Serbo-Croatian Word Order: A Logical Approach

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Vedrana Mihalicek, B.A., M.A.

Graduate Program in Department of Linguistics

The Ohio State University

2012

Dissertation Committee:

Carl Pollard, Advisor
Brian Joseph
Michael White
© Copyright by
Vedrana Mihalicek
2012
Abstract

This dissertation presents a formal theory of Serbo-Croatian grammar. The theory predicts acceptable form/meaning pairs for a substantial chunk of Serbo-Croatian. In particular, we analyze Serbo-Croatian declarative and interrogative main clauses, embedded clauses, a couple of different types of nominal modification, control and predication, as well pro- and encliticization.

Linguistic expressions are represented as triples of typed terms, with each typed term modeling one of the following sets of properties of a linguistic sign: semantic (i.e. truth-conditional meaning), tectogrammatical (i.e. syntactic combinatorial properties), and, finally, phenogrammatical properties which specify the expression’s linearization possibilities.

The focus of our work is on word order in Serbo-Croatian, which is very free in some respects but extremely rigid in others. With phenogrammar and tectogrammar as distinct components, we can isolate word order problems from tectogrammatical and semantic combination, and state theory-internal phenogrammatical generalizations. This is particularly important for the analysis of 2P enclitics, whose placement cannot be adequately characterized tectogrammatically.
The most elaborate component is phenogrammar. We postulate many different phenogrammatical types and modes of combination. This enables us to create islands of fixed word order, while still allowing free reordering of higher-level phenogrammatical objects. Of special significance are phenogrammatical terms which denote sets of strings. Such terms represent possible pronunciations of expressions which can be linearized multiple ways without a change in meaning. Essentially, we are modeling semantically insignificant reordering as phenogrammatical indeterminacy.

Our choice of grammatical architecture is empirically motivated, but methodological in nature. This dissertation purports to show that a decent theory of Serbo-Croatian word order can be given in a framework which does distinguish phenogrammar and tectogrammar; not that a comparable (or superior) theory cannot be given in one which does not. We do, however, believe that this project brings attention to languages with complex word order patterns and their role in delineating a realistic categorial grammar framework for linguistic analysis.
Acknowledgments

A year ago I was sure that I would never complete my thesis. Luckily, I was surrounded by people who had more confidence in me and this project than I did, and who offered practical help, support and encouragement, as well as much needed respite.

I am very grateful to my advisor, Carl Pollard, for not giving up on me, despite my repeated threats to abandon linguistics and graduate school altogether. Over the years, he provided tons of encouragement, practical suggestions and advice. Really, most of what I learned in graduate school I can somehow trace back to Carl’s involvement: courses, conversations, emails, questions that I had to figure out how to answer. Even (or, especially!) frustrating interactions with Carl ultimately lead to my intellectual and personal growth, and I still wonder to what extent these were orchestrated by Carl, for pedagogical purposes.

Brian Joseph has also been my ally for many years, encouraging me to present at conferences, sticking up for me through funding mishaps and always being a sympathetic ear for my graduate school woes. I’m sure he’s supported me behind the scenes in ways I’m not even aware of.
While I barely knew Mike White prior to him joining my thesis committee, I was very lucky that he agreed to take on this role. He was an invaluable asset, whose thoughtful and incisive comments significantly improved my thesis in many ways, and more importantly, forced me to examine and clarify my own attitudes and assumptions concerning various conceptual issues in linguistic theory. His no-nonsense attitude was much appreciated too, and, in retrospect, I wish I had included him in my graduate career earlier.

I am also very grateful to the Linguistics Department staff for their technological (thanks Jim!) and administrative (thanks Julia!) help. A very special thanks goes to Claudia Morettini, Hope Dawson, Joanna Anderson and Julie McGory, who have been my friends and allies for many years, and tremendously helped make my graduate experience pleasant and relatively smooth, whether by offering administrative and teaching support, scrubbing my disgusting tea mug, or having liquid lunches when needed.

I’ve also had the privilege of having many outstanding colleagues in the department over the years, who offered stimulating conversation, friendship and support. Thank you Dahee Kim, Bridget Smith, Scott Martin, Chris Worth, Lia Mansfield, Rachel Klippenstein, Na’im Tyson, and many others, for office parties, rides, cat sitting, sporcle competitions, happy hours, a hand-drawn map of Middle Earth, and all kinds of other good times which I don’t remember; that’s how good they were.
My parents, Jelena and Zlatko Mihaliček, always encouraged me to be independent and do my own thing, and although they couldn’t necessarily understand what I was doing or why, remained confident that I could do it nonetheless. I have always been taught that the only thing I have to do is die, everything else being a matter of choice. Had this not been instilled in me, I doubt very much that I ever would have had the courage to embark on my international education adventures over a decade ago.

I also never would have been able to pursue education abroad without the generous and unconditional financial support of the United World College program, and the Wien Scholarship program at Brandeis University. Along the way, I’ve had many teachers who believed in me, inspired me and helped me develop an appreciation for the intellectual life, particularly Margaret and Eric Mace-Tessler at UWC of the Adriatic, and Ray Jackendoff, Jerry Samet and Jeffrey Poland at Brandeis. I would not be here today were it not for their influence.

Last but not least, it’s unlikely that I would have been able to push through the past couple of years without Bart. He offered help (ranging from housework, to typesetting, to mathematical inspiration), listened to me and tolerated my erratic behavior. His own courage and refusal to be hampered by fear of failure is something I hope to be able to emulate one day. Thank you, Bart, for your affection and companionship.
Vita

2010 ............................................. M.A. The Ohio State University

2007-present ................................. Graduate Teaching Associate,
The Ohio State University

2006-2007 ................................. University Fellow, The Ohio State
University

2006 ............................................. B.A. Brandeis University

2002-2006 ................................. Wien Scholar, Brandeis University

Publications

Mihalicek, V. and C. Pollard. 2012. Distinguishing phenogrammar from tec-
togrammar simplifies the analysis of interrogatives. In deGroote, P. and M. Neder-

Mihalicek, V. 2010. The ordering of wh words in multiple wh extraction in Serbo-

Fields of Study

Major Field: Linguistics
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iv</td>
</tr>
<tr>
<td>Vita</td>
<td>vii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xiii</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Goals and Background</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Empirical Considerations</td>
<td>4</td>
</tr>
<tr>
<td>1.2.1 What Is Meant by Serbo-Croatian</td>
<td>4</td>
</tr>
<tr>
<td>1.2.2 What Is Counted As Data</td>
<td>5</td>
</tr>
<tr>
<td>1.2.3 The Scope of the Project</td>
<td>6</td>
</tr>
<tr>
<td>1.3 Formal Considerations</td>
<td>7</td>
</tr>
<tr>
<td>1.3.1 The Separation of Phenogrammar and Tectogrammar</td>
<td>8</td>
</tr>
<tr>
<td>1.3.2 Phenogrammatical Indeterminacy and ‘Canonical’ Order</td>
<td>14</td>
</tr>
<tr>
<td>1.4 Overview of the Dissertation by Chapters</td>
<td>20</td>
</tr>
<tr>
<td>2. Framework</td>
<td>22</td>
</tr>
<tr>
<td>2.1 Background</td>
<td>22</td>
</tr>
<tr>
<td>2.2 Phenogrammar</td>
<td>25</td>
</tr>
<tr>
<td>2.2.1 Preliminaries</td>
<td>25</td>
</tr>
<tr>
<td>2.2.2 Phenogrammatical Types</td>
<td>30</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Lower Level Types and Functions</td>
</tr>
<tr>
<td>2.2.4</td>
<td>Higher Level Types and Functions</td>
</tr>
<tr>
<td>2.3</td>
<td>Tectogrammar</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Preliminaries</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Representing Inflectional Features</td>
</tr>
<tr>
<td>2.3.3</td>
<td>N and NP type families</td>
</tr>
<tr>
<td>2.3.4</td>
<td>S type family</td>
</tr>
<tr>
<td>2.3.5</td>
<td>( \rightarrow ) types</td>
</tr>
<tr>
<td>2.3.6</td>
<td>( \Rightarrow ) types</td>
</tr>
<tr>
<td>2.4</td>
<td>Semantics</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Preliminaries</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Entailment</td>
</tr>
<tr>
<td>2.4.3</td>
<td>Types</td>
</tr>
<tr>
<td>2.5</td>
<td>Putting it all together</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Signs</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Rules</td>
</tr>
<tr>
<td>3.</td>
<td>Basic Word Order</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>3.2</td>
<td>Data</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Lexical Noun Phrases</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Phrasal Noun Phrases</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Adverbial Modifiers</td>
</tr>
<tr>
<td>3.3</td>
<td>Analysis</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Lexical Noun Phrases</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Phrasal Noun Phrases</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Adverbial Modifiers</td>
</tr>
<tr>
<td>3.4</td>
<td>Conclusion</td>
</tr>
<tr>
<td>4.</td>
<td>Embedding, Predicatives and Control</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>4.2</td>
<td>Embedded Declarative Clauses</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Data</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Analysis</td>
</tr>
<tr>
<td>4.3</td>
<td>Predicatives</td>
</tr>
</tbody>
</table>
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Phenogrammatical type hierarchy.</td>
<td>45</td>
</tr>
<tr>
<td>2.2</td>
<td>Phenogrammatical logical constants.</td>
<td>45</td>
</tr>
<tr>
<td>2.3</td>
<td>Phenogrammatical functions.</td>
<td>46</td>
</tr>
<tr>
<td>2.4</td>
<td>Phenogrammatical types, functions and word order phenomena represented at that level.</td>
<td>47</td>
</tr>
<tr>
<td>2.5</td>
<td>Typesetting conventions.</td>
<td>57</td>
</tr>
<tr>
<td>3.1</td>
<td>Examples of inflectional feature combinations on lexical noun phrases.</td>
<td>62</td>
</tr>
<tr>
<td>3.2</td>
<td>Summary of noun and noun phrase types.</td>
<td>111</td>
</tr>
<tr>
<td>4.1</td>
<td>The verb <em>biti</em> ‘be’ paradigm.</td>
<td>120</td>
</tr>
<tr>
<td>4.2</td>
<td>The verb <em>htjeti</em> ‘want, will’ paradigm.</td>
<td>137</td>
</tr>
<tr>
<td>5.1</td>
<td>The order of enclitics in the cluster.</td>
<td>166</td>
</tr>
<tr>
<td>5.2</td>
<td>Tectogrammatical types of enclitics.</td>
<td>168</td>
</tr>
<tr>
<td>5.3</td>
<td>Interim summary evaluation of grammar with respect to enclitic cluster placement.</td>
<td>204</td>
</tr>
</tbody>
</table>
7.1 Which traces can (√) or cannot (✗) escape from which constituents. 258

7.2 Tectogrammatical types of traces. . . . . . . . . . . . . . . . . . . . . . . 260
Chapter 1: Introduction

1.1 Goals and Background

Much attention has been given to Serbo-Croatian in mainstream generative grammar (henceforth MGG), by which we mean Minimalism (Chomsky, 1995) and its predecessors (Chomsky, 1957, 1965, 1981). Enclitics (e.g. Browne (1974); Progovac (1996); Radanović-Kocić (1996); Schütze (1994); Bošković (2004)) and \( \text{wh} \) questions (e.g. Rudin (1988a,b); Bošković (1997, 2002); Stjepanović (2003)) are among the phenomena that have been of most interest to the mainstream linguistic community.

