Participle movement and second position cliticization in Serbo-Croatian. *Lingua* 96(4). 245–266. DOI: 10.1016/0024-3841(95)00020-Z.
- Bošković, Željko. 2000. Second position cliticisation: Syntax and/or phonology? In Frits Beukema & Marcel den Dikken (eds.), *Clitic phenomena in European languages*, 71–120. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Bošković, Željko. 2001. *On the nature of the syntax-phonology interface: Cliticization and related phenomena*. Amsterdam: Elsevier.
- Bošković, Željko. 2004. Clitic placement in South Slavic. *Journal of Slavic Linguistics* 12(1-2). 37–90. [www.jstor.org/stable/24599868](http://www.jstor.org/stable/24599868).
- Bošković, Željko. 2016. On second position clitics crosslinguistically. In Franc Marušić & Rok Žaucer (eds.), *Formal studies in Slovenian syntax: In honor of Janez Orešnik*, 23–53. Amsterdam: John Benjamins Publishing Company.
- Božović, Đorđe. 2021. *Klitički niz u jezicima balkanskog areala*. Belgrade: Belgrade University.
- Brabec, Ivan. 1964. Enklitika – šta je to. *Jezik: Časopis za Kulturu Hrvatskoga Književnog Jezika* 12(5). 143–151. <https://hrcak.srce.hr/226211>.
- Brabec, Ivan, Mate Hraste & Zreten Živković. 1963. *Gramatika hrvatskosrpskoga jezika*. Zagreb: Školska knjiga.
- Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: Linguistics in search of its evidential base (Studies in generative grammar)*, 75–96. Berlin: Mouton de Gruyter.
- Bresnan, Joan & Tatiana Nikitina. 2009. The gradience of the dative alternation. In Linda Ann Uyechi & Lian-Hee Wee (eds.), *Reality exploration and discovery: Pattern interaction in language and life*, 161–184. Stanford, CA: CSLI Publications.

- Brewer, Marilyn B. 2000. Research design and issues of validity. In Harry T. Reis & Charles M. Judd (eds.), *Handbook of research methods in social psychology*, 3–16. New York: Cambridge University Press.
- Brlobaš, Željka & Mijo Lončarić. 2012. Gornjolonjski dijalekt kajkavskoga narječja. *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje* 38(2). 229–249. <https://hrcak.srce.hr/103505>.
- Browne, Wayles. 1975. Serbo-Croatian enclitics for English-speaking learners. In Rudolf Filipović (ed.), *Kontrastivna analiza engleskog i hrvatskog ili srpskog jezika, svezak i*, 105–134. Zagreb: University of Zagreb.
- Browne, Wayles. 2003. Razlike u redu riječi u zavisnoj rečenici. *Wiener Slavistischer Almanach* Sonderband 57. 39–44. [slavistik-portal.de/datenpool/ebsees-db.html?data=96506](http://slavistik-portal.de/datenpool/ebsees-db.html?data=96506).
- Brozović, Dalibor. 2007. Govor u dolini rijeke Fojnice. *Hrvatski dijalektološki zbornik* 13. 1–183. <https://hrcak.srce.hr/154210>.
- Brozović, Dalibor & Pavle Ivić. 1988. *Jezič Srpskohrvatski Hrvatskosrpski Hrvatski ili Srpski*. Zagreb: Jugoslavenski leksikografski zavod “Miroslav Krleža”.
- Brysbaert, Marc, Pawel Mandera & Emmanuel Keuleers. 2018. The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science* 27(1). 45–50. DOI: <https://doi.org/10.1177/0963721417727521>.
- Buchstaller, Isabelle & Ghada Khattab. 2013. Population samples. In Robert J. Podesva & Devyani Sharma (eds.), *Research methods in linguistics*, 74–95. Cambridge: Cambridge University Press.
- Bukumirić, Mileta. 2003. Govori Severne Metohije. In Aleksandar Mladenović (ed.), *Srpski dijalektološki zbornik*, vol. 50, 3–356. Belgrade: Srpska akademija nauka i umetnosti i Institut za srpski jezik SANU.
- Cacoullos, Rena T. 1999. Construction frequency and reductive change: Diachronic and register variation in Spanish clitic climbing. *Language Variation and Change* 11(2). 143–170. DOI: [10.1017/S095439459911202X](https://doi.org/10.1017/S095439459911202X).
- Caink, Andrew. 2004. Semi-lexical heads and clitic climbing. *Journal of Slavic Linguistics* 12(1-2). 91–133. <https://www.jstor.org/stable/24599869>.
- Čamdžić, Amela & Richard Hudson. 2002. Clitics in Serbo-Croat-Bosnian. In Ad Neeleman & Reiko Vermeulen (eds.), *UCL working papers in linguistics* 14, 321–353. London: Department of Phonetics & Linguistics University College London.
- Ćavar, Damir & Dunja Brozović-Rončević. 2012. Riznica: The Croatian language corpus. *Prace filologiczne* LXIII. 51–65.
- Ćavar, Damir & Chris Wilder. 1994. Clitic third in Croatian. *Linguistics in Potsdam* 1. 25–64. DOI: [10.1515/9783110804010.429](https://doi.org/10.1515/9783110804010.429).

## References

- Čavar, Damir & Dunja Brozović-Rončević. 2011. *Riznica: The Croatian language corpus*. Institut za hrvatski jezik i jezikoslovlje. <https://www.clarin.si/noske/index.html>.
- Čavar, Damir & Melanie Seiss. 2011. Clitic placement, syntactic discontinuity and information structure. In Miriam Butt & Tracy H. King (eds.), *Proceedings of the LFG11 conference*, 131–151. CSLI Publications. <http://web.stanford.edu/group/cslipublications/cslipublications/LFG/16/lfg11.pdf>.
- Čedić, Ibrahim. 2001. *Osnovi gramatike bosanskog jezika*. Sarajevo: Institut za jezik.
- Čermák, František & Alexandr Rosen. 2012. The case of InterCorp, a multilingual parallel corpus. *International Journal of Corpus Linguistics* 17(3). 411–427. DOI: [doi.org/10.1075/ijcl.17.3.05cer](https://doi.org/10.1075/ijcl.17.3.05cer).
- Chafe, Wallace. 1994. *Discourse, consciousness, and time*. Chicago, IL: University of Chicago Press.
- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Ćirković, Svetlana. 2021. Nezamenička upotreba enklitičkog oblika dativa refleksivne zamenice si u timočkim govorima: Sintaksički obrasci. *Zbornik Matice srpske za filologiju i lingvistiku* 1(LXIV). 97–114. DOI: [10.18485/ms\\_zmsfil.2021.64.1.5](https://doi.org/10.18485/ms_zmsfil.2021.64.1.5).
- Clark, Herbert H. 1973. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior* 12(4). 335–359. DOI: [10.1016/S0022-5371\(73\)80014-3](https://doi.org/10.1016/S0022-5371(73)80014-3).
- Clark, Herbert H. & Adrian Bangerter. 2004. Changing ideas about reference. In Ira A. Noveck & Dan Sperber (eds.), *Experimental pragmatics*, 25–49. Houndmills: Palgrave Macmillan.
- Čolak, Majda. 2015. Domene, registri i žanrovi u hrvatskome jeziku. *Tabula* 13(1). 29–44. DOI: [10.32728/tab.13.1.2015.02](https://doi.org/10.32728/tab.13.1.2015.02).
- Coseriu, Eugenio. 1980. „Historische Sprache“ und „Dialekt“. In Joachim Göschel, Pavle Ivić & Kurt Kehr (eds.), *Dialekt und Dialektologie: Ergebnisse des Internationalen Symposiums „Zur Theorie des Dialekts“ (Zeitschrift für Dialektologie und Linguistik. Beihefte. Neue Folge Nr. 26 der Zeitschrift für Mundartforschung)*, 106–122. Wiesbaden: Steiner.
- Cowart, Wayne. 1997. *Experimental syntax: Applying direct object methods to sentence judgments*. Thousand Oaks, CA: Sage Publications.
- Crible, Ludivine. 2016. Discourse markers and disfluencies: Integrating functional and formal annotations. In Harry Bunt (ed.), *Proceedings of the 12th joint ACL-ISO workshop on Interoperable Semantic Annotation*, 38–44. <http://hdl.handle.net/2078.1/179048>.

- Crible, Ludivine. 2018. *Discourse markers and (dis)fluency: Forms and functions across languages and registers*. Amsterdam: John Benjamins Publishing Company.
- Ćurković, Dijana. 2014. *Govor Bitelića*. Filozofski fakultet Sveučilišta u Rijeci. (Doctoral dissertation).
- Cvrček, Vaclav, Zuzana Komrsková, Dávid Lukeš, Petra Poukarová, Anna Řehořková & Adrian J. Zasina. 2018. Variabilita češtiny: Multidimenzionální analýza. *Slovo a slovesnost* 79(4). 293–321. <http://sas.ujc.cas.cz/archiv.php?art=4407>.
- Dąbrowska, Ewa. 2010. Naive v. Expert intuitions: An empirical study of acceptability judgments. *The Linguistic Review* 27. 1–23. DOI: 10.1515/tlir.2010.001.
- Davies, Mark. 1995. Analyzing syntactic variation with computer-based corpora: The case of modern Spanish clitic climbing. *Hispania* 78(2). 370–380. DOI: <https://doi.org/10.2307/345438>.
- Davies, William D. & Stanley Dubinsky. 2004. *The grammar of raising and control: A course in syntactic argumentation*. Malden: Blackwell.
- de Andrade, Arpldo L. 2010. The application of clitic climbing in European Portuguese and the role of register. In Claudia Borgonovo, Manuel Español-Echevarría & Philippe Prévost (eds.), *Selected proceedings of the 12th Hispanic linguistics symposium*, 97–108. Somerville, MA: Cascadilla Proceedings Project.
- Derwing, Bruce L., Roberto G. de Almeida & Assiniboia Hall. 2009. Non-chronometric experiments in linguistics. In David Eddington (ed.), *Experimental and quantitative linguistics*, 234–282. Munich: Lincom.
- Devitt, Michael. 2006. Intuitions in linguistics. *British Journal for the Philosophy of Science* 57(3). 481–513. DOI: 10.1093/bjps/axl017.
- Devitt, Michael. 2010. Linguistic intuitions revisited. *British Journal for the Philosophy of Science* 61(4). 833–865. DOI: 10.1093/bjps/axq018.
- Dickey, Stephen. 2000. *Parameters of Slavic aspect: A cognitive approach*. Stanford, CA: CSLI Publications, Center for the Study of Language & Information.
- Diesing, Molly. 2010. Clitics revisited. In Wayles Browne, Alison Fisher, Esra Kesici, Nikola Predolac & Draga Zec (eds.), *Formal approaches to Slavic linguistics 18*, 104–124. Ann Arbor, MI: Michigan Slavic Publications.
- Diesing, Molly, Dušica Filipović Đurđević & Draga Zec. 2009. Clitic placement in Serbian: Corpus and experimental evidence. In Susanne Winkler & Sam Featherston (eds.), *The fruits of empirical linguistics: Product*, vol. II, 59–73. Berlin, New York: Mouton de Gruyter.
- Diesing, Molly & Draga Zec. 2011. Interface effects: Serbian clitics. In Jeffrey Runner (ed.), *Experiments at the interfaces, syntax and semantics*, vol. 37, 1–30. Leiden: Emerald.