Outside of MGG little work has been done on Serbo-Croatian. In HPSG, Gerald Penn has written about enclitic placement (Penn, 1999a) and word order in \( \text{wh} \) questions (Penn, 1999b), and Zlatić (1997) analyzed noun phrases in Serbo-Croatian in her thesis. Serbo-Croatian constituent questions were also discussed in Vermaat (2005), in Multi Modal Categorial Type Logic. Apart from Vermaat (2005), there is no work on Serbo-Croatian in the categorial grammar tradition that we know of.
The goal of our project is twofold: first, to put together a logical formalism in which a theory of Serbo-Croatian grammar could be expressed; and second, to write a theory of Serbo-Croatian grammar in that formalism. We aim to develop a fairly comprehensive grammar. While our focus is on predicting the correct word order in Serbo-Croatian utterances, syntactic and semantic representations of expressions are simultaneously generated as well.

Empirically, we think this project is distinguished in virtue of its scope. That is, unlike most literature on Serbo-Croatian, including Zlatić (1997), we do not focus on any one specific phenomenon, but try to construct a fragment which covers a substantial chunk of Serbo-Croatian, including declarative and interrogative clauses, noun phrases, prepositional phrases and enclitics.

While the MGG Progovac (2005) aims for a comprehensive description of Serbo-Croatian syntax, more broad that what we attempt here, it does not stand alone as a theory of Serbo-Croatian grammar which systematically predicts the correct form/meaning pairs of the language. In fact, this does not seem to be among the author’s goals at all; instead, Progovac (2005) wants to focus "on the functional projections of the clause" (a theory internal concept), and tie in "syntactic phenomena in Serbian with the theory of Universal Grammar" (p. 1).

In contrast, our project simply aims to construct a theory which predicts Serbo-Croatian form/meaning pairs. Concomitantly, we are constructing a framework, which in a sense constitutes the overarching theory of grammar (not Serbo-Croatian
grammar, just grammar). Certain theoretical constructs emerge as a result, such as the inventory of different ways in which phonological words can attach to one another, or different types of expressions with respect to their linearization properties. So instead of fitting our description of Serbo-Croatian into an existing overarching theory, like Progovac (2005), we are letting the empirical domain guide our design of the overarching theory. Further, we are not concerned with the theory of Universal Grammar at all. In sum, we do not consider Progovac (2005) as overlapping with our project very much at all.

The main theoretical contribution of this project lies in the development of a categorial architecture which is suitable for a detailed analysis of a language with more complex word order patterns than those traditionally analyzed in various categorial systems, such as English, German, Dutch or French. We will return to the discussion of the categorial architecture proposed here in Section 1.3.

It is important to recognize that this thesis does not aim to establish that the proposed formalism is superior, nor that the theory of Serbo-Croatian grammar expressed in it is the right one (it is not). It simply aims to explicitly articulate a formalism and use it to construct a formal theory. Our hope is that subsequent inspection of the concrete theory proposed here and the formalism it’s embedded in, of their advantages and particularly their shortcomings, will lead to a better understanding of the empirical domain, and also help delineate a realistic categorial framework for linguistic description.
In Chomsky’s words (Chomsky (1957); also cited in Reape (1993)):

Precisely constructed models for linguistic structure can play an important role, both negative and positive, in the process of discovery itself. By pushing a precise but inadequate formulation to an unacceptable conclusion, we can often expose the exact source of this inadequacy and, consequently, gain a deeper understanding of the linguistic data. More positively, a formalized theory may automatically provide solutions for many problems other than those for which it was explicitly designed. Obscure and intuition-bound notions can neither lead to obscure conclusions nor provide new and correct ones, and hence they fail to be useful in two important respects. I think that some of those linguists who have questioned the value of precise and technical development of linguistic theory have failed to recognize the productive potential in the method of rigorously stating a proposed theory and applying it strictly to linguistic material with no attempt to avoid unacceptable conclusions by ad hoc adjustments or loose formulation.

A secondary theoretical goal of this thesis, related to the design of the formalism, is to challenge the standard assumption that there is a ‘canonical’ word order from which other possible orders are derived, and that semantically insignificant reordering of its subexpressions individuates linguistic expressions. We discuss both of these in more detail in Section 1.3.

1.2 Empirical Considerations

1.2.1 What Is Meant by Serbo-Croatian

Serbo-Croatian language does not exist anymore, in the sense that as an official language it has been replaced by Bosnian, Croatian, Montenegrin and Serbian. We will, however, continue to use the term Serbo-Croatian throughout, for several reasons. First, we agree with Kordić (2010) that Bosnian, Croatian, Montenegrin
and Serbian are all just variants of the same polycentric language, whose basic
structure is what we aim to describe here. Second, the sort of basic linguistic
phenomena analyzed here we believe to be largely shared by various varieties
of Serbo-Croatian, whether one considers them distinct languages or not. Third,
the author’s judgments will partly constitute the data set and the author’s native
language is in fact Serbo-Croatian.

We will reference literature which purports to be about Bosnian (Leko, 1999),
Serbian (Progovac, 2005), and Serbo-Croatian (Bošković, 1997). The examples
will consistently be given in ijekavian forms because the author’s native and sec-
ondary dialect are both ijekavian, but this should not be taken as desire to exclude
ekavian dialects from consideration or analysis.¹

1.2.2 What Is Counted As Data

The data presented either reflect the author’s judgments or are drawn from the
literature. As mentioned earlier, we will reference literature that is self-reportedly
about different dialects. Altogether, this is an inconsistent data set.

Since no empirical study was conducted as a part of this project, we attempt to
side step the issue of empirical inconsistencies and which empirical generalization
is correct by constructing many different grammars in parallel. That is, if there is

¹There is a systematic regional phonological difference between dialects based on the reflexes
of the Proto-Slavic vowel *jat*. In ekavian dialects, this vowel developed into e, and in ijekavian
dialects it developed into je or ije, depending on length. There are also ikavian dialects, in which
*jat* developed into i, but no standard dialects of Serbo-Croatian are based on ikavian, so we didn’t
mention it above.
both $X$ and $not X$ present in our data set, we show how to construct a grammar that predicts $X$, but also how to construct a distinct grammar that predicts $not X$. In general, however, we pursue in most detail the empirical generalizations which are in accordance with the author’s judgments.

1.2.3 The Scope of the Project

As mentioned earlier, we attempt to construct a theory which predicts form/meaning pairs for a fairly large fragment of Serbo-Croatian. Our grammar generates many different constituents including noun phrases with various modifiers or determiners, prepositional phrases, main and embedded declarative clauses, and polar and constituent questions. We also give an analysis of verb phrase modifiers as well as predicative and control verbs. Finally, our grammar orders enclitics and places the resulting cluster into a variety of clauses.

However, many empirical phenomena are not discussed here at all, including relative clauses, topicalization, coordination, negative concord, anaphora and binding. Also, temporal and modal semantic properties of expressions are largely ignored. For example, we have nothing to say about subjunctive mood in Serbo-Croatian. These unfortunate omissions are due to time constraints and not our belief that the phenomena in question are unimportant or negligible.

Further, we have little to say about prosody or pragmatics here. Our grammar does not represent explicitly any prosodic concepts except for distinguishing
between clitics and non-clitics. Lexical pitch accents and other tones in Serbo-
Croatian, as well as prosodic constituency, are not represented in the grammar at
all. Such considerations are beyond the scope of this project, though our hope
is that one day prosodic information will be standardly represented in categorial
grammars.

The grammar does not explicitly represent any pragmatic concepts and is not
sensitive to pragmatic differences. Therefore, there is no notion of felicity in a
given context in our grammar. We in general assume that if there is a context
in which some sentence $X$ can be felicitously uttered, then $X$ is grammatical.
Throughout, we are solely concerned with truth conditional meaning. This is
not to suggest that pragmatics is not important or that pragmatic considerations
should not eventually be integrated into a theory of Serbo-Croatian grammar; un-
fortunately, however, we cannot pursue these issues here.

Finally, although we aim for a fair amount of linguistic detail in our analysis,
due to the breadth of the project many details have to be omitted. Obviously,
Zlatić (1997), as a monograph on Serbo-Croatian noun phrases, examines noun
phrases in more detail than we can attempt to do here.

1.3 Formal Considerations

A part of this project is proposing a logical formalism, that is, a categorial
framework, in which we can express a theory of Serbo-Croatian grammar. In
doing so, we made a methodological assumption to treat phenogrammar and tectogrammar as distinct grammatical components. This property of our framework is discussed further in 1.3.1.

Once syntactic combination, which determines semantic composition, is disassociated from word order, we can abandon the idea that tectogrammar is involved in managing semantically insignificant reordering. This leads to a theory according to which expressions do not necessarily have a unique pronunciation; rather, a single expression with a given meaning can be linearized multiple different ways. In this way, our theory eschews the idea of canonical word order altogether. This property of our framework is discussed further in 1.3.2.

1.3.1 The Separation of Phenogrammar and Tectogrammar

Background

We assume that word order and combinatorial syntax (i.e. functor/argument relations or argument structures) are represented in the grammar as distinct components, which we refer to as phenogrammar and tectogrammar respectively, following Curry (1961). In making this assumption, we are joining an existing research paradigm in theoretical linguistics. Our project is conceptually aligned with Dowty (1996); Reape (1993, 1996); Kathol (2000); Muskens (2003, 2007b); see also linear precedence rules of Pollard and Sag (1987) and GPSG (Gazdar et al.
Contemporary categorial grammar frameworks which share this assumption are Abstract Categorial Grammar (de Groote, 2001) and Lambda Grammar (Muskens, 2003, 2007b), which our framework is closely related to.

The division of labor between phenogrammar and tectogrammar allows complex word order facts to be described somewhat independently in phenogrammar, without unnecessarily complicating the tectogrammar. In particular, if phenogrammar is a distinct component, semantically insignificant reordering of expressions doesn’t have to be mediated by tectogrammar.

The separation of phenogrammar and tectogrammar has been argued to be useful and important for elegantly analyzing phenomena such as quantifier scoping and medial extraction in English (Oehrle, 1994), as well as German word order (Reape, 1993; Kathol, 2000). Muskens (2003, 2007b) argues for the separation of phenogrammar and tectogrammar on conceptual grounds, and also provides analyses of certain Dutch word order facts in Lambda Grammar. Mihalicek and Pollard (2012) argue that the separation of phenogrammar and tectogrammar considerably simplifies the analysis of interrogatives in English and Chinese, bringing out the underlying tectogrammatical similarities between the two languages and identifying phenogrammar as the locus of cross-linguistic variation with respect to different question forming strategies.

However, mainstream categorial grammar frameworks, such as Multimodal Combinatory Categorial Grammar (Baldrige, 2002), Multimodal Categorial Type
Logics (Moortgat, 1997; Bernardi, 2002; Vermaat, 2005) and Multimodal Type Logical Grammar (Morrill and Solias, 1993; Morrill, 1994; Morrill et al., 2007), do not assume that phenogrammar and tectogrammar are distinct components. Below we will first clarify our position on this issue, and then try to motivate our decision to distinguish phenogrammar from tectogrammar given our stated goals of describing Serbo-Croatian word order.

It is important to recognize that we are making a methodological choice. This thesis does not attempt to demonstrate that frameworks which distinguish phenogrammar and tectogrammar are conceptually, empirically or theoretically superior to those which do not. Given that both research paradigms have produced interesting results, we think it’s fair to conclude at this point that the issue is undecided; the jury is still out. However, a lot more research has been done in the mainstream categorial paradigm, than in the one which assumes a dedicated word order component of the grammar. So, this thesis can be taken as an attempt to reduce this imbalance and show that a fairly general and workable linguistic theory can be expressed in a non-mainstream categorial formalism which distinguishes word order and syntax.

Whether a comparable (or superior) theory of the same empirical domain can be expressed in a mainstream formalism or not is itself an empirical question, one which we do not purport to answer here. We think that an attempt to replicate
the results of this thesis in a mainstream framework such as Multi Modal Combinatory Categorial Grammar (Baldridge, 2002) would be a fruitful enterprise. We would find either that a comparable theory of Serbo-Croatian word order cannot be expressed in a formalism without a designated phenogrammatical component, or that it can.

Whatever the case, we would learn something about the design of a categorial linguistic theory. In particular, distinguishing phenogrammar might turn out to be necessary for a general theory which can be extended to languages such as Serbo-Croatian. Alternatively, we might learn that the two paradigms are essentially equivalent in terms of empirical coverage and then be compelled to discard the non-mainstream one because its architecture is inherently more complex. Of course, in between these two extremes lies a plethora of more moderate positions, and that’s where our project is currently situated.

Motivation

There are good reasons for embracing the more complex, non-standard categorial architecture given what our theory is supposed to be about. Serbo-Croatian word order patterns are complex. Some expressions and groups of expressions can be ordered freely and even occur discontinuously; other expressions’ word order is completely fixed; some expressions can be internally freely ordered but their position with respect to other clausal constituents is fixed; conversely, some
expressions have a fixed internal order, but can freely reorder with respect to other clausal constituents.

Given that our primary goal is to construct a theory of Serbo-Croatian word order, we make the methodological choice to do so in a framework which distinguishes phenogrammar as a separate component. This allows us to essentially isolate the problem: Syntactic combination and semantic composition can proceed in a straightforward fashion, unencumbered by the word order complexities, which are confined entirely to the phenogrammatical component. We can then ask the following question: Given a reasonable syntactic representation coupled with a given semantic representation, how is it possible to order that expression, both internally and externally (i.e. with respect to other expressions)? In other words, we can develop a syntactic and semantic description of a given expression, and then independently specify how that expression can be linearized.

Further, having a distinct phenogrammatical component enables us to state word order generalizations, by which we mean theory-internal phenogrammatical generalizations. This is particularly important for Serbo-Croatian 2P enclitics, whose placement cannot be adequately characterized in tectogrammatical terms.

Serbo-Croatian enclitics have to occur in the second position in the clause, and can be hosted by a variety of expressions, including adjectives, nouns, noun phrases, finite and non-finite verbs, adverbs and complementizers. Therefore, the
set of possible hosts is tectogrammatically heterogeneous. Progovac (1996) maintains a syntactic generalization, whereby the enclitic hosts all share the property of being independently ‘moveable’ around the clause. However, most researchers, including the author, disagree with this claim on empirical grounds. In particular, 2P encliticization can introduce discontinuities in otherwise unbreakable constituents, such as coordinate structures, noun phrases which contain a postnominal modifier, and adverbial phrases which contain a degree.

While Romance clitics have been successfully analyzed in mainstream categorial grammar without a phenogrammatical component (Morrill and Gavarró, 1992; Kraak, 1998), Romance clitics are considerably simpler in that they always procliticize onto the finite verb. In other words, the host is tectogrammatically determined. Further, while Romance clitics essentially only include pronominal clitics, Serbo-Croatian also has clitic forms of control and predicative verbs.

It is an open question whether it is possible to articulate an analysis of Serbo-Croatian 2P clitics in mainstream categorial grammar. Our suspicion is that achieving an empirically adequate analysis would be quite challenging without a phenogrammatical component. Further, even if this was accomplished, we believe that such an analysis would be conceptually unsatisfying since the placement of Serbo-Croatian clitics is simply not a purely tectogrammatical phenomenon.

In our theory, as we will see in Chapter 5, we are able to state a phenogrammatical generalization concerning the placement of the enclitic cluster, namely, that it
always attaches to the last phonological word in the initial length one S-string in the phonogrammatical denotation of the clause in question. Further, this generalization covers both 1W (after the first word) and 1C (after the first constituent) placement. Thus, we believe that Serbo-Croatian 2P clitics empirically strongly suggest and conceptually require that there be a distinct phonogrammatical component.

1.3.2 Phenogrammatical Indeterminacy and ‘Canonical’ Order

In our framework, word order is dissociated from combinatorial syntax and semantics. This opens the door to representing expressions, with a given semantic meaning and syntactic structure, as not having a unique pronunciation. Consider the following Serbo-Croatian example:

(1) a. Marko voli Vesnu.
    Marko{NOM} loves Vesna{ACC}
    ‘Marko loves Vesna’

b. Marko Vesnu voli.

c. Voli Marko Vesnu.

d. Voli Vesnu Marko.

e. Vesnu Marko voli.

f. Vesnu voli Marko.

The verb, the subject and the object can freely order with respect to one another and the truth-conditional semantic meaning is not affected. A linguistic theory
can treat the six strings above as corresponding to six distinct expressions, which are syntactically all declarative sentences, with a shared meaning represented by the term (love vesna marko). Alternatively, a theory can treat the six strings above as different pronunciations of a single expression, which is a declarative sentence expressing the proposition (love vesna marko). We pursue the latter approach, which we refer to as phenogrammatical indeterminacy, because it implies that an expression’s pronunciation is not uniquely determined. Rather, each expression is associated with a set of possible pronunciations.

First we’d like to point out that phenogrammatical indeterminacy is not a necessary consequence of distinguishing phrogrammar and tectogrammar. However, once we’ve separated these two components, and more importantly, sets of properties of linguistic expressions, a question of what individuates expressions naturally arises. In particular, we may ask whether linguistic expressions are individuated by syntactically and semantically insignificant reordering of their subparts. In this thesis, for better or worse, we answer this question with a resounding “No!”

Phenogrammatical indeterminacy is conceptually appealing, assuming that there truly are instances of reordering which are not conditioned by nor have consequences for the syntactic and semantic properties. We believe that there indeed is such reordering in Serbo-Croatian.
Not only do we assume phenogrammatical indeterminacy, but within the set of possible pronunciations of a given expression, we do not treat any one pronunciation as somehow basic or ‘canonical’, from which other pronunciations are somehow derived. For us, a phenogrammatical representation of an expression which can be internally ordered multiple different ways is just a bag of equally good pronunciations.

In this way, our work is a direct continuation of Reape (1993), who (we believe accurately) describes the received, mainstream view in linguistic theory as positing a certain ‘canonical’ order of constituents, and then adding special mechanisms to generate ‘non-canonical’ orders.

In addition, the deviations from the ‘canonical’ order are supposed to be justified in some way, typically pragmatically, as it assumed to be impossible for a ‘non-canonical’ order to occur without any special reason. For example, Bošković (1997) allows ‘non-canonical’ (Superiority violating) relative ordering of phrasal wh expressions in Serbo-Croatian only if they are D-linked. In contrast, we assume that certain expressions can freely order simply because they can freely order, not because of some special syntactic, semantic or pragmatic property or consequence.

In sum, this thesis, much like Reape (1993), challenges the received view of linearization of constituents. Instead of asking under what conditions it is possible to deviate from some fixed, ‘canonical’ word order, we try to approach this
issue from the other end, that is, assume that anything goes and ask under what conditions it is necessary to restrict or completely fix the order of expressions.

The assumption of phenogrammatical indeterminacy and the eschewing of canonical word order in this thesis should be understood in the same way as our choosing a grammatical architecture which contains a designated word order component. In other words, it is a methodological assumption guided by our consideration of the empirical domain, namely Serbo-Croatian word order.

While these methodological assumptions enable us to give a reasonable theory of Serbo-Croatian word order, one may wonder (i) whether the framework can be extended to represent pragmatic and prosodic information, and (ii) whether systematic correspondences between word order and pragmatic information could be expressed in the framework. We believe that both (i) and (ii) are in principle possible, though we are not prepared to demonstrate that, by implementing these extensions.

We see no obstacle to extending our framework by adding prosodic information to phenogrammar, and pragmatic information to semantics, or to adding prosody and pragmatics as distinct components to the existing architecture.

The more interesting question is whether the framework could be extended to represent systematic correspondences between prosody and word order on the one hand, and pragmatic and semantic information on the other. If expressions’
phenogrammatical representations are undifferentiated bags of their possible pronunciations, how could we pick out a single pronunciation and correlate it with some pragmatic property?

With respect to Serbo-Croatian, we are not convinced that there are any such systematic correspondences. In this thesis, we assume the null hypothesis, i.e. that there are no such systematic correspondences. Therefore, the burden of proof is on those who believe that such correspondences really do exist to demonstrate them through a quantitative empirical study. We know of no such body of work.

Godjevac (2000) conducted a production study and mapped out the prosodic inventory of Serbo-Croatian. She also established correspondences between word order and prosodic structure (e.g. the placement of phrase accents). However, in our judgment, she asserted but did not establish that there exist systematic correspondences between word order and prosodic structure on the one hand, and pragmatic properties on the other hand. In the absence of a definitive answer as to whether there are correlations between word order and prosody, and pragmatic information in Serbo-Croatian, we are entitled to maintain our null hypothesis.

However, even if it’s true that there aren’t such correspondences in Serbo-Croatian, they may indeed exist in other languages. For example, Gärtner (to appear) argues that linear order plays a role in allowing scope extension of negative quantifiers in English, in other words, that there is a correspondence between word order and semantics in English. Also, it is generally accepted that prosodic
properties of English expressions, such as the placement of pitch accents, correlate with pragmatic information. Can the framework proposed here be used to describe a language like English?

The answer is affirmative. First, even in our current set up, we assume that differences in truth conditional meaning individuate expressions. A reordering which triggers such different interpretations is semantically significant, and would not be modeled in terms of phenogrammatical indeterminacy, as one of many possible phenogrammatical guises of a single expression.

Second, phenogrammatical indeterminacy is compatible with our framework and something we embrace given certain facts about Serbo-Croatian, but it is not its necessary property. It emerges from the lexical entries we give Serbo-Croatian expressions, not from some more fundamental property of the formalism.

In principle, the same formalism could be used to express a theory with no phenogrammatical indeterminacy, or with more restricted phenogrammatical indeterminacy than what we assume. In other words, how we individuate expressions in this framework is largely determined by specific lexical entries, not intrinsic properties of the formalism.

If we wanted to give a theory of English grammar in our framework, we could represent each expression as having a unique pronunciation, which opens the door for representing correspondences between word order and prosody, and
word order and prosody on the one hand and pragmatics on the other. For example, the theory of constituent questions proposed in Mihalicek and Pollard (2012) is essentially expressed in the framework assumed here, modulo minor notational changes and a slightly different theory of interrogative meaning, and there is no phenogrammatical indeterminacy in Mihalicek and Pollard (2012).

In sum, we think phenogrammatical indeterminacy is an interesting, even provocative, hypothesis which we embrace as reasonable for Serbo-Croatian. Our framework enables us to implement phenogrammatical indeterminacy but it does not require us to do so. We join Mihalicek and Pollard (2012) in suggesting that phenogrammar is a locus of cross-linguistic variation. In particular, some languages, like Serbo-Croatian, might have phenogrammatical indeterminacy, while others, such as English, do not. Our framework, given that it provides us with the notion of phenogrammatical indeterminacy, enables us to entertain such generalizations.

1.4 Overview of the Dissertation by Chapters

Chapter 2 describes each component of the framework in more detail.

Chapter 3 analyses basic word order patterns in main declarative clauses consisting of a verb together with its noun phrase arguments. Determiners, attributive adjectives and other noun modifiers are also analyzed, as well as verb phrase adverbials, including prepositional phrases.
Chapter 4 presents our analysis of more complex structures in Serbo-Croatian. In particular, we analyze embedded declarative clauses, as well as subject and object control, and predicatives.

Chapter 5 is dedicated entirely to enclitics. We show how to assemble the clitic cluster and how to ensure it winds up in the correct slot in a clause. Both 1W and 1C placement of the cluster are discussed.