## References

- Diesing, Molly & Draga Zec. 2017. Getting in the first word: Prosody and predicate initial sentences in Serbian. *Glossa: A Journal of General Linguistics* 2(1). 1–25. DOI: 10.5334/gjgl.204.
- Divjak, Dagmar. 2017. The role of lexical frequency in the acceptability of syntactic variants: Evidence from that-clauses in Polish. *Cognitive Science* 41(2). 354–382. DOI: 10.1111/cogs.12335.
- Dobrić, Nikola. 2009. Savremene tendencije u lingvostatistici. *Statistička revija* 58(1-2). 45–50. <https://ssrn.com/abstract=2309957>.
- Dobrić, Nikola. 2012. Savremeni jezički korpusi na zapadnom Balkanu: Istorijat, trenutno stanje i budućnost. *Statistička revija* 60(4). 677–692. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2309948](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2309948).
- Donders, Franciscus Cornelis. 1868. On the speed of mental processes. *Acta Psychologica* 30. 412–431. <http://www2.psychology.uiowa.edu/faculty/mordkoff/InfoProc/pdfs/Donders%201868.pdf>.
- Dotlačil, Jakub. 2017. ŠPLHání KLITIK. In Petr Karlík, Marek Nekula & Jana Pleskalová (eds.), *CzechEncy: Nový encyklopedický slovník češtiny*. Brno. <https://www.czechency.org/slovník/%C5%A0PLH%C3%81N%C3%8D%20KLITIK>.
- Dotlačil, Jakub. 2004. *The syntax of infinitives in Czech*. University in Tromsø. (MA thesis).
- Dragičević, Milan. 2007. O govoru Srba Banje Vručice kraj Teslića I (Fonetske i morfološke osobine). In Aleksandar Mladenović (ed.), *Srpski dijalektološki zbornik*, vol. 54, 325–401. Belgrade: Srpska akademija nauka i umetnosti i Institut za srpski jezik SANU.
- Dukanović, Vlado. 1994. Glagoli trebati i valjati i njihova infinitivno/prezentska dopuna. *Naučni sastanak slavista u Vukove dane* 22(2). 119–126.
- EAGLES (Expert Advisory Group on Language Engineering Standards). 1996. *EAGLES. Preliminary recommendations on text typology*. [EAGLES document EAG-TCWG-TTYP/P].
- Eisenbeiss, Sonja. 2010. Production methods. In Elma Blom & Sharon Unsworth (eds.), *Experimental methods in language acquisition research*, 11–34. Amsterdam: John Benjamins Publishing Company.
- Fanselow, Gisbert. 2007. Carrots: Perfect as vegetables, but please not as a main dish. *Theoretical Linguistics* 33(3). 353–367. DOI: 10.1515/TL.2007.023.
- Farkaš, Loretana & Tena Babić. 2011. Govor Babine Grede. *Šokačka rič* 8. 133–154.
- Featherston, Sam. 2005. Universals and grammaticality: Wh-constraints in German and English. *Linguistics* 43(4). 667–711. DOI: 10.1515/ling.2005.43.4.667.
- Featherston, Sam. 2007. Reply. *Theoretical Linguistics* 33(3). 401–413. DOI: 10.1515/TL.2007.027.

- Featherston, Sam. 2008. Thermometer judgments as linguistic evidence. In Claudia Maria Riehl & Astrid Rothe (eds.), *Was ist linguistische Evidenz?*, 69–89. Aachen: Shaker Verlag.
- Featherston, Sam. 2009. A scale for measuring well-formedness: Why syntax needs boiling and freezing points. In Sam Featherston & Susanne Winkler (eds.), *The fruits of empirical linguistics*, vol. 1, Process, 47–74. Berlin: Mouton de Gruyter.
- Fechner, Gustav Theodor. 1860. *Elemente der Psychophysik*. Leipzig: Breitkopf und Härtel.
- Fehrmann, Dorothee, Uwe Junghanns & Denisa Lenertová. 2010. Two reflexive markers in Slavic. *Russian Linguistics* 34(3). 203–238. [www.jstor.org/stable/40927420](http://www.jstor.org/stable/40927420).
- Ferreira, Fernanda. 2005. Psycholinguistics, formal grammars, and cognitive science. *The Linguistic Review* 22(2-4). 364–380. DOI: 10.1515/tlir.2005.22.2-4.365.
- Filipan-Žiganić, Blaženka. 2013. *Jezik moje bake i internet: Posuđenice u međimurskom kajkavskom dijalektu – jučer, danas, sutra*. Čakovec: Matica hrvatska Ogranak Čakovec.
- Filipović Đurđević, Dušica. 2021. *Slobodni i otvoreni softver u laboratorijskim i mrežnim istraživanjima u eksperimentalnoj psihologiji*. Belgrade. DOI: 10.5281/zenodo.4748564.
- Ford, Marilyn & Joan Bresnan. 2013. Using convergent evidence from psycholinguistics and usage. In Manfred Krug & Julia Schlüter (eds.), *Research methods in language variation and change*, 295–312. Cambridge: Cambridge University Press.
- Frančić, Anđela, Lana Hudeček & Milica Mihaljević. 2006. *Normativnost i više-funkcionalnost u hrvatskome standardnom jeziku*. Zagreb: Hrvatska sveučilišna naklada.
- Frančić, Anđela & Bernardina Petrović. 2013. *Hrvatski jezik i jezična kultura*. Veleučilište Zaprešić: Baltazar.
- Franks, Steven. 1997. South Slavic clitic placement is still syntactic. *U. Penn Working Papers in Linguistics* 4(2). 111–126. <https://repository.upenn.edu/pwpl/vol4/iss2/8>.
- Franks, Steven. 2000. Clitics at the interface: An introduction to clitics phenomena in European languages. In Frits Beukema & Marcel den Dikken (eds.), *Clitic phenomena in European languages*, 1–46. Philadelphia, PA: John Benjamins Publishing Company.
- Franks, Steven. 2017. *Syntax and spell-out in Slavic*. Bloomington, IN: Slavica.

## References

- Franks, Steven, Uwe Junghanns & Paul Law. 2004. Pronominal clitics in Slavic. *Journal of Slavic Linguistics* 12(1-2). 3–36. <https://www.jstor.org/stable/24599867>.
- Franks, Steven & Tracy H. King. 2000. *A handbook of Slavic clitics*. Oxford: Oxford University Press.
- Franks, Steven & Anita Peti-Stantić. 2006. Splitting puzzles in South Slavic. In *Presented at FDSL 6.5*. University of Nova Gorica.
- Franks, Steven & Ljiljana Progovac. 1994. On the placement of Serbo-Croatian clitics. In *Proceedings of the 9th Biennial Conference on Balkan and South Slavic Linguistics, Literature and Folklore* (Indiana Slavic Studies 69-78).
- Fried, Mirjam. 1994. Second-position clitics in Czech: Syntactic or phonological? *Lingua* 94(2–3). 155–175. DOI: 10.1016/0024-3841(94)90024-8.
- Fried, Mirjam & Jan-Ola Östman. 2004. Construction Grammar: A thumbnail sketch. In Mirjam Fried & Jan-Ola Östman (eds.), *Construction Grammar in a cross-language perspective*, 11–86. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Fukuda, Shin, Grant Goodall, Dan Michael & Henry Beecher. 2012. Is magnitude estimation worth the trouble? In *Proceedings of the 29th West Coast Conference on formal linguistics (Somerville, MA: Cascadilla Proceedings Project)*, 328–336.
- Gadžijeva, Sofija, Ana Kovačević, Milan Mihaljević, Sandra Požar, Johannes Reinhart, Marinka Šimić & Jasna Vince. 2014. *Hrvatski crkvenoslavenski jezik*. Zagreb: Hrvatska sveučilišna naklada/Staroslavenski institut.
- Gato, Maristella. 2014. *Web as Corpus: Theory and practice*. London, New Delhi, New York, Sydney: Bloomsbury.
- George, Leland & Jindřich Toman. 1976. Czech clitics in universal grammar. In Salikoko S. Mufwene, Carol A. Walker & Sanford B. Steever (eds.), *Papers from the 12th regional meeting Chicago linguistic society*, 235–249. Chicago, IL: Chicago Linguistic Society.
- Gibson, Edward & Evelina Fedorenko. 2010. Weak quantitative standards in linguistics research. *Trends in Cognitive Sciences* 14(6). 233–234. DOI: 10.1016/j.tics.2010.03.005.
- Gibson, Edward & Evelina Fedorenko. 2013. The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes* 28(1-2). 88–124. DOI: 10.1080/01690965.2010.515080.
- Gnjatović, Tena & Ranko Matasović. 2013. Some observations on verbs with obligatory control in Croatian. *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje* 39(2). 405–422. <https://hrcaj.srce.hr/117870>.
- Gołąb, Zbigniew. 1964. The problem of verbal moods in Slavic languages. *International Journal of Slavic Linguistics and Poetics* 8. 1–36.



- Goldstein, Bruce. 2010. *Sensation and perception*. Wadsworth: Cengage Learning.
- Golić, Latinka. 1993. *Suvremeni donjomiholjački govor*. Osijek: Gradska tiskara.
- Gordon, Peter C. & Randall Hendrick. 1997. Intuitive knowledge of linguistic co-reference. *Cognition* 62(3). 325–370. DOI: 10.1016/S0010-0277(96)00788-3.
- Gorjanac, Živko. 2011. Neki podaci o šokačkom selu Vršendi u Republici Mađarskoj. 103–114.
- Grewendorf, Günther. 2007. Empirical evidence and theoretical reasoning in generative grammar. *Theoretical Linguistics* 33(3). 369–380. DOI: 10.1515/TL.2007.024.
- Grickat, Irena. 1972. *Aktuelni jezički i tekstološki problemi u starim srpskim ćirilskim spomenicima*. Belgrade: Narodna biblioteka SR Srbije.
- Gries, Stefan Th. 2010. Corpus linguistics and theoretical linguistics. A love-hate relationship? Not necessarily... *International Journal of Corpus Linguistics* 15(3). 327–343. DOI: 10.1075/ijcl.15.3.02gri.
- Gries, Stefan Th. 2013. Basic significance testing. In Robert J. Podesva & Devyani Sharma (eds.), *Research methods in linguistics*, 316–336. Cambridge: Cambridge University Press.
- Gries, Stefan Th. & John Newman. 2013. Creating and using corpora. In Robert J. Podesva & Devyani Sharma (eds.), *Research methods in linguistics*, 257–287. Cambridge: Cambridge University Press.
- Halilović, Senahid. 1990. Govor muslimana Tuholja. In Asim Peco (ed.), *Bosansko-hercegovački dijalektološki zbornik*, vol. 6, 249–358. Sarajevo: Birografika.
- Halilović, Senahid. 1996. Govorni tipovi u međuriječju Neretve i Rijeke dubrovačke. In *Bosanskohercegovački dijalektološki zbornik*, vol. 7, 1–286. Sarajevo: Institut za Jezik u Sarajevu.
- Halilović, Senahid. 2005. Bosanskohercegovački govori. In Svein Mønnesland (ed.), *Jezik u Bosni i Hercegovini*, 15–52. Sarajevo: Institut za Jezik u Sarajevu.
- Halilović, Senahid, Ilijas Tanović & Amela Šehović. 2009. *Govor grada Sarajeva i razgovorni bosanski jezik*. Sarajevo: Biblioteka Bosnistika.
- Ham, Sandra, Jadranka Mlikota, Borko Baraban & Alen Orlić. 2014. *Hrvatski jezični savjeti*. Zagreb: Školska knjiga.
- Hana, Jiřka. 2007. *Czech clitics in higher order grammar*. Charles University in Prague. (Doctoral dissertation).
- Hansen, Björn, Zrinka Kolaković & Edyta Jurkiewicz-Rohrbacher. 2018. Clitic climbing and stacked infinitives in Bosnian, Croatian and Serbian: A corpus-driven study. In Eric Fuß, Marek Konopka, Beata Trawiński & Ulrich H. Waßner (eds.), *Grammar and corpora 2016*, 259–268. Heidelberg: Heidelberg University Publishing (heiUP).