Chapter 6 analyzes polar and constituent questions. We show how to generate two different kinds of polar interrogatives. Both single and multiple constituent questions are discussed, with respect to word order possibilities and scope.

Chapter 7 addresses locality-based overgeneration of the grammar and substantially refines the system proposed in earlier chapters. We propose a simple tectogrammatical mechanism for enforcing locality constraints in the grammar, thereby preventing clitic climbing, long distance control and extraction from questions.

Chapter 8 concludes by summarizing and evaluating the project.
Chapter 2: Framework

2.1 Background

In this chapter, we describe the framework that our theory of Serbo-Croatian grammar is expressed. First we provide some historical and conceptual background.

While obviously natural language expressions have both purely combinatorial syntactic properties (what sorts of arguments do they require? what sorts of expressions can they be arguments of?), and ordering properties (do they have to occur in some specific place in a clause or not? do they have to occur immediately to the left or right of some other expressions or not? how can they be internally ordered?), in many logical frameworks these two sets of properties are represented jointly, by a single component of the framework.

For example, in many mainstream versions of Categorial Grammar (Multimodal Combinatory Categorial Grammar (MMCCG; Baldridge (2002)), Multimodal Categorial Type Logics (MMCTL; Moortgat (1997); Bernardi (2002); Vermaat (2005)) and Multimodal Type Logical Grammar (MMTL; Morrill and Solias

22
(1993); Morrill (1994); Morrill et al. (2007)), and non-multimodal versions of these frameworks), both combinatorial syntax and word order of expressions are represented by a single component of the framework. This results in a certain degree of inflexibility when it comes to dealing with freer word order. We see the development of multimodal versions of these frameworks as an attempt to compensate for the inflexibility that stems from combinatorial syntax and word order being represented jointly, by the same component (see also (Muskens, 2003)).

Curry (1961) proposed a different architecture, in which combinatorial syntax and word order are represented by distinct components. He referred to these two components as tectogrammar and phenogrammar. While Curry’s original paper was largely programmatic in nature, since then many concrete frameworks which embody the distinction between phenogrammar and tectogrammar have been proposed.

Cresswell (1973) was the first to use lambda calculus to formalize the tectogrammatical component. Oehrle (1994) and Ranta (1994) both developed frameworks which embody the phenogrammar/tectogrammar distinction and use typed lambda calculi to represent each component. While Ranta (1994) uses Martin-Löf Type Theory (Martin-Löf, 1984) for the type system, Oehrle (1994) uses linear logic as the basis for the tectogrammatical type system.

We will ambiguously use phenogrammar and tectogrammar to refer to sets of properties of linguistic expressions, but also the components of the theory that represent these sets of properties.
The use of typed lambda calculi in pheno- and tectogrammar, and the use of linear logic in the formalization of tectogrammar (Oehrle, 1994) is adopted by Abstract Categorial Grammar (de Groote, 2001) and Lambda Grammar (Muskens, 2003, 2007b), and by our framework here as well. Other frameworks which embody similar ideas, and out of which the framework that we use here developed, include Higher-Order Grammar (Pollard, 2004), CVG (Pollard, 2011), what has previously been called Linear Logic Based Grammar (Mihalicek, 2010b) and Pheno-Tecto Distinguished Categorial Grammar (originally in Smith (2010); also Mihalicek and Pollard (2012)).

We choose to implement our theory of Serbo-Croatian in a framework which distinguishes tectogrammar and phenogrammar. This choice is methodological in nature and guided by complex word order patterns that Serbo-Croatian exhibits, particularly the 2P placement of the enclitic cluster. Having a dedicated phenogrammatical component allows us to isolate the problem: We can give a reasonable tectogrammatical and semantic representation to an expression, and then independently ask what the possible pronunciations of that expression are, i.e. what its phenogrammatical representation is. Further, we are able to state phenogrammatical generalizations, which is crucial for the analysis of 2P clitics. We direct the reader to Section 1.3.1 for a more detailed discussion of the motivation for distinguishing phenogrammar from tectogrammar.
In addition to phenogrammar and tectogrammar, we also assume a semantic component. Our framework consists of three term/type calculi, which independently represent phenogrammar, tectogrammar and semantics. Each linguistic expression is then represented in our theory as a triple of term/type pairs, corresponding to the representation of that expression’s word order, combinatorial syntactic properties, and its meaning. In the remainder of this chapter, we describe each component separately and then sketch how the three components work together.

2.2 Phenogrammar

2.2.1 Preliminaries

The phenogrammatical component of our framework is the most elaborate one. We propose a variety of types, and functions on terms of various types. The reason for the complexity of this component is the complexity of the empirical domain that it’s modeling, that is, Serbo-Croatian word order. The details of the linearization properties of Serbo-Croatian expressions will be presented in substantive chapters later on. In this section, we briefly survey different kinds of linearization properties of Serbo-Croatian expressions, in order to motivate the introduction of specific phenogrammatical mechanisms later on.

3The contents of this section are largely based on Carl Pollard’s lectures in LING 602.02 taught at The Ohio State University, January-March 2012.
It is instructive to consider linearization properties of an expression with respect to the following two dimensions:

- the degree of flexibility of the ordering of that expression’s subparts (internal ordering); and,

- the degree of flexibility of the ordering of that expression with respect to other expressions in the clause or other phrase that it’s a subpart of (external ordering).

For example, a noun phrase consisting of a noun and a postnominal modifier is rigidly internally ordered: The postnominal modifier must occur immediately to the right of the noun it modifies. However, such a noun phrase, as a verbal argument, shows a great deal of flexibility in its external ordering: It can freely order with respect to the verb and any other nominal arguments.

On the other hand, a noun phrase consisting of a noun and an attributive adjective is flexibly ordered both internally and externally, so much so that the noun and the adjective may even occur discontinuously in a clause.

An embedded clause is internally ordered semi-flexibly: Most of its subparts can freely order, but the complementizer must occur on its left periphery. In terms of its external ordering, it is quite rigid: The embedded clause must remain contiguous in the matrix clause and tends to occur on the right edge of the matrix clause, so its external ordering properties are quite rigid.
We can recast these linearization differences among Serbo-Croatian expressions as consequences of their varying degrees of the strength of attachment to other expressions:

- how strongly does an expression attach to other expressions it combines with?

The most extreme form of attachment is clitic attachment: A clitic combined with a phonological word results in another phonological word, which is unbreakable. This bond is so deep that the boundary between a clitic and its host is invisible already at level of phonological words, let alone at higher levels, such as the one at which phrases are ordered.

The attachment of a postnominal modifier to the noun it modifies is strong, but not quite as strong as the clitic attachment. The postnominal modifier must occur immediately to the right of the noun, but the sequence of the noun and the postnominal modifier can be broken up by enclitics.

The attachment of an attributive adjective the noun it modifies is even weaker: Internal reordering of such noun phrases, and even discontinuities of the noun phrase are allowed. That is, the sequence of the adjective and the noun can be broken up not just by enclitics, but also by other phonological words and sequences of phonological words. However, their attachment is not completely loose because under certain conditions, namely when there is also a postnominal modifier
in addition to the attributive adjective, the adjective must remain adjacent to the noun.

Finally, when we look at the attachment of verbs to their noun phrase arguments in a main clause, we find that it is very loose; they can freely order with respect to one another. However, in freely ordering verbs and their noun phrase arguments, we have to ensure that lower level attachments are respected, for example, that the clitics are not separated from their hosts, and that postnominal modifiers are not separated from their nouns.

At the level of embedded clauses, although the attachment of expressions in main clauses is very weak, the complementizer attaches to its main clause argument very strongly. Not only is their relative order fixed, but the entire resulting embedded clause must occur in a relatively fixed position in the matrix clause, and be protected from intrusions from the matrix clause.

In trying to model these various levels of attachment and degrees of ordering flexibility, we are lead to postulate a hierarchy of phenogrammatical types and accompanying functions. Each Serbo-Croatian expression is represented in the grammar by a phenogrammatical term. The type of that term locates that expression somewhere on the phenogrammatical type hierarchy, from the lowest (most basic) types of clitics and phonological words, to the highest level types of loosely attached phrases. The higher level types are all constructed out of the basic type of phonological words.
The phenogrammatical operations essentially model the strength of attachment between expressions. Operations defined on lower level types model very strong attachment. Operations defined on higher level types model varieties of less strong attachment. These higher level operations are oblivious to finer grained phenogrammatical structure of their arguments, and they cannot destroy any existing ordering enforced on a lower level. For example, operations defined on loosely attached phrases simply do not see, and therefore cannot annihilate, lower level clitic attachment.

Finally, a host of phenogrammatial functions help us traverse the type hierarchy. For example, an expression might have to be strictly ordered internally, but flexibly ordered with respect to other expressions, as in the case of noun phrases containing a postnominal modifier. Such noun phrases are internally glued at a fairly low level, but externally, from the clausal vantage point, appear to be of a higher level type which enables them to loosely attach to other expressions in the clause.

At the same time, recall that a clitic can break up the sequence of a noun and its postnominal modifier. In that case, although from the clausal perspective we are dealing with a higher level object and cannot see its lower level structure, we have to be able to look inside that higher level object and see its internal structure so we can break it up and correctly place the clitic.
In sum, the grammar provides us with mechanisms (i) to look at a term of a lower level type and ‘zoom out’, i.e. elevate it to a higher level type and protect its lower level structure; and (ii) to look at a term of a higher level type and ‘zoom in’, i.e. look at and possibly alter its lower level structure. Defining such functions which can phenogrammatically zoom in or out is possible because all higher level types are constructed out of the basic type of phonological words.

The hierarchy of phenogrammatical types and the operations defined on these types, together with functions which enable us to raise or lower the phenogrammatical type of a given expression, allow us to exert a great deal of control over linearization.

2.2.2 Phenogrammatical Types

We recursively define the set of phenogrammatical types as follows. A more informal presentation follows.

(2)  
\begin{enumerate}
\item \( p, c, \) and \( t \) are basic types.
\item If \( P \) and \( P' \) are types, then so are:
  \begin{enumerate}
  \item \( N_P \) (the type of null \( P \)-strings)
  \item \( C_P \) (the type of non-null \( P \)-strings)
  \item \( P \rightarrow P' \)
  \item \( P + P' \)
  \end{enumerate}
\item Nothing else is a type.
\end{enumerate}
2.2.3 Lower Level Types and Functions

Informally, the lowest level phenogrammatical types are the type of clitics, \( c \), and the type of phonological words, \( p \). For example, a phonological word such as \( \text{grad} \) ‘city’, and clitics, such as the preposition \( u \) ‘in, to’, and the auxiliary \( je \) ‘is’ are represented in the grammar as follows:

\[(3)\]

a. \( \vdash \text{grad} : p \)
   read as: ‘\( \text{grad} \) is a term of type \( p \)’,
   i.e. the expression \( \text{grad} \) is a phonological word
b. \( \vdash u : c \)
   read as: ‘\( u \) is a term of type \( c \)’,
   i.e. the expression \( u \) is a clitic
c. \( \vdash je : c \)
   read as: ‘\( je \) is a term of type \( c \)’,
   i.e. the expression \( je \) is a clitic

There is a pair of functions which model the strongest kind of attachment, namely clitic attachment. These functions combine a clitic with a phonological word, and output another phonological word.

\[(4)\]

a. \( \vdash \#_{pc} : c \rightarrow p \rightarrow p \) [proclitic attachment]
b. \( \vdash \#_{ec} : p \rightarrow c \rightarrow p \) [enclitic attachment]

For example, if \( \vdash \text{grad} : p \) (\( \text{grad} \) is a phonological word), and \( \vdash u : c \) (\( u \) is a clitic), then \( \vdash (u\#_{pc}\text{grad}) : p \), i.e. \( u \text{grad} \) is a phonological word consisting of the clitic \( u \) procliticized onto the phonological word \( \text{grad} \).

Similarly, if \( \vdash \text{grad} : p \), and \( \vdash je : c \), then \( \vdash (\text{grad}\#_{ec}je) : p \), i.e. \( \text{grad} \text{je} \) is a phonological word consisting of the clitic \( \text{je} \) encliticized onto the phonological word \( \text{grad} \). In practice, we drop the subscripts on \( \# \).
Next up in the hierarchy of types are the $\mathbf{p}$-string types: the type of non-null $\mathbf{p}$-strings ($\mathbf{C}_\mathbf{p}$), the type of null $\mathbf{p}$-strings ($\mathbf{N}_\mathbf{p}$), and the type of all $\mathbf{p}$-strings is $\mathbf{N}_\mathbf{p} + \mathbf{C}_\mathbf{p}$.

The type of non-null $\mathbf{p}$-strings is built out of phonological words. For $\mathcal{P}$ a type, the following function maps terms of type $\mathcal{P}$ to corresponding non-null $\mathcal{P}$-strings:

$$
(5) \quad \vdash \text{tos}_\mathcal{P} : \mathcal{P} \rightarrow \mathbf{C}_\mathbf{p}
$$

We make use of the following abbreviation:

$$
(6) \quad \text{toS} = \text{def} \ \text{tos}_\mathcal{P}
$$

Interchangeably with the above abbreviations, and depending on legibility, we use the following abbreviations as well:

$$
(7) \quad \text{for } \phi \text{ a term of type } \mathbf{p}, \ \phi_s = \text{def} \ (\text{toS } \phi)
$$

Intuitively, the function $\vdash \text{tos} : \mathbf{p} \rightarrow \mathbf{C}_\mathbf{p}$ takes a phonological word (a term of type $\mathbf{p}$) and takes it up one level, by constructing a length one string out of it. The type $\mathbf{C}_\mathbf{p}$ is the type of non-null strings of phonological words. So, the expression $\text{grad} \ 'city'$, in addition to being represented as $\vdash \text{city} : \mathbf{p}$, can also be viewed as the string $\vdash \text{grad}_s : \mathbf{C}_\mathbf{p}$.

For $\mathcal{P}$ a type, $\mathbf{N}_\mathcal{P}$ is the type of null $\mathcal{P}$ strings. Each such type has the unique inhabitant $\vdash \epsilon_\mathcal{P} : \mathbf{N}_\mathcal{P}$. In particular, we make use of the following null string:

$$
(8) \quad a. \quad \vdash \epsilon_\mathcal{P} : \mathbf{N}_\mathcal{P}
$$

It is important to distinguish between the type of null and non-null strings for formal reasons, but also for theoretical reasons, in particular in order to prevent
overgeneration. Consider the functional terms below. They allow us to build up and tear down terms of various string types. For $\mathcal{P}$ a type:

(9)  
\begin{align*}
& a. \vdash \text{cns}_p : \mathcal{P} \to (\mathcal{N}_p + \mathcal{C}_p) \to \mathcal{C}_p \\
& b. \vdash \text{fst}_p : \mathcal{C}_p \to \mathcal{P} \\
& c. \vdash \text{rst}_p : \mathcal{C}_p \to (\mathcal{N}_p + \mathcal{C}_p)
\end{align*}

(10)  
\begin{align*}
& a. \vdash \text{snc}_p : (\mathcal{N}_p + \mathcal{C}_p) \to \mathcal{P} \to \mathcal{C}_p \\
& b. \vdash \text{lst}_p : \mathcal{C}_p \to \mathcal{P} \\
& c. \vdash \text{tsr}_p : \mathcal{C}_p \to (\mathcal{N}_p + \mathcal{C}_p)
\end{align*}

Intuitively, $\text{cns}_p$ takes a phonological word and prefixes it to a string, resulting in a non-null string. $\text{fst}_p$ takes a non-null $\mathcal{p}$-string and outputs the initial phonological word. Conversely, $\text{snc}_p$ suffixes a phonological word to a $\mathcal{p}$-string, and $\text{lst}_p$ takes a non-null $\mathcal{p}$-string and outputs the last phonological word. For example, $\text{fst}_p(\text{marko}s \cdot \text{spava}s) = \text{marko}$, and $\text{lst}_p(\text{marko}s \cdot \text{spava}s) = \text{spava}$.

For any non-null string $c$, if you take off its prefix ($\text{fst}_p c$) and then reattach that prefix to what’s left of $c$ ($\text{rst}_p c$), you just get the same $\mathcal{p}$-string $c$ back. Analogously, if you take off its suffix ($\text{lst}_p c$) and then reattach the suffix to what’s left of $c$ ($\text{tsr}_p c$), you get $c$ back. This is stated more formally below.

(11)  
\begin{align*}
& a. \forall_{c \in \mathcal{C}_p}[(\text{cns}_p (\text{fst}_p c) (\text{rst}_p c)) = c] \\
& b. \forall_{c \in \mathcal{C}_p}[(\text{snc}_p (\text{tsr}_p c) (\text{lst}_p c)) = c]
\end{align*}

We will use the functions $\text{fst}_p$ and $\text{lst}_p$ to look inside a string of words and pick out a phonological word that a clitic procliticizes (respectively, encliticizes) onto.

\[ \text{Here, we are borrowing and renaming basic Lisp functions: cons is just our \text{cns}_p, car is our \text{fst}_p, and cdr is our \text{rst}_p.} \]
It is, therefore, crucial that the domain of these functions are non-null strings; otherwise, the clitics might be stranded without a host, which is unacceptable. So, distinguishing null and non-null string types directly helps us capture an important generalization about Serbo-Croatian.

Some functions are defined on both null and non-null strings of a given type. For each type $\mathcal{P}$, the type $\text{Str}_{\mathcal{P}}$ is the type of null and non-null $\mathcal{P}$-strings, i.e. it is the coproduct of $\mathcal{C}_{\mathcal{P}}$ and $\mathcal{N}_{\mathcal{P}}$. The following functional terms are to be interpreted as canonical injections into $\mathcal{N}_{\mathcal{P}} + \mathcal{C}_{\mathcal{P}}$:

(12) a. $\lambda(i_{\mathcal{P}} x) : \mathcal{N}_{\mathcal{P}} \rightarrow (\mathcal{N}_{\mathcal{P}} + \mathcal{C}_{\mathcal{P}})$  
b. $\lambda(j_{\mathcal{P}} y) : \mathcal{C}_{\mathcal{P}} \rightarrow (\mathcal{N}_{\mathcal{P}} + \mathcal{C}_{\mathcal{P}})$

We will use the following type abbreviation:

(13) $s = \text{def } \mathcal{N}_{\mathcal{P}} + \mathcal{C}_{\mathcal{P}}$

In practice we will omit the term constructors $i$ and $j$, except when introducing formal definitions. For example, we will write terms such $\vdash e_{\mathcal{P}} : s$ and $\vdash e_{\mathcal{S}} : z$, instead of the technically correct $\vdash (i_{e_{\mathcal{P}}}) : s$ and $\vdash (i_{e_{\mathcal{S}}}) : z$; similarly for injected non-null strings.

Next we define concatenation for arbitrary $\mathcal{P}$-strings.

(14) $\vdash \cdot_{\mathcal{P}} : (\mathcal{N}_{\mathcal{P}} + \mathcal{C}_{\mathcal{P}}) \rightarrow (\mathcal{N}_{\mathcal{P}} + \mathcal{C}_{\mathcal{P}}) \rightarrow (\mathcal{N}_{\mathcal{P}} + \mathcal{C}_{\mathcal{P}})$

In practice, we use the following constant for $\mathcal{P}$-string concatenation, and omit the subscript.

(15) $\vdash \cdot : s \rightarrow s \rightarrow s$
Our basic higher order theory of string concatenation can then be stated as follows:

\begin{equation}
\begin{aligned}
a. & \quad \forall_a : p [ (tosp \ a) = (cns_p \ a \ (i \ e_p) ) ] \\
b. & \quad \forall_a : p [ (tosp \ a) = (snc_p \ (i \ e_p) \ a) ] \\
c. & \quad \forall_{s : n_p + c_p} [ (i \ e_p) \cdot p \cdot s = s ] \\
d. & \quad \forall_{s : n_p + c_p} [ s \cdot p \cdot (i \ e_p) = s ] \\
e. & \quad \forall_{s : n_p + c_p} \forall_{c : c_p} [ (j \ c) \cdot p \cdot s = (j \ (cns_p \ (fst_p \ c) \ ((rst_p \ c) \cdot p \cdot s)) ) ]
\end{aligned}
\end{equation}

Concatenation allows us to combine strings into larger strings. For example, if $\vdash marko : s$ and $\vdash spava : s$, i.e. they are both $p$-strings, then we can concatenate them and construct a $p$-string $\vdash marko \cdot spava : s$.

Concatenation is associative, meaning that if $s, t, u$ are strings, $(s \cdot t) \cdot u = s \cdot (t \cdot u)$, i.e. rebracketing is allowed. However, concatenation is not commutative, so a string $s \cdot t$ is not the same as the string $t \cdot s$, i.e. reordering of smaller strings inside a string is not allowed.

Given the type of all $p$-strings, $s$, our framework provides us with the notion of $p$-languages, that is, sets of $p$-strings. Since $t$ is the type of truth values, $s \rightarrow t$ is the type of sets of $p$-strings, or the type of $p$-languages. We abbreviate $s \rightarrow t$ as $S$.

For example, since $\vdash grad : s$, $\vdash \lambda_s. s = grad : s$ is the singleton $p$-language which contains the string $grad$. Similarly, since $\vdash marko \cdot spava : s$, $\vdash \lambda_s. s = marko \cdot spava : s$ is the singleton $p$-language which contains the string $marko \cdot spava$. 
We sometimes abbreviate terms corresponding to singleton \( p \)-languages by writing the \( p \)-string in uppercase. For example:

(17) \( \text{GRAD} = \text{def} \lambda_s.s = \text{grad}_s \)

For \( p \)-languages, we define the operation of \( p \)-language fusion:

(18) a. \( \vdash S \cdot : S \rightarrow S \rightarrow S \)

b. \( S \cdot = \text{def} \lambda_S.\exists_{st}[(S s) \land (T t) \land u = s \cdot t] \)

\( \vdash 0_S : S \) is the unique empty \( p \)-language. \( \vdash 1_S : S \) is the unique \( p \)-language which contains only the null \( p \)-string.

This operation takes some two sets of \( p \)-strings and outputs another set of \( p \)-strings by concatenating all the strings in the input sets of strings. For example, if \( S \) denotes the set of strings \{markos, anas\}, and \( T \) denotes the set of strings \{spavas, radis\}, then \( S \cdot T \) denotes the set of strings:

(19) \{markos \cdot spavas, markos \cdot radis, anas \cdot spavas, anas \cdot radis\}

So, the operation of \( p \)-language fusion is itself defined in terms of concatenation. The type \( S \), given that it denotes sets of \( p \)-strings, allows us to express phenogrammatical indeterminacy. Below we give a toy example, to illustrate how the phenogrammatical machinery laid out so far can be used to control word order patterns.

Suppose a noun and an adjective could reorder with respect to one another but had to remain adjacent, so their attachment was still relatively strong. For example, \textit{dobar} ‘good’, and \textit{drug} ‘friend’, could be pronounced both as \textit{dobar drug}
and *drug dobar*. We can represent this expression in the grammar as a term of type $S$, i.e. a set of two $p$-strings, $\{dobar_s \cdot drug_s, drug_s \cdot dobar_s\}$.