## References

- Hansen, Björn, Daniel Romić & Zrinka Kolaković. 2013. Okviri za istraživanje sintaktičkih struktura govornika druge generacije bosanskoga, hrvatskoga i srpskoga jezika u Njemačkoj. *Lahor: časopis za hrvatski kao materinski, drugi i strani jezik* 1(15). 9–45. <https://hrcak.srce.hr/125643>.
- Hansen, Björn, Veronika Wald & Zrinka Kolaković. 2018. Subjektkasus und Finitheit: Eine korpusbasierte Studie zur Mikrovariation und zur Entwicklung kroatischer Modalkonstruktionen. Teil I: Mikrovariation im modernen Kroatischen. *Zeitschrift für Slavische Philologie* 74(1). 113–195. <https://zsph.winter-verlag.de/article/ZSPH/2018/1/5>.
- Hansen, Björn, Zrinka Kolaković & Edyta Jurkiewicz-Rohrbacher. 2016. Clitic climbing in complement taking predicates in Serbian: A study on clause integration. Variation in space and time: Clausal complementation in South Slavic. In *Clitic climbing in complement taking predicates in Serbian: A study on clause integration. Variation in space and time: Clausal complementation in South Slavic*.
- Heggie, Lorie & Francisco Ordóñez. 2005. Clitic ordering phenomena. The path to generalizations. In Lorie Heggie & Francisco Ordóñez (eds.), *Clitic and affix combinations: Theoretical perspectives*, 1–29. Amsterdam: John Benjamins.
- Hiramatsu, Kazuko. 1999. What syntactic satiation can tell us about islands. *Papers from the Regional Meetings, Chicago Linguistic Society (CLS)* 35. 141–151.
- Hoffmann, Thomas. 2013. Obtaining introspective acceptability judgments. In Manfred Krug & Julia Schlüter (eds.), *Research methods in language variation and change*, 99–118. Cambridge: Cambridge University Press.
- Horvat, Joža. 2018. *Toponimija ludbreške Podravin*. Filozofski fakultet Sveučilišta u Rijeci. (Doctoral dissertation).
- Hoyt, Alexander D. 2012. *Hrvatski jezik u Zagrebu: Sociolingvistički pogled*. Zagreb: Hrvatska sveučilišna naklada.
- Ivić, Milka. 1970. O upotrebi glagolskih vremena u zavisnoj rečenici: Present u rečenici s veznikom *da*. *Zbornik za filologiju i lingvistiku* 13(1). 43–54.
- Ivić, Milka. 1972. Problematika srpskohrvatskog infinitiva. *Zbornik za filologiju i lingvistiku* 15(2). 117–138.
- Ivić, Pavle. 1957. O govoru Galipoljskih Srba. In Aleksandar Belić (ed.), *Srpski dijalektološki zbornik*, vol. 12, 1–520. Belgrade: Naučna knjiga.
- Ivić, Pavle, Ivan Klajn, Mitar Pešikan & Branislav Brborić. 2011. *Srpski jezički priručnik*. Belgrade: Službeni glasnik.
- Ivić, Pavle, Dragolub Petrović & Milorad Radovanović. 2001. *Dijalektologija srpskohrvatskog jezika: Uvod u štokavsko narečje*. Sremski Karlovci; Novi Sad: Izdavačka knjižarnica Zorana Stojanovića.

- Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4). 434–446. DOI: 10.1016/j.jml.2007.11.007.
- Jahić, Dževdad, Senahid Halilović & Ismail Palić. 2000. *Gramatika bosanskoga jezika*. Zenica: Dom štampe.
- Jang, Yoonhee, John T. Wixted & David E. Huber. 2009. Testing signal-detection models of yes/no and two-alternative forced-choice recognition memory. *Journal of Experimental Psychology* 138(2). 291–306. DOI: 10.1037/a0015525.
- Janse, Mark. 1998. Cappadocian clitics and the syntax-morphology interface. In Brian D. Joseph, Geoffrey C. Horrocks & Irene Philippaki-Warbuton (eds.), *Themes in Greek linguistics II*, 106–122. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Jonke, Ljudevit. 1953. O pomicanju enklitike prema početku rečenice. *Jezik: Časopis za Kulturu Hrvatskoga Književnog Jezika* 2(5). 149–150. <https://hrcak.srce.hr/51355>.
- Joseph, Brian D. 1983. *The synchrony and diachrony of the Balkan infinitive: A study in areal, general, and historical linguistics*. Cambridge: Cambridge University Press.
- Jung, Hakyung & Krzysztof Migdalski. 2015. On the degrammaticalization of pronominal clitics in Slavic. In Malgorzata Szajbel-Keck, Roslyn Burns & Darya Kavitskaya (eds.), *Formal approaches to Slavic linguistics: The first Berkeley meeting 2014*, 143–162. Ann Arbor, MI: Michigan Slavic Publications.
- Junghanns, Uwe. 2002. Clitic Climbing im Tschechischen. *Linguistische Arbeitsberichte* 80. 57–90.
- Jurkiewicz-Rohrbacher, Edyta, Björn Hansen & Zrinka Kolaković. 2017. Clitic climbing, finiteness and the raising-control distinction: A corpus-based study. *Jazykovedny časopis* 68(2). 179–190. DOI: 10.1515/jazcas-2017-0028.
- Jurkiewicz-Rohrbacher, Edyta, Zrinka Kolaković & Björn Hansen. 2017. Web corpora – the best possible solution for tracking rare phenomena in under-resourced languages: Clitics in Bosnian, Croatian and Serbian. In Piotr Bański, Marc Kupietz, Harald Lungen, Paul Rayson, Hanno Biber, Evelyn Breiteneder, Simon Clematide, John Mariani, Mark Stevenson & Theresa Sick (eds.), *Proceedings of the workshop on challenges in the management of large corpora and big data and natural language processing (CMLC-5+BigNLP) 2017 including the papers from the web-as-corpus (WAC-XI) guest section. Birmingham, 24 July 2017*, 49–55. Mannheim: Institut für Deutsche Sprache.
- Kalsbeek, Janneke. 2003. Jedno poglavlje iz sintakse istarske čakavštine: Upotreba glagolskih oblika. *Hrvatski dijalektološki zbornik* 12. 103–113. <https://hrcak.srce.hr/168568>.

## References

- Karlík, Petr. 2017a. ANAKOLUT. In Petr Karlík, Marek Nekula & Jana Pleskalová (eds.), *CzechEncy - nový encyklopedický slovník češtiny*. Brno. <https://www.czechency.org/slovník/ANAKOLUT> (18 July, 2019).
- Karlík, Petr. 2017b. APOZIOPEZE. In Petr Karlík, Marek Nekula & Jana Pleskalová (eds.), *CzechEncy - nový encyklopedický slovník češtiny*. Brno. <https://www.czechency.org/slovník/APOZIOPEZE> (18 July, 2019).
- Karlsson, Matti Miestamo, Fred & Kaius Sinnemäki. 2008. Introduction: The problem of language complexity. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change. (Studies in Language Companion Series 94)*, vii–xiv. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Katičić, Radoslav. 1986. *Sintaksa hrvatskoga književnog jezika*. Zagreb: Jugoslavenska akademija znanosti i umjetnosti. Nakladni zavod Globus.
- Kedveš, Ana & Ana Werkmann. 2013. Zanaglasnice u nastavi hrvatskoga jezika – između predodžbe i prakse. *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje* 39(2). 457–473. <https://hrcak.srce.hr/117874>.
- Keller, Frank. 2000. *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. University of Edinburgh. (Doctoral dissertation).
- Kibrik, Andrej A. & Vera I. Podlesskaya. 2003. K sozdaniju korpusov ustnoj ruskoj reči: Principy transkribovanija. *Naučno-tehničkaskaja informacija* 2(10). 5–13. [http://www.philol.msu.ru/~otipl/new/main/people/kibrik-aa/files/Corpora\\_speech\\_transcription@S&I\\_2003.pdf](http://www.philol.msu.ru/~otipl/new/main/people/kibrik-aa/files/Corpora_speech_transcription@S&I_2003.pdf) (21 January, 2018).
- Kibrik, Andrej A. & Vera I. Podlesskaya. 2006. Problema segmentacii ustnogo diskursa i kognitivnaja sistema govorjaščego. In Solov'jeva (ed.), *Kognitivnye issledovanija 1*, vol. 1, 138–158. Moskva: Institut psihologii RAN. [http://iling-ran.ru/kibrik/Segmentation\\_discourse@Cognitive\\_studies\\_2006.pdf](http://iling-ran.ru/kibrik/Segmentation_discourse@Cognitive_studies_2006.pdf) (21 January, 2018).
- Kilgarriff, Adam. 2001. Web as corpus. In Milla Luodonpää-Manni, Esa Penttilä & Johanna Viimaranta (eds.), *Proceedings of the Corpus Linguistics Conference (CL 2001)*. University Centre for Computer Research on Language technical paper, vol. 13, 342–344. UK: Lancaster University.
- Kilgarriff, Adam & Gregory Grefenstette. 2003. Introduction to the special issue to the Web as Corpus. *Computational Linguistics* 29(3). 333–347. DOI: 10.1162/089120103322711569.
- Klaić, Adolf B. 1959. *Bizovačko narječje*. Bizovac: Matica hrvatska.
- Kolaković, Zrinka, Edyta Jurkiewicz-Rohrbacher & Björn Hansen. 2019. Clitic climbing, the raising-control dichotomy and diaphasic variation in Croatian.

- Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje* 45(2). 505–522. DOI: 10.31724/rihjj.45.2.13.
- Kolenić, Ljiljana. 1999. Morfološko-tvorbene osobine ilačkoga govora u okviru slavonskoga dijalekta. *Hrvatski dijalektološki zbornik* 11. 41–47. <https://doaj.org/article/3ecd628209e1416a8f6522ce222792a4>.
- Kolenić, Ljiljana & Anica Bilić. 2004. Mjesni govor Retkovaca. In Anica Bilić (ed.), *Šokačka rič 2. Zbornik radova znanstvenoga skupa slavonski dijalekt*, 5–23. Vinkovci: ZAKUD.
- Kosek, Pavel, Radek Čech & Olga Navrátilová. 2018. Starobyła dativni enklitika *mi, si, ti* ve staročeske bibli 1. Redakce. In Petr Malčík (ed.), *Vesper Slavicus: Sborník k nedožitým devadesátinám prof. Radoslava Večerky*, 137–151. Praha: Nakladatelství Lidové noviny.
- Kostić, Aleksandar. 2003. Elektronski korpus srpskog jezika Đorđa Kostića. *Zbornik Matice srpske za slavistiku* 64. 260–264.
- Krstev, Cvetana & Duško Vitas. 2005. Corpus and lexicon - mutual incompleteness. In Pernilla Danielsson & Martijn Wagenmakers (eds.), *Proceedings of the Corpus Linguistics Conference 14-17 July 2005*. Birmingham. <https://www.birmingham.ac.uk/research/activity/corpus/publications/conference-archives/2005-conf-e-journal.aspx>.
- Krug, Manfred & Katrin Sell. 2013. Designing and conducting interviews and questionnaires. In Manfred Krug & Julia Schlüter (eds.), *Research methods in language variation and change*, 69–98. Cambridge: Cambridge University Press.
- Kučanda, Dubravko. 1998. *Rečenični subjekt u hrvatskom i engleskom jeziku*. Filozofski fakultet Sveučilište u Zagrebu. (Doctoral dissertation).
- Kurtović Budja, Ivana. 2009. *Govori Makarskoga primorja (Fonologija i morfologija)*. Filozofski fakultet Sveučilišta u Zagrebu. (Doctoral dissertation).
- Kuvač Kraljević, Jelena & Gordana Hržica. 2016. Croatian adult spoken language corpus (HrAL). *Fluminensia: Journal for Philological Research* 28(2). 87–102. <https://hrcak.srce.hr/174013>.
- Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26. DOI: 10.18637/jss.v082.i13.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.
- Landau, Idan. 2000. *Elements of control: Structure and meaning in infinitival constructions*. (Studies in natural language and linguistic theory). Dordrecht: Kluwer Academic.