Suppose we now want to add a postnominal modifier to this phrasal noun, and the postnominal modifier must occur to the right of the noun+adjective combination, whichever order they are pronounced in. For example, if we add *iz Beograda* ‘from Belgrade’, there are two ways we could pronounce the resulting phrasal noun: *dobar drug iz Beograda*, or *drug dobar iz Beograda*, with the meaning ‘good friend from Belgrade’.

Since *iz* ‘from’ is a proclitic, we represent *iz Beograda* ‘from Belgrade’ as a phonological word, *iz#beograda*. Recall that *dobar drug* ‘good friend’ is represented as a set of two $p$-strings. To be able to use $p$-language fusion, we have to ‘zoom out’ and reconceptualize *iz Beograda* ‘from Belgrade’ as a singleton $p$-language. The following term represents this prepositional phrase in its disguise as a $p$-language with only one member: $\lambda_s.s = (iz#beograda)_s$.

Now when we use $p$-language fusion to combine the set $\{dobar_s \cdot drug_s, drug_s \cdot dobar_s\}$ with the set $\lambda_s.s = (iz#beograda)_s$. Fusion requires us to concatenate each member of the first set with each member of the second set of strings, so we wind up with the following set of exactly two $p$-strings, as we wanted:

(20) $\{dobar_s \cdot drug_s \cdot (iz#beograda)_s, drug_s \cdot dobar_s \cdot (iz#beograda)_s\}$

The phenogrammatical types we have discussed so far, $p$, $c$, $s$ and $S$, are the lower types in the type hierarchy. Since functions on these types are very restrictive,
at this level we enforce strict ordering of expressions, such as clitic attachment (via the function #) and postnominal modifier placement (via p-string concatenation). However, the type S already allows us to implement a limited notion of phenogrammatical indeterminacy, by allowing us to phenogrammatically treat expressions as sets of their possible pronunciation.

As we move up in the type hierarchy, we will encounter more permissive functions, which can reorder and otherwise transform the linear order of expressions. Since these ‘disordering’ functions are defined on higher level types, they cannot destroy any existing structure, established on the level of clitics and phonological words, or strings of phonological words.

2.2.4 Higher Level Types and Functions

Just as we built up the type of p-strings (Np, Cp, and s), we construct the types of S-strings, i.e. strings of p-languages, namely NS, CS, and NS+CS. We have a function toZ which maps p-languages to strings of p-languages, defined analogously to toS. We use the following abbreviations:

(21)  a.  z =def NS + CS

b.  toZ =def toS

The unique member of NS is the null S-string ⊨ eS : NS. We define concatenation for S-strings, analogous to p-string concatenation, and notate it as follows:

(22)  ⊨ o : z → z → z
The functions which add and remove prefixes and suffixes of $p$-strings, also have their $S$-string counterparts. For example, the analogue of $\text{cns}_p$ is $\text{cns}_S$.

For strings of $S$-languages we also employ the following notational convention:

\[(\text{23}) \quad \text{for } \phi \text{ a term of type } S, \phi_z \overset{\text{def}}{=} (\text{to}Z \phi)\]

For example, viewed as a length one string of $p$-languages, the expression $\text{grad} '\text{city}'$ is represented as follows:

\[(\text{24}) \quad \vdash \text{GRAD}_z : z\]

Finally, just as we could define sets of $p$-strings (i.e. $p$-languages), we can define sets of $S$-strings (i.e. $S$-languages). The type of $S$-languages is $z \rightarrow t$ abbreviated as $Z$.

\[(\text{25}) \quad z \rightarrow t =_{\text{def}} Z\]

Seen as a set of $S$-languages, the expression $\text{grad} '\text{city}'$ is represented as follows:

\[(\text{26}) \quad \vdash \lambda z.z = \text{GRAD}_z : Z\]

The counterpart of $p$-language fusion is $\vdash \bullet : Z \rightarrow Z \rightarrow Z$, the $S$-language fusion. The empty $S$-language is $\vdash 0_Z : Z$ and the singleton language which contains only the null $S$-string, is $\vdash 1_Z : Z$.

We use a shuffle function $\vdash \odot : Z \rightarrow Z \rightarrow Z$, which outputs a set of all possible shuffles of its two arguments. We reconceptualize domain union described in Reape (1993) as a function which outputs the set of all possible strings resulting
from interleaving the two argument strings. For example, if \( x = v \circ w \), and \( x' = y \circ z \), then \( x \odot x' \) denotes the following set of strings:

\[
\{ v \circ w \circ y \circ z, v \circ y \circ w \circ z, v \circ y \circ z \circ w, y \circ v \circ w \circ z, y \circ v \circ z \circ w, y \circ z \circ v \circ w \} 
\]

So, shuffling \( x \) into \( x' \) means constructing a set of \( S \)-strings in which smaller strings in \( x \) and \( x' \) can reorder with respect to one another so long as the relative order of the smaller strings in \( x \) and \( x' \) is retained. Going back to the example above, \( w \circ v \circ y \circ z \) is not in the set denoted by \( x \odot x' \), because in \( x \), \( v \) precedes \( w \).

To define \( \odot \) more formally, we need a pair of auxiliary functions, called language union, which construct a single language out of two languages.

\[
\begin{align*}
\text{a. } & \cup_S = \text{def } \lambda_{STs}((Ss) \cup (Ts)) \\
\text{b. } & \cup_Z = \text{def } \lambda_{VWx}((Vx) \cup (Wx))
\end{align*}
\]

For example, if \( P \) contains the strings \( s \cdot t \) and \( u \), and \( T \) contains the strings \( p \cdot u \) and \( t \), then \( P \cup_S T \) is the set of all strings that are either in \( P \) or \( T \), i.e. it denotes the set \( \{ s \cdot t, u, p \cdot u, t \} \).

Now, we define \( \odot \) as follows (see Jędrzejowicz and Szepietowski (2001)):

\[
\begin{align*}
\text{a. } & \vdash \forall_{v,z}:[(v \odot (i e_S)) = ((i e_S) \odot v) = (\lambda_{x}.x = v)] \\
\text{b. } & \vdash \forall_{u,v,z,a,b,c}:[((j a \circ u) \odot (j b \circ v)) = [(\lambda_{x}.x = j a) \bullet (u \odot (j b \circ v))] \cup_Z [(\lambda_{x}.x = j b) \bullet ((j a \circ u) \odot v)]
\end{align*}
\]

We will use \( \odot \) to introduce adverbials into a clause. By shuffling them in, we allow them to occur pretty much anywhere without giving them the power to reorder the clause. Further, multi-word prepositional adverbials may occur discontinuously in a clause, so long as the phonological word of the noun phrase to which
the preposition procliticized precedes the rest of the noun phrase. By shuffling in an adverbal prepositional phrase, we allow such discontinuities but require that the relative order of expressions in the prepositional phrase be maintained.

We also introduce the function \( \text{PER} : \mathbf{z} \rightarrow \mathbf{z} \), called permutation, which takes some \( \mathbf{S} \)-string (of type \( \mathbf{z} \)) and constructs an \( \mathbf{S} \)-languages (of type \( \mathbf{z} \)) consisting of all possible reorderings of the smaller \( \mathbf{S} \)-strings in the original \( \mathbf{S} \)-string, i.e. the set of all permutations of its argument. For example, \( \text{PER}(x \circ y \circ z) \) denotes a set of six \( \mathbf{S} \)-strings, namely \( \{x \circ y \circ z, x \circ z \circ y, y \circ x \circ z, z \circ x \circ y, z \circ y \circ x\} \).

To define \( \text{PER} \) more formally, we first generalize \( \odot \) to language shuffle operation. Intuitively, shuffling two languages together results in a set which contains shuffles of all the strings in the two argument languages.

\[
\begin{align*}
(30) & \quad a. \quad \odot_1 : \mathbf{Z} \rightarrow \mathbf{Z} \rightarrow \mathbf{Z} \\
& \quad b. \quad \odot_1 = \text{def} \, \lambda_{\mathbf{vwX}}. \exists_{\mathbf{vw}} (\mathbf{v} \circ \mathbf{w}) \land (\mathbf{w} \circ \mathbf{v})
\end{align*}
\]

Now we recursively define \( \text{PER} \) in terms of \( \odot_1 \):

\[
(31) \quad a. \quad \vdash (\text{PER} i \in \mathbf{S}) = 1_\mathbf{S} \\
& \quad b. \quad \vdash \forall_{\mathbf{a}, \mathbf{c}_\mathbf{s}, \mathbf{v}, \mathbf{z}}. \text{PER}(\mathbf{j}_\mathbf{a} \circ \mathbf{v}) = [(\lambda_{\mathbf{x}}. \mathbf{x} = \mathbf{j}_\mathbf{a}) \odot_1 (\text{PER} \, \mathbf{v})]
\]

This function allows us to model free word order. It is crucial to note that it only allows free reordering of terms of type \( \mathbf{z} \). Any lower level structure within an individual term of type \( \mathbf{z} \), such as attached clitics or concatenated \( \mathbf{p} \)-strings, cannot be destroyed by \( \text{PER} \), because \( \text{PER} \) is oblivious to such lower level restrictions. Below we give two examples to illustrate how this function works.
If we view the name Marko and the verb spava ‘sleeps’ as S-strings, i.e. as terms \( \vdash \text{MARKO}_z : z \) and \( \vdash \text{SPAVA}_z : z \), respectively, then we can concatenate them into an S-string \( \vdash \text{MARKO}_z \circ \text{SPAVA}_z : z \). If we apply PER to this string, the result is the following set of S-strings: \( \{ \text{MARKO}_z \circ \text{SPAVA}_z, \text{SPAVA}_z \circ \text{MARKO}_z \} \). We can take this term to phenogrammatically represent the sentence Marko spava ‘Marko sleeps’; in doing so, we are again introducing phenogrammatical indeterminacy by claiming that there is a single expression which can be pronounced multiple different ways.

Consider the expression drug iz Beograda ‘friend from Belgrade’. Recall that iz ‘from’ is a proclitic, and that the postnominal modifier iz Beograda ‘from Belgrade’ must occur immediately to the right of the noun drug ‘friend’. The clitic attaches to its host Beograda via #, and then we construct a p-string out of the resulting phonological word, which gives us \((iz\#\text{beograda})_s\). We also construct a p-string out of the noun, which gives us \((\text{drug})_s\), and then we concatenate it with \((iz\#\text{beograda})_s\). This results in the p-string \((\text{drug})_s \cdot (iz\#\text{beograda})_s\).

Next, we can proceed to construct a higher level term out of this p-string, first by constructing a p-language which contains precisely that string, then by mapping that p-language to a length one S-string, resulting in the following term:

\[ \vdash (\lambda s. s = \text{drug}_s \cdot (iz\#\text{beograda})_s)_z : z \]

Now we can concatenate this S-string with SPAVA\(_z\), and apply PER to the resulting S-string, which gives us the following term:

\[ \vdash \text{PER}((\lambda s. s = \text{drug}_s \cdot (iz\#\text{beograda})_s)_z \circ \text{SPAVA}_z) : z \]
However, \textsc{per} ‘sees’ only individual \textit{s}-strings and not their internal structure. Therefore, it can only place \textit{spava}_z to the left or to the right of \((\lambda_s.s = \text{drug}_s \cdot (iz\#\text{beograda})_s)_z\), but it cannot make the latter discontinuous, nor can it change its internal order, because \textsc{per} ‘sees’ it as a length one \textit{s}-string.