## References

- Landau, Idan. 2004. The scale of finiteness and the calculus of control. *Natural Language & Linguistic Theory* 22. 811–877. DOI: 10.1007/s11049-004-4265-5.
- Landau, Idan. 2013. *Control in generative grammar: A research companion*. Cambridge: Cambridge University Press.
- Langston, Keith & Anita Peti-Stantić. 2014. *Language planning and national identity in Croatia*. New York: Palgrave Macmillan.
- Lau, Jey Han, Alexander Clark & Shalom Lappin. 2017. Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge. *Cognitive Science* 41(5). 1202–1241. DOI: 10.1111/cogs.12414.
- Lenertová, Denisa. 2001. On clitic placement, topicalization and CP-structure in Czech. In Gerhild Zybatow, Uwe Junghanns, Grit Mehlhorn & Lika Szucsich (eds.), *Current issues in formal Slavic linguistics*, 294–305. Berlin: Peter Lang.
- Lenertová, Denisa. 2004. Czech pronominal clitics. *Journal of Slavic Linguistics* 12(1/2). 135–171. <https://www.jstor.org/stable/24599870>.
- Lenth, Russel V. 2016. Least-squares means: The R package lsmeans. *Journal of Statistical Software* 69(1). 1–33. DOI: 10.18637/jss.v069.i01.
- Lešnerová, Šárka & Marko Malink. 2008. Clitic climbing and theta-roles in Upper Sorbian and Czech. In Gerhild Zybatow, Luka Szucsich, Uwe Junghanns & Roland Meyer (eds.), *Formal description of Slavic languages: FDSL 5*, 396–407. Frankfurt am Main: Peter Lang. [philosophy.uchicago.edu/faculty/files/malink/14%20Clitic%20Climbing%202008.pdf](http://philosophy.uchicago.edu/faculty/files/malink/14%20Clitic%20Climbing%202008.pdf) (16 May, 2017).
- Lisac, Josip. 2003. *Hrvatska dijalektologija 1. Hrvatski dijalekti i govori štokavskoga narječja i hrvatski govori torlačkoga narječja*. Zagreb: Golden marketing – Tehnička knjiga.
- Lisac, Josip. 2009. *Hrvatska dijalektologija 2. Čakavsko narječje*. Zagreb: Golden marketing – Tehnička knjiga.
- Lisac, Josip. 2012. *Dvije strane medalje - dijalektološki i jezičnopovijesni spisi o hrvatskom jeziku*. Split: Književni krug.
- Ljubešić, Nikola & Tomaž Erjavec. 2011. HrWaC and slWaC: Compiling web corpora for Croatian and Slovene. In Ivan Habernal & Václav Matoušek (eds.), *Text, speech and dialogue. TSD 2011. Series: Lecture notes in computer science*, vol. 6836, 395–402. Berlin: Springer.
- Ljubešić, Nikola & Tomaž Erjavec. 2016a. Corpus vs. Lexicon supervision in morphosyntactic tagging: The case of Slovene. In *Proceedings of the tenth international conference on Language Resources and Evaluation (LREC 2016)*, 1527–1531. Paris, France: ELRA.
- Ljubešić, Nikola & Tomaž Erjavec. 2016b. *BsWaC v1.2*. <https://www.clarin.si/noske/index.html>.

- Ljubešić, Nikola & Tomaž Erjavec. 2016c. *HrWaC v2.2*. <https://www.clarin.si/noske/index.html>.
- Ljubešić, Nikola & Tomaž Erjavec. 2016d. *SrWaC v1.2*. <https://www.clarin.si/noske/index.html>.
- Ljubešić, Nikola & Filip Klubička. 2014. bs,hr,srWaC – web corpora of Bosnian, Croatian and Serbian. In Felix Bildhauer & Roland Schäfer (eds.), *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*, 29–35. Gothenburg: Association for Computational Linguistics.
- Ljubešić, Nikola, Filip Klubička, Željko Agić & Ivo-Pavao Jazbec. 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)*, 4264–4270. Paris: ELRA.
- Lončarić, Mijo. 1994. O kajkavskoj sintaksi. *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje* 20(1). 137–154. <https://hrcak.srce.hr/69295>.
- Luce, Robert D. 1986. *Response times*. New York: Oxford University Press.
- Mamić, Mile. 1995. Upotreba akuzativnih oblika zamjenice ona. *Jezik: časopis za kulturu hrvatskoga književnog jezika* 43(5). 183–188. DOI: <https://hrcak.srce.hr/204054>.
- Manning, Christopher & Hinrich Schütze. 1999. *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.
- Manning, Christopher D. 2002. Probabilistic syntax. In Rens Bod, Jennifer Hay & Stefanie Jannedy (eds.), *Probabilistic linguistics*, 289–342. Cambridge, MA: MIT Press.
- Marasović-Alujević, Marina & Katarina Lozić Knezović. 2018. *Toponimija otoka Drvenika i Ploče*. Split: Filozofski fakultet u Splitu.
- Maresić, Jela. 2011. Podravski kajkavski dijalekt. *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje* 37(2). 451–466. <https://hrcak.srce.hr/83989>.
- Marinko, Božović & Snežana Bašćarević. 2005. Ratarska leksika Ibarskog Kolašina. *Zbornik radova Filozofskog fakulteta u Prištini* 35. 57–64.
- Marković, Svetozar. 1955. Položaj zamjeničke enklitike u vezi sa naporednom upotrebom infinitiva i prezenta sa svezicom *da*. *Naš jezik* 1-2. 33–40.
- Marušić, Franc & Rok Žaucer. 2014. The involuntary state/FEEL-LIKE construction: What aspect cannot do. *Journal of Slavic Linguistics* 22(2). 185–213. DOI: 10.1353/jsl.2014.0010.
- Mathôt, Sebastiaan, Daniel Schreij & Jan Theeuwes. 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods* 44. 314–324. DOI: 10.3758/s13428-011-0168-7.

## References

- Matuschek, Hannes, Reinhold Kliegl, Shravan Vasishth, R. Harald Baayen & Douglas Bates. 2017. Balancing type I error and power in linear mixed models. *Journal of Memory and Language* 94. 305–315. DOI: 10.1016/j.jml.2017.01.001.
- McEnery, Tony & Andrew Wilson. 1996. *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- Meermann, Anastasia & Barbara Sonnenhauser. 2016. Das Perfekt im Serbischen zwischen slavischer und balkanslavischer Entwicklung. In Alena Bazhutkina & Barbara Sonnenhauser (eds.), *Linguistische Beiträge zur Slavistik. XXII. JungslavistInnen-Treffen in München 12.-14. September 2013*, 83–110. Leipzig: Bibliion Media.
- Meillet, Antoine & André Veillant. 1924. *Grammaire de la langue serbo-croate*. Paris: Champion.
- Menac-Mihalić, Mira. 1989. Glagolski oblici u čakavskom narječju i u hrvatskom književnom jeziku. *Filologija* 17. 81–109. <https://hrcak.srce.hr/165407>.
- Menac-Mihalić, Mira & Anita Celinić. 2012. *Ozvučena čitanka iz hrvatske dijalektologije: Sveučilišni udžbenik s nosačem zvuka*. Zagreb: Knjigra.
- Meyer, Charles F. 2014. Corpus-based and corpus-driven approaches to linguistic analysis: One and the same? In Irma Taavitsainen, Merja Kytö, Claudia Claridge & Jeremy Smith (eds.), *Developments in English: Expanding electronic evidence (Studies in English language)*, 14–28. Cambridge: Cambridge University Press.
- Migdalski, Krzysztof. 2016. *Second position effects in the syntax of Germanic and Slavic languages*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.
- Migdalski, Krzysztof. 2020. The auxiliary ‘be’ as a bundle of phi-features. In Tania Ionin & Jonathan MacDonald (eds.), *Formal approaches to Slavic linguistics: The Urbana-Champaign Meeting 2017*, 246–263. Ann Arbor, MI: Michigan Slavic Publications.
- Migdalski, Krzysztof. 2021. Parametrizing second position effects in Slavic. *Journal of Slavic Linguistics* 28. in print.
- Milićević, Jasmina. 2007. Co-occurrence of Serbian second-position clitics: Syntactic and morphological constraints. In Nabil Nathout & Fabio Montermini (eds.), *Morphologie à Toulouse. Actes du colloque international de morphologie 4e Décembrettes*, 99–121. Munich: Lincom Europa.
- Mišeska Tomić, Olga. 2006. *Balkan sprachbund morpho-syntactic features*. Dordrecht: Springer.
- Mišeska Tomić, Olga. 2008. Variation in clitic-doubling in South Slavic. In Sjeff Barbiers, Olaf Koenenman, Marika Lekakou & Margreet van der Ham (eds.), *Microvariation in syntactic doubling*, 443–468. Bingley: Emerald.



- Mišeska Tomić, Olga. 1996. The Balkan Slavic nominal clitics. In Aaron Halpern & Arnold M. Zwicky (eds.), *Approaching second: Second position clitics and related phenomena*, 511–535. Stanford, CA: Stanford University Press.
- Mladenović, Radivoje. 2010. *Zamenice u govorima jugozapadnog dela Kosova i Metohije*. Belgrade: Institut za srpski jezik SANU.
- Moskovljević, Jasmina. 2008. *Ogledi o glagolskoj potkategorizaciji*. Belgrade: Čigoja.
- Moulton, Erin Elizabeth. 2015. *Rethinking reflexivity: SJA/SE-verbs in Russian and Bosnian/Croatian/Serbian*. University of Kansas. (Doctoral dissertation).
- Mrazović, Pavica & Zora Vukadinović. 2009. *Gramatika srpskog jezika za strance*. Sremski Karlovci, Novi Sad: Izdavačka kuća Zorana Stojanovića.
- Murelli, Adriano. 2011. *Relative constructions in European non-standard varieties*. Berlin: De Gruyter Mouton.
- Myers, James. 2017. Acceptability judgments. In Mark Aronoff (ed.), *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press. DOI: 10.1093/acrefore/9780199384655.013.333.
- Neeleman, Ad & Hans van de Koot. 2006. Syntactic haplology. In Martin Everaert & Henk van Riemsdijk (eds.), *The Blackwell companion to syntax*, vol. IV, 684–710. Oxford: Blackwell Publishers.
- Newmeyer, Frederick J. 1983. *Grammatical theory, its limits and its possibilities*. Chicago, IL: University of Chicago Press.
- Newmeyer, Frederick J. 2007. Commentary on Sam Featherston, 'Data in generative grammar: The stick and the carrot.' *Theoretical Linguistics* 33(3). 395–399. DOI: 10.1515/TL.2007.026.
- Nikolić, Berislav M. 1964. Sremski govor. In Mihailo Stevanović (ed.), *Srpski dijalektološki zbornik*, vol. 14, 201–413. Belgrade: Naučno delo.
- Nikolić, Berislav M. 1966. Mačvanski govor. In Mihailo Stevanović (ed.), *Srpski dijalektološki zbornik*, vol. 16, 181–313. Belgrade: Naučno delo.
- Noonan, Michael. 1985. Complementation. In Timothy Shopen (ed.), *Language typology and syntactic description: Complex constructions*, 42–140. Cambridge: Cambridge University Press.
- Okuka, Miloš. 2008. *Srpski dijalekti*. Zagreb: SKD Prosvjeta.
- Ondrus, Pavel. 1957. Zur Stellung der Enklitika im serbischen Satz. *Zeitschrift für Slawistik* 2(1). 513–522. DOI: 10.1524/slav.1957.2.1.513.
- Pallotti, Gabriella. 2015. A simple view of linguistic complexity. *Second Language Research* 31(1). 117–134. DOI: 10.1177/0267658314536435.
- Pavlović, Slobodan. 2013. Konstituisanje sistema enklitika u srpskom jeziku. *Naučni sastanak slavista u Vukove dane* 1(42). 346–355.

## References

- Peco, Asim. 1985. Govori sjeverne i sjeveroistočne Bosne. Morfologija. In Asim Peco (ed.), *Bosanskohercegovački dijalektološki zbornik*, vol. 5, 201–336. Sarajevo: RO Minerva.
- Peco, Asim. 1990. Govori centralne, jugoistočne i jugozapadne Bosne. Morfološke osobine govora u centralnom pojasu Bosne. In Asim Peco (ed.), *Bosanskohercegovački dijalektološki zbornik*, vol. 6, 125–247. Sarajevo: Birografika.
- Peco, Asim. 2007a. *Izabrana djela I–VI. Knjiga I: Govori istočne i centralne Hercegovine*. Sarajevo: Bosansko filološko društvo.
- Peco, Asim. 2007b. *Izabrana djela I–VI. Knjiga II: Govori zapadne Hercegovine*. Sarajevo: Bosansko filološko društvo.
- Pell, Godfrey. 2005. Use and misuse of Likert scales. *Medical Education* 39(9). 970. DOI: 10.1111/j.1365-2929.2005.02237.x.
- Pešikan, Mitar. 1958. O mestu enklitike u rečenici. *Naš jezik* 9(7-10). 305–311.
- Pešikan, Mitar B. 1965. Starocrnogorski srednjokatunski i lješanski govori. In Mihailo Stevanović (ed.), *Srpski dijalektološki zbornik*, vol. 15, 1–294. Belgrade: Naučno delo.
- Peti-Stantić, Anita. 2005. Sintaksa savremenoga srpskog jezika. In Krešimir Bagić (ed.), *Važno je imati stila*, 165–179. Belgrade: Institut za Srpski Jezik SANU.
- Peti-Stantić, Anita. 2007. Wackernagelovo pravilo – norma ili mogućnost. In Branko Kuna (ed.), *Sintaktičke kategorije*, 173–187. Osijek, Zagreb: Filozofski fakultet/Institut za hrvatski jezik i jezikoslovlje.
- Phillips, Colin. 2010. Should we impeach armchair linguists. *Japanese/Korean Linguistics* 17. 49–64. <https://web.stanford.edu/group/cslipublications/cslipublications/ja-ko-contents/toc-jako17.shtml>.
- Phillips, Colin & Howard Lasnik. 2003. Linguistics and empirical evidence. *Trends in Cognitive Sciences* 7(2). 61–62. DOI: 10.1016/s1364-6613(02)00046-3.
- Piper, Predrag. 2009. O prirodi gramatičkih razlika srpskog i hrvatskog jezika. In Vjara Maldžijeva, Zuzana Topolinjska, Maja Đukanović & Predrag Piper (eds.), *Južnoslovenski jezici: Gramatičke strukture i funkcije*, 538–552. Belgrade: Beogradska knjiga.
- Piper, Predrag, Ivana Antonić, Vladislava Ružić, Sreto Tanasić, Ljudmila Popović & Branko Tošović. 2005. *Sintaksa savremenoga srpskog jezika*. Belgrade: Institut za Srpski Jezik SANU.
- Piper, Predrag & Ivan Klajn. 2014. *Normativna gramatika srpskog jezika*. Novi Sad: Matica srpska.
- Plotnikova, Anna. 1997. Mitološka leksika timočkog govora. *Facta universitatis - series: Linguistics and Literature* 1997 1(4). 267–282. <http://facta.junis.ni.ac.rs/lal/lal97/lal97-04.html>.

- Popović, Ljubomir. 2004. *Red reči u rečenici*. Belgrade: Društvo za srpski jezik i književnost Srbije.
- Progovac, Ljiljana. 1993a. *Clitics in Serbian/Croatian: Comp as the second position*. Wayne State University.
- Progovac, Ljiljana. 1993b. Locality and subjunctive-like complements in Serbo-Croatian. *Journal of Slavic Linguistics* 1(1). 116–144. <https://www.jstor.org/stable/24599042>.
- Progovac, Ljiljana. 1994. *Negative and positive polarity: A binding approach*. Cambridge: Cambridge University Press.
- Progovac, Ljiljana. 1996. Clitics in Serbian/Croatian: Comp as the second position. In Aaron Halpern & Arnold M. Zwicky (eds.), *Approaching second: Second position clitics and related phenomena*, 411–428. Stanford, CA: Stanford University Press.
- Progovac, Ljiljana. 2005. *A syntax of Serbian: Clausal architecture*. Bloomington, IN: Slavica Publishers.
- Przepiórkowski, Adam & Alexandr Rosen. 2005. Czech and Polish raising/control with or without structure sharing. *Research in Language* 3. 33–66. <http://lingo.stanford.edu/sag/L221b/prz-ros.control.pdf> (31 May, 2017).
- Radanović-Kocić, Vesna. 1996. The placement of Serbo-Croatian clitics: A prosodic approach. In Aaron Halpern & Arnold M. Zwicky (eds.), *Approaching second: Second position clitics and related phenomena*, 429–445. Stanford, CA: Stanford University Press.
- Radanović-Kocić, Vesna. 1988. *The grammar of Serbo-Croatian clitics: A synchronic and diachronic perspective*. University of Illinois at Urbana-Champaign. (Doctoral dissertation).
- Radovanović, Dragana. 2006. Morfološke osobenosti govora srednje Kolubare. In Aleksandar Mladenović (ed.), *Srpski dijalektološki zbornik*, 189–374. Belgrade: Srpska akademija nauka i umetnosti i Institut za srpski jezik SANU.
- Raguž, Dragutin. 1997. *Praktična hrvatska gramatika*. Zagreb: Medicinska naklada.
- Raguž, Marija. 2016. Govor Mjesta Bogdánovci. *Život i škola: časopis za teoriju i praksu odgoja i obrazovanja* 62(1). 267–277.
- Reinkowski, Ljiljana. 2001. *Syntaktischer Wandel im Kroatischen am Beispiel der Enklitika*. Munich: Otto Sagner.
- Reinkowski, Ljiljana. 2008. Zum Erlernen der Syntax im Kroatischen: Die Problematik der Enklitika. In Biljana Golubović & Jochen Raecke (eds.), *Bosnisch, Kroatisch, Serbisch als Fremdsprachen an den Universitäten der Welt*, 131–141. Munich: Sagner.

## References

- Remetić, Slobodan. 1985. Govori centralne Šumadije. In Pavle Ivić (ed.), *Srpski dijalektološki zbornik*, 189–374. Belgrade: Srpska akademija nauka i umetnosti i Institut za srpskohrvatski jezik.
- Rescher, Nicholas. 1998. *Complexity: A philosophical overview*. New Brunswick: Transaction.
- Rezac, Milan. 1999. *The syntactic implementation of clitic movement*. University of Toronto. (MA thesis).
- Rezac, Milan. 2005. The syntax of clitic climbing in Czech. In Lorie Heggie & Francisco Ordóñez (eds.), *Clitics and affix combinations. Theoretical perspectives*, 103–140. Amsterdam: John Benjamins Publishing Company.
- Ridjanović, Midhat. 2012. *Bosnian for foreigners with a comprehensive grammar*. Sarajevo: Rabic Publishing Company.
- Riemer, Nick. 2009. Grammaticality as evidence and as prediction in a Galilean linguistics. *Language Sciences* 31(5). 612–633. DOI: 10.1016/j.langsci.2008.04.001.
- Romaine, Suzanne. 1981. On the problem of syntactic variation. *Working Papers in Sociolinguistics* 82. 1–38.
- Rosen, Alexandr. 2001. *A constraint-based approach to dependency syntax applied to some issues of Czech word order*. Prague: Charles University, Faculty of Arts.
- Rosen, Alexandr. 2014. Haplology of reflexive clitics in Czech. In Elżbieta Kaczmarek & Motoki Nomachi (eds.), *Slavic and German in contact: Studies from areal and contrastive linguistics. Slavic Eurasian Studies* 26, 97–116. Hokkaido: Slavic Research Center.
- Rosen, Alexandr & Jiřka Hana. 2017. *Czech clitics: Word order and haplology*. Presentation at Round table on Mechanisms and Constraints on Clitic Climbing, University of Regensburg, Germany.
- Rosenbach, Anette. 2013. Combining elicitation data with corpus data. In Manfred Krug & Julia Schlüter (eds.), *Research methods in language variation and change*, 278–294. Cambridge: Cambridge University Press.
- Ross, John R. 1967. *Constraints on variables in syntax*. MIT. (Doctoral dissertation).
- Runić, Jelena. 2014. *A new look at clitics, clitic doubling, and argument ellipsis: Evidence from Slavic*. University of Connecticut. (Doctoral dissertation).
- Samardžija, Marko. 2015. *Srpsko-hrvatski objasnidbeni rječnik*. Zagreb: Matica hrvatska.
- Sampson, Geoffrey. 1975. *The form of language*. London: Weidenfeld & Nicolson.
- Sampson, Geoffrey. 2001. *Empirical linguistics*. London: Continuum.
- Santos, Diana. 1998. Providing access to language resources through the world wide web: The Oslo corpus of Bosnian texts. In *Proceedings of the first inter-*

- national conference on language resources and evaluation (Granada, 28–30 May 1998)*, 475–481.
- Schäfer, Roland & Felix Bildhauer. 2013. *Web corpus construction*. San Rafael, CA: Morgan & Claypool Publishers.
- Schütze, Carson T. 1994. Serbo-Croatian second position clitic placement and the phonology-syntax interface. In Andrew Carnie, Heidi Harley & Tony Bures (eds.), *Papers on phonology and morphology* (MIT working papers in linguistics 21), 373–473. Cambridge: MIT. <http://www.ai.mit.edu/projects/dm/schutze-secondposition.pdf> (10 August, 2016).
- Schütze, Carson T. 2016. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Berlin: Language Science Press.
- Schütze, Carson T. & Jon Sprouse. 2013. Judgment data. In Robert J. Podesva & Devyani Sharma (eds.), *Research methods in linguistics*, 27–50. Cambridge: Cambridge University Press.
- Sekereš, Stjepan. 1966. Govor našičkog kraja. In Mate Hraste (ed.), *Hrvatski dijalektološki zbornik*, vol. 2, 209–301. Zagreb: Jugoslavenska akademija znanosti i umjetnosti.
- Sekereš, Stjepan. 1977. Govor Hrvata u južnoj Baranji. In Božidar Finka (ed.), *Hrvatski dijalektološki zbornik*, vol. 4, 323–484. Zagreb: Jugoslavenska akademija znanosti i umjetnosti.
- Selting, Margret, Peter Auer, Birgit Barden, Jörg Bergmann, Elizabeth Couper-Kuhlen, Susanne Günthner, Christoph Meier, Uta Quasthoff, Peter Schlobinski & Susanne Uhmann. 1998. Gesprächsanalytisches Transkriptionssystem (GAT). *Linguistische Berichte* 173. 91–122.
- Sgall, Petr, Eva Hajičová & Jarmila Panevová. 1986. *The meaning of the sentence in its semantic and pragmatic aspects*. Dordrecht: Reidel.
- Siewierska, Anna & Ludmila Uhlířová. 1998. An overview of word order in Slavic languages. In Anna Siewierska (ed.), *Constituent order in the languages of Europe*, 105–150. Berlin: Mouton de Gruyter.
- Silić, Josip. 1978. The basic, or grammatical, word order in the Croatian literary language. In Rudolf Filipović (ed.), *Contrastive analysis of English and Serbo-Croatian*, vol. 2, 359–393. Zagreb: Institute of Linguistics. Faculty of Philosophy. University of Zagreb.
- Silić, Josip. 1984. *Od rečenice do teksta (Teoretsko-metodološke pretpostavke nadrečeničnog jedinstva)*. Zagreb: Sveučilišna naklada Liber.
- Silić, Josip. 2006. *Funkcionalni stilovi hrvatskogta jezika*. Zagreb: Dispu.
- Silić, Josip & Ivo Pranjković. 2007. *Gramatika hrvatskoga jezika: Za gimnazije i visoka učilišta*. Zagreb: Školska knjiga.