Next, we define a function \(\vdash L : z \rightarrow s\), called linguification, which takes an \textit{s}-string and outputs a set of \textit{p}-strings.

\[(32)\]
\[
\begin{align*}
\text{a. } (L e_s) &= 1_s \\
\text{b. } \forall_{P_z} (L (cns_s P z)) &= P \cdot (L z)
\end{align*}
\]

For example, \((L (\text{marko}_z \circ \text{spava}_z))\) is \((\text{marko} \cdot \text{spava})\) which is the set of \textit{p}-strings that contains exactly one string, \textit{marko} \cdot \textit{spava}. This function will help us create multi-word ‘fortresses’, which cannot be made discontinuous, out of \textit{s}-strings. Recall that some functions, such as \textsc{per} and \(\odot\) ‘see’ only \textit{s}-strings. Suppose we have an \textit{s}-string of length greater than two, but we want to keep it intact in the presence of a ‘disordering’ function. We can apply \(L\) to that string, turning it into a \textit{p}-language. Then we can use \(\text{toz}\) to map the resulting \textit{p}-language into a length one \textit{s}-string. As a result, \textsc{per} and \(\odot\) will be unable to make it discontinuous or reorder it internally.

We define another function, \(k\), which serves a similar purpose to \(L\), but applies to \textit{s}-languages, not \textit{s}-strings. \(\vdash k : z \rightarrow s\), called compaction, takes a set of \textit{s}-strings and then unions the linguifications of all \textit{s}-strings in that set resulting in a \textit{p}-language.

\[(33)\]
\[
\begin{align*}
\text{a. } \vdash (k 0_z) &= 0_s
\end{align*}
\]
b. $\forall_{x_\cdot}(k(x \cup_{x} \lambda_{x}.z = v)) = (k \cdot \cup_{x}(L \cdot v))$

For example, $k(\text{PER} (\text{MARKO}_{z} \circ \text{SPAVA}_{z}))$ denotes the set of $p$-strings $\{\text{marko}_{s} \cdot \text{spava}_{s}, \text{spava}_{s} \cdot \text{marko}_{s}\}$.

We analyze finite verbs as building $S$-languages, phenogrammatically. So, an embedded clause will be represented as a term of type $z$. When en embedding verb picks up its complement clause, it will compact it via $k$, then turn the resulting $p$-language into a length $S$-string via $toZ$. This procedure creates another word order ‘fortress’ out of an $S$-language (much like $L$ and $toZ$ do out of $S$-strings), and protects the embedded clause from matrix clause intrusions.

Table 2.1 summarizes the phenogrammatical type hierarchy. Table 2.2 lists all logical constants that we introduced, and Table 2.3 summarizes all the phenogrammatical functions. Table 2.4 gives a more informal correspondence between different phenogrammatical types and the sort of word order phenomenon modeled at that type level, in addition to functions defined at that type level and functions which traverse the type hierarchy.
<table>
<thead>
<tr>
<th>TYPE:</th>
<th>REFERRED TO AS:</th>
<th>EXAMPLE:</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOWEST</td>
<td>( c )</td>
<td>clitics</td>
</tr>
<tr>
<td></td>
<td>( p )</td>
<td>phonological words</td>
</tr>
<tr>
<td></td>
<td>( s )</td>
<td>strings of phonological words; ( p )-strings</td>
</tr>
<tr>
<td></td>
<td>( S )</td>
<td>sets of ( p )-strings; ( p )-languages</td>
</tr>
<tr>
<td></td>
<td>( z )</td>
<td>strings of ( p )-languages; ( S )-strings</td>
</tr>
<tr>
<td>HIGHEST</td>
<td>( Z )</td>
<td>sets of ( S )-strings; ( S )-languages</td>
</tr>
</tbody>
</table>

Table 2.1: Phenogrammatical type hierarchy.

<table>
<thead>
<tr>
<th>CONSTANT</th>
<th>NOTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \vdash e_p : s )</td>
<td>null ( p )-string; identity for ( p )-string concatenation (( \cdot ))</td>
</tr>
<tr>
<td>( \vdash e_S : z )</td>
<td>null ( S )-string; identity for ( S )-string concatenation (( \circ ))</td>
</tr>
<tr>
<td>( \vdash 0_S : S )</td>
<td>the empty ( p )-language</td>
</tr>
<tr>
<td>( \vdash 0_Z : Z )</td>
<td>the empty ( S )-language</td>
</tr>
<tr>
<td>( \vdash 1_S : S )</td>
<td>the singleton ( p )-language; identity for ( p )-language fusion (( \circ_S ))</td>
</tr>
<tr>
<td>( \vdash 1_Z : Z )</td>
<td>the singleton ( S )-language; identity for ( S )-language fusion (( \circ_Z ))</td>
</tr>
</tbody>
</table>

Table 2.2: Phenogrammatical logical constants.
<table>
<thead>
<tr>
<th>FUNCTION</th>
<th>NOTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>† #ₚc : c → p → p</td>
<td>procliticization to a phonological word</td>
</tr>
<tr>
<td>† #ₑc : p → c → p</td>
<td>encliticization to a phonological word</td>
</tr>
<tr>
<td>† · : s → s → s</td>
<td>concatenation for p-strings</td>
</tr>
<tr>
<td>† o : z → z → z</td>
<td>concatenation for S-strings</td>
</tr>
<tr>
<td>† ⋊ₚ : S → S → S</td>
<td>p-language fusion</td>
</tr>
<tr>
<td>† ⋊ₗ : Z → Z → Z</td>
<td>S-language fusion</td>
</tr>
<tr>
<td>† Uₚ : S → S → S</td>
<td>p-language union</td>
</tr>
<tr>
<td>† Uₗ : Z → Z → Z</td>
<td>S-language union</td>
</tr>
<tr>
<td>† cⁿˢₚ : p → s → Cₚ</td>
<td>prefixes a p-string</td>
</tr>
<tr>
<td>† cⁿˢₛ : S → z → Cₛ</td>
<td>prefixes an S-string</td>
</tr>
<tr>
<td>† fˢᵗₚ : Cₚ → p</td>
<td>outputs the prefix of a p-string</td>
</tr>
<tr>
<td>† fˢᵗₛ : Cₛ → S</td>
<td>outputs the prefix of an S-string</td>
</tr>
<tr>
<td>† rˢᵗₚ : Cₚ → s</td>
<td>outputs a p-string without its prefix</td>
</tr>
<tr>
<td>† rˢᵗₛ : Cₛ → z</td>
<td>outputs an S-string without its prefix</td>
</tr>
<tr>
<td>† sⁿˢₚ : p → s → Cₚ</td>
<td>suffixes a p-string</td>
</tr>
<tr>
<td>† sⁿˢₛ : S → z → Cₛ</td>
<td>suffixes an S-string</td>
</tr>
<tr>
<td>† lˢᵗₚ : Cₚ → p</td>
<td>outputs the suffix of a p-string</td>
</tr>
<tr>
<td>† lˢᵗₛ : Cₛ → S</td>
<td>outputs the suffix of an S-string</td>
</tr>
<tr>
<td>† tˢʳₚ : Cₚ → s</td>
<td>outputs a p-string without its suffix</td>
</tr>
<tr>
<td>† tˢʳₛ : Cₛ → z</td>
<td>outputs an S-string without its suffix</td>
</tr>
<tr>
<td>† L : z → S</td>
<td>constructs a p-language out of a string of p-languages</td>
</tr>
<tr>
<td>† k : Z → S</td>
<td>constructs a p-language out of an S-language</td>
</tr>
<tr>
<td>† PER : z → Z</td>
<td>constructs the set of all permutations of some S-string</td>
</tr>
<tr>
<td>† o : Z → z → Z</td>
<td>constructs an S-language by interleaving two S-strings</td>
</tr>
</tbody>
</table>

Table 2.3: Phenogrammatical functions.
Table 2.4: Phenogrammatical types, functions and word order phenomena represented at that level.

2.3 Tectogrammar

2.3.1 Preliminaries

The tectogrammatical component of the grammar is not concerned with word order but primarily with argument/functor relations. Since expressions have different argument requirements, we distinguish between various tectogrammatical types of expressions, where each type roughly corresponds to a syntactic category.

We take the stance that since inflectional features, such as case, gender, number and person, influence the argument requirements of an expression, they need to be represented in the tectogrammar. For example, the verb *spavam* ‘sleep’ needs a
first person singular subject, but spava ‘sleeps’ needs a third person singular sub-
ject. In our grammar, they will be treated as distinct expressions assigned to dif-
ferent tectogrammatical types since they have different argument requirements.

2.3.2 Representing Inflectional Features

To represent inflectional features, we introduce special tectogrammatical types
for each kind of feature. These types are special because no linguistic expressions
are assigned to these types—they don’t directly correspond to syntactic categories.
However, they will help us encode inflectional information on other tectogram-
matical types, that is, terms of the inflectional types will serve as parameters that
define families of tectogrammatical types.5

Cse is the type of case features. Terms of this type represent specific case val-
ues: nom for nominative, gen for genitive, dat for dative, acc for accusative and
inst for instrumental.

Gdr is the type of gender features. Terms of this type represent specific gender
values: m for masculine, f for feminine and n for neuter.

Terms of type Prs are 1, 2 and 3 for first, second and third person.

Finally, terms of type Num are sg for singular and pl for plural.

5See de Groote and Maarek (2007) for a similar representation of inflectional features in a
dependent type system.
2.3.3 N and NP type families

All nouns are assigned to a type in the N family. The types in this family are parametrized by terms of type Cse, Gdr and Num, since these features are relevant for determiner and adjective agreement with nouns. So, for example, $N_{nom,f,sg}$ is the type of nominative singular feminine nouns, $N_{inst,m,pl}$ is the type of instrumental plural masculine nouns, and so on.

All noun phrases, whether lexical or phrasal, are assigned to a type in the NP family. The types in this type family are parametrized by terms of type Cse, Gdr, Num and Prs, since these features are relevant for subject-verb agreement, and object selection. $NP_{dat,n,sg,3}$ is, for example, the type of dative neuter third person singular noun phrases, and $NP_{acc,f,pl,3}$ is the type of accusative feminine third person plural noun phrases.

For a more detailed exposition, as well as arguments for representing all these features on noun and noun phrase types, we direct the reader to Chapter 3.

2.3.4 S type family

All clauses are assigned to a type in the S family. This family of types is parametrized by terms of type K and Nat. Terms of type K encode different types of clauses: m(ain), e(mbedded), q(uestions) or inf(initival).

The type Nat is the type of natural numbers. As parameters, they will be used to enforce the placement and ordering of enclitics in a clause. This strategy will be
explained in more detail in Chapter 5 (see also Morrill and Gavarró (1992) for a similar use of natural number parameters). For now we just mention that clauses with no enclitics placed inside of them are associated with the parameter 6, and as more and more enclitics are placed in the clause, the natural number parameter is lowered, so that a clause whose parameter is 0 cannot have any more clitics placed inside of it.

2.3.5 \( \rightarrow o \) types

We can construct more complex tectogrammatical types out of types in the \( N, NP \) and \( S \) family with the type constructor \( \rightarrow o \), the linear implication.\(^6\) Such implicative types encode syntactic dependencies between expressions and their arguments.

For example, an intransitive verb which needs a subject noun phrase to construct a sentence would be associated with the type \( NP_{nom} \rightarrow o \ S_m \), a complementizer which converts a main clause into an embedded clause with the type \( S_m \rightarrow o \ S_e \), and a determiner which constructs noun phrases out of nouns with the type \( N \rightarrow o \ NP \).\(^7\)

\(^6\)Among categorial frameworks which also use linear implication as the type constructor in the tectogrammatical component are ACG (de Groote (2001)) and Lambda Grammar (Muskens (2003, 2007b)), though the practice goes back to Oehrle (1994). Since linear implication is insensitive to the order of hypotheses, it can only be used as the main tectogrammatical type constructor in frameworks which distinguish between phenogrammar and tectogrammar, where phenogrammar is designated to keep track of linear order, and so the tectogrammatical component need not be order-sensitive.