## References

- Šimík, Radek. 2011. *Modal existential wh-constructions*. University of Groningen. (Doctoral dissertation).
- Šimundić, Mate. 1971. *Govor Imotske krajine i Bekije*. Vol. 11 (Djela akademije nauka i umjetnosti Knjiga 26). Sarajevo: Akademija nauka i umjetnosti Bosne i Hercegovine.
- Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Ślodołowicz, Szymon. 2008. *Control in Polish complement clauses*. Munich: Sagner.
- Smith, Christopher Upham Murray. 2008. *Biology of sensory systems*. Chichester, Hoboken: John Wiley & Sons.
- Snyder, William. 2000. An experimental investigation of syntactic satiation effects. *Linguistic Inquiry* 31(3). 575–582. DOI: 10.1162/002438900554479.
- Sobkowiak, Włodzimierz. 1993. Unmarked-before-marked as a freezing principle. *Language and Speech* 36(4). 393–414. DOI: 10.1177/002383099303600403.
- Spencer, Andrew & Ana R. Luís. 2012. *Clitics: An introduction*. Cambridge: Cambridge University Press.
- Spencer, Nancy J. 1973. Differences between linguists and nonlinguists in intuitions of grammaticality-acceptability. *Journal of Psycholinguistic Research* 2(2). 83–98. DOI: 10.1007/BF01067203.
- Sprouse, Jon & Diogo Almeida. 2012. Assessing the reliability of textbook data in syntax: Adger's core syntax. *Journal of Linguistics* 48(3). 609–652. DOI: 10.1017/S0022226712000011.
- Sprouse, Jon, Carson T. Schütze & Diogo Almeida. 2013. A comparison of informal and formal acceptability judgments using a random sample from *Linguistic Inquiry* 2001–2010. *Lingua* 134. 219–248. DOI: 10.1016/j.lingua.2013.07.002.
- Stanojčić, Živojin & Ljubomir Popović. 2002. *Gramatika srpskog jezika. Udžbenik za I, II, III i IV razred srednje škole*. Belgrade: Zavod za udžbenike i nastavna sredstva.
- Stefanowitsch, Anatol. 2003. Constructional semantics as a limit to grammatical alternation: The two genitives of English. In Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of grammatical variation in English (Topics in English linguistics 43)*, 413–444. Berlin & New York: Mouton de Gruyters.
- Stefanowitsch, Anatol. 2006. Negative evidence and the raw frequency fallacy. *Corpus Linguistics and Linguistic Theory* 2(1). 61–77. DOI: 10.1515/CLLT.2006.003.
- Stefanowitsch, Anatol. 2007. Linguistics beyond grammaticality. *Corpus Linguistics and Linguistic Theory* 3(1). 57–71. DOI: 10.1515/CLLT.2007.004.
- Stefanowitsch, Anatol. 2020. *Corpus linguistics (Textbooks in Language Sciences 7)*. Berlin: Language Science Press. DOI: 10.5281/zenodo.3735822.

- Stevanović, Mihailo. 1950. Đakovački govor. In Aleksandar Belić (ed.), *Srpski dijalektološki zbornik*, vol. 11, 1–152. Belgrade: Naučna knjiga.
- Stevanović, Mihailo. 1975. *Savremeni srpskohrvatski jezik I*. Belgrade: Naučna knjiga.
- Stevanović, Slavica. 1999. Predstavljanje projekta "Korpusno istraživanje upotrebe lokalnih i temporalnih deiktika (spontano) govorenog i (reflektirano) pisanog bosanskog/ hrvatskog/srpskog jezika. *Lingvističke aktuelnosti* 1(2). 86–93.
- Stiebels, Barbara. 2007. Studies in complement control. *ZAS Papers in Linguistics* 47. 1–80. <https://d-nb.info/1060429071/34>.
- Stiebels, Barbara. 2015. Control. In Tibor Kiss & Artemis Alexiadou (eds.), *Syntax - theory and analysis. An international handbook (HSK 42)*, 412–446. Berlin: Mouton de Gruyter.
- Stjepanović, Sandra. 2004. Clitic climbing and restructuring with "finite clause" and infinitive complements. *Journal of Slavic Linguistics* 12(1-2). 173–212. [www.jstor.org/stable/24599871](http://www.jstor.org/stable/24599871).
- Stowe, Laurie & Edith Kaan. 2006. *Developing an experiment: Techniques and design*. Gainesville: University of Florida Digital Collections. <http://ufdc.ufl.edu/IR00005517/00001> (21 September, 2017).
- Tadić, Marko. 1998. Raspon, opseg i sastav korpusa suvremenoga hrvatskoga jezika. *Filologija* 30-31. 337–347. <https://hrcak.srce.hr/173511>.
- Tadić, Marko. 2002. Building the Croatian National Corpus. In Manuel González Rodríguez & Carmen Paz Suarez Araujo (eds.), *Proceedings of the third international Conference on Language Resources and Evaluation (LREC'02)*, vol. II, 441–446. Paris-Las Palmas: ELRA.
- Tadić, Marko. 2020. *Hrvatski nacionalni korpus v3.0 beta*. [http://filip.ffzg.hr/cgi-bin/run.cgi/first\\_form?corpname=HNK\\_v30;align=](http://filip.ffzg.hr/cgi-bin/run.cgi/first_form?corpname=HNK_v30;align=).
- Težak, Stjepko. 1985. Kajkavsko-čakavski utjecaj na štokavske govore Žumberka. *Hrvatski dijalektološki zbornik* 1(7). 255–274. <https://hrcak.srce.hr/file/301064>.
- Težak, Stjepko & Stjepan Babić. 1996. *Gramatika hrvatskoga jezika. Priručnik za osnovno jezično obrazovanje*. Zagreb: Školska knjiga.
- Thompson, Sandra & Elizabeth Couper-Kuhlen. 2005. The clause as a locus of grammar and interaction. *Discourse Studies* 7(4-5). 481–505. DOI: 10.1177/1461445605054403.
- Thorpe, Alana I. 1991. *Clitic placement in complex sentences in Czech*. Providence, RI Brown University. (Doctoral dissertation).
- Todorović, Nataša. 2011. da in the subjunctive and indicative complements. In Katerina Chatzopoulou, Alexandra Ioannidou & Suwon Yoon (eds.), *ICGL 2009 Proceedings*. Chicago: University of Chicago.

## References

- Todorović, Nataša. 2015. *The indicative and subjunctive da-complements in Serbian: A syntactic-semantic approach* (Potsdamer Linguistische Untersuchungen Band 16). Frankfurt am Main: Peter Lang.
- Todorović, Nataša. 2012. *The indicative and subjunctive da-complements in Serbian: A syntactic-semantic approach*. University of Illinois at Chicago. (Doctoral dissertation). <http://indigo.uic.edu/bitstream/handle/10027/9600/Todorovic,%20Nataša.pdf?sequence=1> (16 May, 2017).
- Toury, Gideon. 1995. *Descriptive translation studies and beyond*. Amsterdam: John Benjamins Publishing Company.
- Tummers, Jose, Kris Heylen & Dirk Geeraerts. 2005. Usage-based approaches in cognitive linguistics: A technical state of the art. *Corpus Linguistics and Linguistic Theory* 1(2). 225–261. DOI: 10.1515/cllt.2005.1.2.225.
- Turano, Giuseppina. 2017. Modal particles in Albanian subjunctive, infinitive and supine constructions: Presence vs absence of clitic climbing. *Quaderni di Linguistica e Studi Orientali* 3. 61–86. DOI: 10.13128/QULSO-2421-7220-21339.
- Utvić, Miloš. 2011. Annotating the corpus of contemporary Serbian. *INFOtheca: Journal of informatics & librarianship* XII(2). 36a–47a. <http://jerteh.rs/biblisha/AboutDocument.aspx?DID=1.2011.2.3>.
- van der Auwera, Johan & Vladimir A. Plungian. 1998. Modality's semantic map. *Linguistic Typology* 2(1). 79–124. DOI: <https://doi.org/10.1515/lity.1998.2.1.79>.
- van Marle, Jaap. 1997. Dialect versus standard language: Nature versus culture. In Jenny Cheshire & Dieter Stein (eds.), *Taming the vernacular: From dialect to written standard language*, 13–34. London, New York: Longman.
- Vitas, Duško, Cvetana Krstev & Gordana Pavlović-Lažetić. 2000. Recent results in Serbian computational lexicography. In Neda Bokan (ed.), *Proceedings of the symposium "Contemporary Mathematics", 18–20 December 1998*, 8. Belgrade: University of Belgrade.
- von Waldenfels, Ruprecht & Maciej Eder. 2016. A stylometric approach to the study of differences between standard variants of Bosnian/Croatian/Serbian, or: Is the Hobbit in Serbian more Hobbit or more Serbian? *Russian Linguistics* 40(1). 11–31. DOI: 10.1007/s11185-015-9155-4.
- Vranić, Silvana. 2003. Iz sintakse paških čakavskih govora. *Hrvatski dijalektološki zbornik* 12. 151–167. <https://hrcak.srce.hr/file/248819>.
- Vuković, Teodora. 2021. *Representing variation in a spoken corpus of an endangered dialect: The case of Torlak*. 10.1007/s10579-020-09522-4.
- Vukša Nahod, Perina. 2014. *Fonologija i morfologija mjesnih govora Slivanjskoga područja*. Filozofski fakultet Sveučilišta u Rijeci. (Doctoral dissertation).
- Vuletić, Frane. 1890. *Gramatika bosanskoga jezika za srednje škole*. Sarajevo: Zemaljska vlada BiH.



- Vuletić, Nikola & Vladimir Skračić. 2018. *Leksik morske faune u sjevernoj Dalmaciji. Prvi dio: Rječnik pučkoga nazivlja*. Zadar: Sveučilište u Zadru.
- Vulić, Sanja. 2009. Iz današnjega govora šokačkih Hrvata u Rekašu u Rumunjskoj. *Šokačka rič 6: Zbornik radova Znanstvenoga skupa Slavonski dijalekt 6* 6. 161–175.
- Wackernagel, Jakob. 1892. Über ein Gesetz der indogermanischen Wortstellung. *Indogermanische Forschungen* 1(1). 333–436. DOI: 10.1515/9783110242430.333.
- Walkden, George, Christina Sevdali & Morgan Macleod. 2020. *On a law of Indo-European word order: Über ein Gesetz der indogermanischen Wortstellung*. Berlin: Language Science Press. DOI: 10.5281/zenodo.3978908.
- Walker, James. 2010. *Variation in linguistic systems*. New York, London: Routledge.
- Walker, James. 2013. Variation analysis. In Robert Podesva & Devyani Sharma (eds.), *Research methods in linguistics*, 440–459. Cambridge: Cambridge University Press.
- Wallis, Sean. 2014. *Coping with imperfect data*. corplingstats.wordpress.com/2014/04/10/imperfect-data/ (31 January, 2017).
- Wasow, Thomas & Jennifer Arnold. 2005. Intuitions in linguistic argumentation. *Lingua* 115(11). 1481–1496. DOI: 10.1016/j.lingua.2004.07.001.
- Weber, Adolfo. 1859. *Skladnja ilirskoga jezika, za niže gimnazije*. Reprint from 2005 made by Institut za Hrvatski jezik i jezikoslovlje. Vienna: Carska kraljevska naklada školskih knjigah.
- Weber, Ernst Heinrich. 1834. *De pulsū, resorptione, auditu et tactu. Annotationes anatomicae et physiologicae*. Lipsiae: Koehler.
- Weskott, Thomas & Gisbert Fanselow. 2011. On the informativity of different measures of linguistic acceptability. *Language* 87(2). 249–273. DOI: 10.1353/lan.2011.0041.
- Wiemer, Björn & Björn Hansen. 2012. Assessing the range of contact-induced grammaticalization in Slavonic. In Björn Wiemer, Bernhard Wälchli & Björn Hansen (eds.), *Grammatical replication and borrowability in language contact. (Trends in Linguistics. Studies and Monographs 242)*, 67–155. Berlin: Mouton de Gruyter.
- Wilder, Chris & Damir Ćavar. 1994a. Long head movement? Verb movement and cliticization in Croatian. *Lingua* 93(1). 1–58. DOI: 10.1016/0024-3841(94)90352-2.
- Wilder, Chris & Damir Ćavar. 1994b. Word order variation, verb movement and economy principles. *Studia Linguistica* 48(1). 46–86. DOI: 10.1111/j.1467-9582.1994.tb00849.x.

## References

- Wulff, Stefanie. 2003. A multifactorial corpus analysis of adjective order in English. *International Journal of Corpus Linguistics* 8(2). 245–282. DOI: 10.1075/ijcl.8.2.04wul.
- Wurmbrand, Susi. 1999. Modal verbs must be raising verbs. In Sonya Bird, Andrew Carnie, Jason D. Haugen & Peter Norquest (eds.), *Proceedings of the 18th West Coast Conference on Formal Linguistics (WCCFL 18)*, 599–612. Somerville, MA: Cascadilla Press.
- Yeasmin, Sabina & Khan Ferdousour Rahman. 2012. “Triangulation” research method as the tool of social science research. *BUP Journal* 1(1). 154–163.
- Zaliznjak, Andrej Anatol’evič. 1993. K izučeníju jazyka berestjanyh gramot. In Valentin Lavrent’evič Janin & Andrej Anatol’evič Zaliznjak (eds.), 191–321. Moskva: Nauka.
- Zasina, Adrian J., David Lukeš, Zuzana Komrsková, Petra Poukarová & Anna Řehořková. 2018. *Koditex: Korpus diverzifikovaných textů*. Prague. <http://www.korpus.cz>.
- Zec, Draga & Dušica Filipović-Đurđević. 2017. The role of prosody in clitic placement. In Vera Griбанова & Stephanie Shih (eds.), *The morphosyntax-phonology connection: Locality and directionality at the interface*, 171–196. New York: Oxford University Press.
- Zec, Draga & Sharon Inkelas. 1990. Prosodically constrained syntax. In Draga Zec & Sharon Inkelas (eds.), *The phonology-syntax connection*, 365–378. Chicago, IL: University of Chicago Press.
- Zimmerling, Anton & Peter Kosta. 2013. Slavic clitics: A typology. *STUF: Language Typology and Universals* 66(2). 178–214. DOI: 10.1524/stuf.2013.0009.
- Zwicky, Arnold M. 1977. *On clitics*. Bloomington, IN: Indiana University Linguistics Club.
- Zwicky, Arnold M. 1994. What is a clitic? In Joel A. Nevis, Brian D. Joseph, Dieter Wanner & Arnold M. Zwicky (eds.), *Clitics: A comprehensive bibliography (1892–1991)*, xii–xx. Amsterdam: John Benjamins Publishing Company.

# Name index

- Abbuhl, Rebekha, 341, 345  
Adamovičová, Ana, 291  
Aladrović, Katarina, 145, 148, 155  
Alexander, Ronelle, 28, 113–115, 119,  
124, 174  
Aljović, Nadira, 14, 225, 229–232,  
252, 254, 274, 275, 277, 279,  
281, 299, 325, 365, 384, 389,  
394, 395  
Almeida, Diogo, 57, 348  
Angouri, Jo, 54  
Arnold, Jennifer, 57, 348  
Avgustinova, Tania, 27
- Baayen, R. Harald, 323, 335, 354, 355  
Babić, Stjepan, 92, 95, 97, 100, 104–  
108, 111, 115, 120, 124, 125,  
194  
Babić, Tena, 142, 167  
Bader, Markus, 66  
Balvet, Antonio, 75  
Bangerter, Adrian, 51, 58  
Barac-Grum, Vida, 99, 111  
Barić, Eugenija, 6, 44, 92, 93, 95–100,  
103–106, 109, 110, 113, 117–  
120, 124, 194  
Barjaktarević, Danilo, 141, 143, 147,  
167  
Baroni, Marco, 77  
Barr, Dale J., 355  
Baščarević, Snežana, 132  
Bates, Douglas, 323, 354, 355
- Benko, Vladimír, 77, 79  
Bermel, Neil, 389  
Biber, Douglas, 57, 79, 84  
Bierwisch, Manfred, 43  
Bildhauer, Felix, 79, 81, 84  
Bilić, Anica, 163  
Birtić, Matea, 341  
Birzer, Sandra, 181  
Blaikie, Norman, 63  
Bošković, Željko, 11, 18–20, 23, 25, 29,  
121, 151, 226, 275, 281  
Božović, Đorđe, 7  
Brabec, Ivan, 98, 111, 114  
Bresnan, Joan, 54, 57  
Brewer, Marilyn B., 55  
Brlobaš, Željka, 145, 146, 160  
Browne, Wayles, 26, 40, 41, 298, 305  
Brozović, Dalibor, 128, 138, 145, 147,  
148, 162  
Brozović-Rončević, Dunja, 72, 74  
Brysbart, Marc, 335  
Buchstaller, Isabelle, 53  
Bukumirić, Mileta, 137, 141, 144, 145
- Cacoullou, Rena T., 312  
Caink, Andrew, 225  
Čamdžić, Amela, 225, 230  
Ćavar, Damir, 5, 24–29, 72, 74, 121,  
226, 230, 237, 240–242, 245,  
298, 305, 389  
Čedić, Ibrahim, 115  
Celinić, Anita, 133, 134

*Name index*

- Čermák, František, 291, 292  
Chafe, Wallace, 181  
Chomsky, Noam, 60  
Čirković, Svetlana, 146  
Clark, Herbert H., 51, 58, 67  
Čolak, Majda, 312  
Coseriu, Eugenio, 14  
Couper-Kuhlen, Elizabeth, 177, 179  
Cowart, Wayne, 51, 52, 61, 341, 342,  
345, 346, 349  
Crible, Ludivine, 179, 180, 183  
Ćurković, Dijana, 136, 138, 139, 146–  
150, 152, 166  
Cvrček, Vaclav, 79  
  
Dąbrowska, Ewa, 51, 52, 62, 63, 348  
Davies, Mark, 312  
Davies, William D., 36  
de Andrade, Arpldo L., 312, 313  
Derwing, Bruce L., 59, 60  
Devitt, Michael, 348  
Dickey, Stephen, 334  
Diesing, Molly, 5, 8, 25, 26, 205, 214,  
412, 413  
Divjak, Dagmar, 389  
Dobrić, Nikola, 75  
Donders, Franciscus Cornelis, 59  
Dotlačil, Jakub, 225, 228, 229, 240,  
248–251, 253–255, 258–260,  
262, 276–281, 299, 365, 381,  
386  
Dragičević, Milan, 137, 145, 146  
Dubinsky, Stanley, 36  
Đukanović, Vlado, 41  
  
Eder, Maciej, 4  
Egbert, Jesse, 79  
Eisenbeiss, Sonja, 65, 66  
Erjavec, Tomaž, 74, 77, 78  
  
Fanselow, Gisbert, 51, 66, 348  
Farkaš, Loretana, 142, 167  
Featherston, Sam, 51, 61, 62  
Fechner, Gustav Theodor, 64  
Fedorenko, Evelina, 348  
Fehrmann, Dorothee, 42–48  
Ferreira, Fernanda, 348  
Filipan-Žigniće, Blaženka, 133  
Filipović Đurđević, Dušica, 347  
Filipović-Đurđević, Dušica, 8, 25, 26,  
413  
Ford, Marilyn, 54  
Frančić, Anđela, 95, 97, 111, 113, 115,  
120, 313  
Franks, Steven, 4, 7, 13, 17, 19, 23–25,  
27, 31, 115, 118, 121, 122, 124,  
193, 226, 234, 238, 298, 392,  
411  
Fried, Mirjam, 23, 36  
Fukuda, Shin, 66  
  
Gadžijeva, Sofija, 150  
Gato, Maristella, 79, 81, 84  
George, Leland, 253, 254, 260, 261,  
279  
Gibson, Edward, 348  
Gnjatović, Tena, 347  
Gołąb, Zbigniew, 40  
Goldstein, Bruce, 64  
Golić, Latinka, 145, 152, 157  
Gordon, Peter C., 348  
Gorjanac, Živko, 159  
Grefenstette, Gregory, 58, 84  
Grewendorf, Günther, 51  
Grickat, Irena, 20, 150  
Gries, Stefan Th., 56–58, 62  
  
Halilović, Senahid, 137–139, 142, 148,  
152, 154, 162

- Ham, Sandra, 6  
Hana, Jiřka, 6, 22, 225, 228–230, 253,  
254, 258, 262, 265–267, 270,  
271, 279–281, 299, 365, 367,  
369, 386  
Hansen, Björn, 33, 35, 41, 47, 86, 257,  
262, 291, 297, 299, 300, 308,  
311, 314, 325, 327, 329, 333,  
365, 394, 396, 405, 418  
Häussler, Jana, 66  
Heggie, Lorie, 18  
Hendrick, Randall, 348  
Hiramatsu, Kazuko, 52  
Hoffmann, Thomas, 54, 62–65, 352,  
389  
Horvat, Joža, 133  
Hoyt, Alexander D., 136, 146  
Hrřica, Gordana, 82  
Hudson, Richard, 225, 230  
  
Inkelas, Sharon, 28–30  
Ivić, Milka, 40, 41  
Ivić, Pavle, 111, 112, 128, 129, 133, 135,  
141, 144, 146, 149–151, 156,  
159, 167, 168  
  
Jaeger, T. Florian, 354  
Jahić, Dževdad, 92, 95–97, 100, 103,  
104, 106, 107, 109, 110, 113,  
120, 194  
Jang, Yoonhee, 64  
Janse, Mark, 34  
Jonke, Ljudevit, 113  
Joseph, Brian D., 395  
Jung, Hakyung, 23  
Junghanns, Uwe, 21, 33, 165, 225, 226,  
228, 229, 234, 235, 238–253,  
263–265, 267, 272, 275–281,  
288, 330, 367, 371, 386, 387  
  
Jurkiewicz-Rohrbacher, Edyta, 35,  
86, 257, 262, 291, 297, 299,  
300, 308, 311, 314, 325, 327,  
329, 333, 365, 394, 405, 418  
  
Kaan, Edith, 341, 345  
Kalsbeek, Janneke, 163  
Karlík, Petr, 182  
Karlsson Fred, Matti Miestamo, 390  
Katićić, Radoslav, 44, 46, 96, 97, 104–  
106, 110, 111, 113, 116, 120, 124  
Kedveš, Ana, 174  
Keller, Frank, 51  
Khattab, Ghada, 53  
Kibrik, Andrej A., 176, 179, 181  
Kilgarriř, Adam, 58, 84  
King, Tracy H., 7, 13, 17, 19, 27, 31, 124,  
193, 226, 234, 238, 298, 392  
Klaić, Adolf B., 154  
Klajn, Ivan, 92–98, 100, 101, 103, 104,  
106–113, 115, 117, 118, 124,  
125, 191, 194, 218, 413, 415  
Kliegl, Reinhold, 323, 355  
Klubićka, Filip, 71, 77, 79–81  
Knezović, Katarina Lozić, 133  
Knittl, Luděk, 389  
Kolaković, Zrinka, 41, 47, 86, 257,  
262, 291, 297, 299, 300, 308,  
311, 314, 325, 327, 329, 333,  
365, 394, 405, 418  
Kolenić, Ljiljana, 148, 163  
Kosek, Pavel, 30, 31, 90, 187, 199, 201,  
221, 417  
Kosta, Peter, 18, 20, 27–29, 34, 150  
Kostić, Aleksandar, 75  
Krstev, Cvetana, 75  
Krug, Manfred, 64, 65, 346, 348, 349,  
351, 352, 414  
Kućanda, Dubravko, 46, 47

## *Name index*

- Kurtović Budja, Ivana, 145  
Kuvač Kraljević, Jelena, 82  
Kuznetsova, Alexandra, 354
- Labov, William, 53  
Landau, Idan, 7, 37, 38  
Langston, Keith, 313  
Lasnik, Howard, 51  
Lau, Jey Han, 63, 64  
Lenertová, Denisa, 7, 225, 253, 255,  
256, 262, 264, 270  
Lenth, Russel V., 354  
Lešnerová, Šárka, 262, 280  
Lisac, Josip, 128, 131–137, 141, 143–  
145, 148, 149, 160–163, 168  
Ljubešić, Nikola, 71, 74, 77–81  
Lončarić, Mijo, 145, 146, 159, 160  
Luce, Robert D., 59  
Luís, Ana R., 11, 13, 17, 19, 22, 228, 229
- Mächler, Martin, 354  
Malink, Marko, 262, 280  
Mamić, Mile, 94  
Manning, Christopher, 71, 84, 85  
Manning, Christopher D., 389  
Marasović-Alujević, Marina, 133  
Maresić, Jela, 147  
Marinko, Božović, 132  
Marković, Svetozar, 298–300, 396  
Marušić, Franc, 47  
Matasović, Ranko, 347  
Mathôt, Sebastiaan, 346  
Matuschek, Hannes, 355  
McEnery, Tony, 69  
Meermann, Anastasia, 20, 105  
Meillet, Antoine, 114  
Menac-Mihalić, Mira, 133, 134, 145  
Meyer, Charles F., 57  
Migdalski, Krzysztof, 7, 20, 23
- Milićević, Jasmina, 93, 101, 102  
Milin, Petar, 354  
Mišeska Tomić, Olga, 19, 20, 118, 119,  
121, 122, 167  
Mladenović, Radivoje, 138, 139, 141,  
144, 146  
Moskovljević, Jasmina, 37  
Moulton, Erin Elizabeth, 42, 45, 46  
Mrazović, Pavica, 93–103, 112, 113,  
124  
Murelli, Adriano, 313  
Myers, James, 55, 56, 59, 60, 332
- Neeleman, Ad, 21  
Newman, John, 56, 58  
Newmeyer, Frederick J., 51, 61, 348  
Nikitina, Tatiana, 57  
Nikolić, Berislav M., 148–151, 155,  
156, 158, 159, 162, 164–166  
Noonan, Michael, 34
- Okuka, Miloš, 31, 128, 129, 133–135,  
138–141, 143, 144, 146–153,  
155, 156, 158–164
- Oliva, Karel, 27  
Ondrus, Pavel, 102  
Ordóñez, Francisco, 18  
Östman, Jan-Ola, 36
- Pallotti, Gabriella, 390  
Pavlović, Slobodan, 20  
Peco, Asim, 137, 139–147, 151, 153, 154,  
156, 157, 164  
Pell, Godfrey, 63  
Pešikan, Mitar, 111–114, 117, 118, 120,  
121, 123, 414  
Pešikan, Mitar B., 140, 143, 145, 150,  
151, 158  
Peti-Stantić, Anita, 11, 115, 122, 123,  
205, 313

- Petrović, Bernardina, 95, 120  
 Phillips, Colin, 51  
 Piper, Predrag, 15, 16, 92–98, 100, 101, 103, 104, 106–113, 115, 117, 118, 123–125, 191, 194, 218, 413, 415  
 Plotnikova, Anna, 132  
 Plungian, Vladimir A., 47  
 Podlesskaya, Vera I., 176, 179, 181  
 Popović, Ljubomir, 47, 92, 95, 97, 98, 100, 102, 107, 109, 110, 113, 115, 120, 122, 123, 125  
 Pranjković, Ivo, 44, 92, 95, 116, 125, 404  
 Progovac, Ljiljana, 5, 11, 19, 23, 24, 38, 39, 45, 116, 118, 119, 122, 158, 170, 206, 220, 226, 230, 232, 233, 298, 389, 417  
 Przepiórkowski, Adam, 35  
 Radanović-Kocić, Vesna, 11, 24, 29, 93, 94, 98, 112, 114–117, 119, 121, 122, 158, 162, 170, 200, 206, 207, 220, 226, 417  
 Radovanović, Dragana, 136, 145, 146  
 Raguž, Dragutin, 111  
 Raguž, Marija, 154  
 Rahman, Khan Ferdousour, 53  
 Reinkowski, Ljiljana, 11, 28, 29, 114, 214  
 Remetić, Slobodan, 144, 145  
 Rescher, Nicholas, 390–392, 419  
 Rezac, Milan, 225, 229–231, 234, 237, 240, 251, 253–256, 258–260, 272, 273, 277, 279, 281, 299, 307, 309, 365, 367, 381, 386  
 Ridjanović, Midhat, 22, 23, 92, 97, 99–101, 103–105, 109, 118, 122–124, 191, 194, 195, 217, 218, 250, 416  
 Riemer, Nick, 61  
 Romaine, Suzanne, 14  
 Rosen, Alexandr, 6, 22, 35, 225, 234, 240, 264–268, 270, 280, 281, 291, 292, 330, 374, 406, 419  
 Rosenbach, Anette, 54, 64, 65, 353  
 Ross, John R., 238, 393  
 Runić, Jelena, 23  
 Samardžija, Marko, 15  
 Sampson, Geoffrey, 53, 57  
 Santos, Diana, 71, 72  
 Schäfer, Roland, 79, 81, 84  
 Schütze, Carson T., 11, 25, 28, 29, 51–53, 60–65, 122, 226, 342, 346, 350  
 Schütze, Hinrich, 71, 84, 85  
 Seiss, Melanie, 25–27  
 Sekereš, Stjepan, 136, 155, 157  
 Sell, Katrin, 64, 65, 346, 348, 349, 351, 352, 414  
 Selting, Margret, 176  
 Sgall, Petr, 27  
 Siewierska, Anna, 197  
 Silić, Josip, 44, 92, 95, 116, 125, 174, 404  
 Šimik, Radek, 251  
 Šimundić, Mate, 136, 138, 146, 152, 153, 164  
 Sinclair, John, 70  
 Sinnemäki, Kaius, 390  
 Skračić, Vladimir, 133  
 Słodowicz, Szymon, 37, 229, 230  
 Smith, Christopher Upham Murray, 64  
 Snyder, William, 52  
 Sobkowiak, Włodzimierz, 31

## *Name index*

- Sonnenhauser, Barbara, 20, 105  
Spencer, Andrew, 11, 13, 17, 19, 22,  
228, 229  
Spencer, Nancy J., 348  
Sprouse, Jon, 51, 52, 57, 60–65, 342,  
346, 348, 350  
Stanojčić, Živojin, 47, 92, 95, 98, 107,  
109, 123, 125  
Stefanowitsch, Anatol, 30, 31, 57, 58,  
413, 414  
Stevanović, Mihailo, 92, 94, 95, 103,  
104, 138, 141, 144, 151, 161,  
167  
Stevanović, Slavica, xii, 82, 175  
Stiebels, Barbara, 34, 37, 38  
Stjepanović, Sandra, 5, 33, 40, 225,  
230–233, 272, 275, 281, 298,  
307, 308, 384, 389  
Stowe, Laurie, 341, 345  
  
Tadić, Marko, 72–74  
Težak, Stjepko, 92, 95, 97, 100, 104–  
108, 115, 124, 125, 146, 194  
Thompson, Sandra, 177, 179  
Thorpe, Alana I., 253, 279, 386  
Todorović, Nataša, 38, 41, 225, 232,  
233, 274, 281, 298, 306  
Toman, Jindřich, 253, 254, 260, 261,  
279  
Toury, Gideon, 71  
Tummers, Jose, 56  
Turano, Giuseppina, 395  
  
Uhlířová, Ludmila, 197  
Utvić, Miloš, 75  
  
van de Koot, Hans, 21  
van der Auwera, Johan, 47  
van Marle, Jaap, 313  
  
Vasishth, Shraavan, 323, 355  
Vavřín, Martin, 291  
Veillant, André, 114  
Vitas, Duško, 75  
von Waldenfels, Ruprecht, 4  
Vranić, Silvana, 146, 152, 153  
Vukadinović, Zora, 93–103, 112, 113,  
124  
Vuković, Teodora, 83  
Vukša Nahod, Perina, 137–140, 146,  
157, 162, 163, 165  
Vuletić, Frane, 92  
Vuletić, Nikola, 133  
Vulić, Sanja, 161  
  
Wackernagel, Jakob, 3  
Wald, Veronika, 41, 47  
Walkden, George, 3  
Walker, James, 13–15  
Wallis, Sean, 300  
Wasow, Thomas, 57, 348  
Weber, Adolfo, 113  
Weber, Ernst Heinrich, 64  
Werkmann, Ana, 174  
Weskott, Thomas, 66  
Wiemer, Björn, 396  
Wilder, Chris, 5, 24, 25, 28, 29, 121,  
226, 230, 237, 240–242, 245,  
298, 305, 389  
Wilson, Andrew, 69  
Wulff, Stefanie, 31  
Wurmbrand, Susi, 36, 232  
  
Yeasmin, Sabina, 53  
  
Zaloznjak, Andrej Anatol'evič, 27  
Zasina, Adrian J., 86  
Žaucer, Rok, 47  
Zec, Draga, 8, 25, 26, 28–30, 214, 413



Zimmerling, Anton, 18, 20, 27–29,  
34, 150

Zwicky, Arnold M., 3, 4, 12, 17





# Clitics in the wild

This collective monograph is the first data-oriented, empirical in-depth study of the system of clitics on Bosnian, Croatian and Serbian. It fills the gap between the theoretical and normative literature by including solid data on variation found in dialects and spoken language and obtained from massive Web Corpora and speakers' acceptability judgements. The authors investigate three primary sources of variation: inventory, placement and morphological processes. A separate part of the book is dedicated to the phenomenon of clitic climbing, the major challenge for any syntactic theory. The theory of complexity serves as the explanation for the very diverse constraints on clitic climbing established in the empirical studies. It allows to construct a series of hierarchies where the factors relevant for predicting clitic climbing interact with each other. Thus, the study pushes our understanding of clitics away from fine-grained descriptions and syntactic generalisations towards a probabilistic modelling of syntax.