\(^7\)In these examples of functional types we abstracted away from many type parameters for illustrative purposes.
2.3.6 $\prod$ types

While it is necessary to keep track of inflectional features of expressions, many expressions are vague with respect to some subset of inflectional features. For example, present tense verbs in Serbo-Croatian require a nominative subject of a specific number and person, but they do not care what the gender of their subject noun phrase is. Past participles, on the other hand, require that their subject be nominative and of a specific gender and number, but do not care about the subject’s person. Prepositions require a noun phrase argument of a specific case, but do not care about their argument’s number, gender or person.

Continuing with the example of present tense verbs, one option would be to simply list each version of the verb. For example, spava ‘sleeps’ requires a third person singular nominative subject of any gender. Since there are three genders, we could list three tectogrammatical versions of this verb, one for each gender:

(34)  

a. $\vdash$ spava$^f$ : $\text{NP}_{\text{nom},f,sg,3} \rightarrow S_{m,6}$  
b. $\vdash$ spava$^m$ : $\text{NP}_{\text{nom},m,sg,3} \rightarrow S_{m,6}$  
c. $\vdash$ spava$^n$ : $\text{NP}_{\text{nom},n,sg,3} \rightarrow S_{m,6}$

However, listing each version of a vague expression in the lexicon is not just an inelegant solution that substantially increases the number of lexical entries in the grammar; if we were to do that we would in a sense also be missing a linguistic generalization.

Following de Groote and Maarek (2007), we address this problem by introducing dependent product types, which, intuitively, help us abstract away from the
value of some parameter, and allow us to obtain more specific versions of lexical entries by supplying that value.

So, a more expedient way to represent the same verb *spava* ‘sleeps’ tectogrammatically is as follows:

\[(35) \text{a. } \vdash \lambda_{g:Gdr}spava_g : \Pi_{g:Gdr}[NP_{\text{nom},g,sg,3} \rightarrow S_{m,o}]\]

This term is not specified for gender. But, given the three terms of type $Gdr$, m, f, and n we can obtain three more specific versions of this term via the product elimination rule (which we return to in later in the chapter). So, instead of being non-logical axioms that have to be asserted, the three tectogrammatical versions of this verb listed in (34) are now theorems.

The product types should be understood as universally quantifying over all terms of a given type. So the tectogrammatical term in (35) above means that for each $\tau$ of type $Gdr$, there’s a more specific version of the term that has precisely $\tau$ as its gender parameter.

While the parametrization of types essentially allowed us to simulate subtyping by allowing us to define type families, product types let us ‘underspecify’ tectogrammatical terms for certain features.

### 2.4 Semantics

#### 2.4.1 Preliminaries

We assume a hyperintensional semantic theory along the lines of Pollard (2008a) and Plummer and Pollard (2012) (see Thomason (1980) and Muskens (2005),
Muskens (2007a) for versions of hyperintensional semantics with somewhat different technical assumptions). While we believe this choice to be well motivated (we direct the reader to Pollard (2008a) for a detailed discussion of problems with traditional possible world semantics), our choice of semantic theory is inessential in the context of this project; our theory of Serbo-Croatian grammar is equally compatible with a more mainstream Montague-style possible world semantics.

2.4.2 Entailment

While in standard possible world semantics, the type of possible worlds is treated as basic, and propositions are analyzed as sets of possible worlds, on our approach propositions are taken as basic and the type of possible worlds is defined to be a certain sets of propositions.

This set up has (desirable) consequences regarding entailment. In standard possible world semantics, entailment is modeled as subset inclusion, an antisymmetric relation on sets of worlds. This prevents us from distinguishing mutually entailing propositions, since they are represented in the theory as the same set of possible worlds.

On our approach, however, propositions are modeled as members of a pre-boolean algebra pre-ordered by entailment. Entailment is axiomatized as a reflexive, transitive, but not antisymmetric relation on propositions. This way, it is possible for equivalent (mutually entailing) propositions to be distinct.
2.4.3 Types

The hyperintensional semantic theory that we are using is expressed in classical higher order logic. The underlying logic provides us with the basic types $\mathbf{t}$ (truth values) and $\mathbf{n}$ (natural numbers).

In addition to these types, we introduce as basic types $\mathbf{p}$ (propositions) and $\mathbf{e}$ (individual entities).\(^8\) We call the types $\mathbf{p}$ and $\mathbf{e}$, and any implicative types constructed out of these by means of the type constructor $\to$, hyperintensional types. These types are used to model linguistic meanings.

We recursively define the function $\text{Ext}$ mapping hyperintensional types to the corresponding extensional types. Here, $S$ and $S'$ are metavariables over hyperintensional types:

\[
\begin{align*}
\text{a. } \text{Ext}(\mathbf{e}) &= \mathbf{e} \\
\text{b. } \text{Ext}(\mathbf{p}) &= \mathbf{t} \\
\text{c. } \text{Ext}(S \to S') &= S \to (\text{Ext}S')
\end{align*}
\]

The type of possible worlds $\mathbf{w}$ is constructed out of the basic types in such a way that the interpretation of the type $\mathbf{w}$ is the set of ultrafilters of the pre-boolean prealgebra that interprets the type $\mathbf{p}$. Specifically, $\mathbf{w} =_{\text{def}} [\mathbf{p} \to \mathbf{t}]_u$, where $\vdash u : (\mathbf{p} \to \mathbf{t}) \to \mathbf{t}$ is a predicate on sets of propositions that picks out those sets of propositions that are ultrafilters (see Pollard (2008a) for details of this construction).

\(^8\)For expository simplicity, we depart from Pollard (2008a) in not distinguishing between the extensional type $\mathbf{e}$ and the corresponding hyperintensional type $\mathbf{i}$ (individual concepts). In particular, the meaning of a name is the same as its reference. Plummer and Pollard (2012) adopt the same practice as us, so that $\text{Ext}(\mathbf{e}) = \mathbf{e}$.
Concomitantly, we introduce a family of constants \( \text{ext}_S : S \rightarrow w \rightarrow (\text{Ext}S) \)
(where \( S \) is a variable ranges over the hyperintensional types) interpreted as a
polymorphic function that maps a hyperintension and a world to the extension of
that hyperintension at that world, as follows:

\[
\begin{align*}
\text{(37)} & \quad \forall x : e \forall w : w \ [(\text{ext}_e x w) = x] \\
& \quad \forall p : p \forall w : w \ [(\text{ext}_p p w) = p@w] \\
& \quad \forall f : S \rightarrow S' \forall w : w \ [(\text{ext}_{S \rightarrow S'} f w) = \lambda x : S (\text{ext}_S (f x) w)]
\end{align*}
\]

Here the notation ‘\( p@w \)’ abbreviates \( (\mu_u wp) \), where \( \mu_u \) denotes the embedding
of the set of worlds into the set of sets of propositions.\(^9\)

In general, however, we will only refer to extensional types in axioms which
relate special hyperintensional constants (e.g. \( \exists \), \( \& \), etc.) to their extensional
counterparts. Representations of linguistic meaning in the grammar are hyperin-
tensional terms, and here we only mention how these can be extensionalized for
completeness’ sake.

2.5 Putting it all together

2.5.1 Signs

Lexical entries are triples of typed lambda terms, however, in practice, we of-
ten omit tectogrammatical terms altogether and write lexical entries in the follow-
ing format:

\(^9\)The kind of HOL we employ follows Lambek and Scott (1986) in having machinery for form-
ing (separation-style) subtypes. Thus, if \( S \) is a type and \( \sigma \) an \( S \)-predicate (term of type \( S \rightarrow t \)), then
there is a type \( \mathcal{S}_\sigma \) interpreted as the subset of the interpretation of \( S \) that has the interpretation of
\( \sigma \) as its characteristic function; and there is a constant \( \mu_\sigma \) that denotes the subset embedding.
We call a sign any such triple of typed terms, including lexical entries (non-logical axioms) as well as larger expressions constructed out of the lexical entries via the inference rules (essentially, theorems of our grammar).

Sometimes, we will also omit the phenogrammatical and the semantic type and write signs in the following form:

(39) $\Gamma \vdash \text{pheno term}; \text{tecto type}; \text{semantic term}$

Table 2.5 declares typesetting conventions for object language or metalanguage expressions for each of the three calculi, which we have thus far been using implicitly.

### 2.5.2 Rules

The rules operate on signs. The grammar has three logical rules governing the behavior of the implicative type constructors in the type system of each component. We will, however, present them omitting the phenogrammatical and semantic type, and the name we use for the rules references the tectogrammatical type constructor $\rightarrow^o$.

$\Gamma$ is a metavariable over contexts which are multisets of triples of typed variables, and while lexical entries, as non-logical axioms in our theory, typically do not depend on any assumptions, our analysis of the inherent reflexive in Chapter 5 will require that a lexical entry have a non-empty context, i.e. introduce a hypothesis.
<table>
<thead>
<tr>
<th>TECTOGRAMMAR</th>
<th>SEMANTICS</th>
<th>PHENOGRAMMAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>terms</td>
<td>acc, pl</td>
<td>ana, (love ana)</td>
</tr>
<tr>
<td>types</td>
<td>N, NP, S, Gdr, Num, Prs, Cs</td>
<td>e, p, t</td>
</tr>
<tr>
<td>term</td>
<td>u, v, w, x, y, z</td>
<td>x, y, z; P, Q, p, q</td>
</tr>
<tr>
<td>variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>type</td>
<td>T', T'', ...</td>
<td>S', S'', ...</td>
</tr>
<tr>
<td>variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>term</td>
<td>t', t'', ...</td>
<td>t', t'', ...</td>
</tr>
<tr>
<td>metavariables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>variable</td>
<td>v', v'', ...</td>
<td>v', v'', ...</td>
</tr>
<tr>
<td>metavariables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>type</td>
<td>T', T'', ...</td>
<td>S', S'', ...</td>
</tr>
<tr>
<td>metavariables</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.5: Typesetting conventions.
The \([\text{Ax}]\) rule allows us to introduce hypotheses (i.e. triples of typed variables). Once a hypothesis is introduced, it is stored in the context. Informally, this rule allows us to introduce traces or gaps.

\[
(40) \quad \frac{\nu \vdash T; \nu}{\nu; T \vdash \nu} \quad \text{[Ax]}
\]

Informally, the \([\neg \circ \text{E}]\) rule allows two signs to combine into a larger sign, so long as the tectogrammatical type of one sign \((T)\) is the argument type of the other sign \((T' \to T')\). It is our analogue of Merge in MGG.

More formally, this rule is just the implication elimination rule for each of the three calculi, and in the term calculi it is accompanied by function application. \(\cup\) is supposed to denote multiset union.

\[
(41) \quad \frac{\Gamma \vdash \phi; T \to T'; \sigma \quad \Delta \vdash \phi'; T; \sigma'}{\Gamma \cup \Delta \vdash (\phi \phi'); T'; (\sigma \sigma')} \quad \text{[\neg \circ \text{E}]}
\]

The \([\neg \circ \text{I}]\) rule allows us to discharge any hypotheses, i.e. bind traces. This rule consists of implication introduction in each of the three calculi, and in the term calculi it is accompanied by \(\lambda\) abstraction.

\[
(42) \quad \frac{\Gamma \cup \nu; T; \nu \vdash \phi; T'; \sigma}{\Gamma \vdash \lambda \nu. \phi; T \to T'; \lambda \nu. \sigma} \quad \text{[\neg \circ \text{I}]}
\]

For us, a declarative or interrogative sentence’s phenogrammatical type is \(Z\), i.e. phenogrammatically they denote \(S\)-languages. However, we want to establish a correspondence between \(\mathbf{p}\)-strings and pronunciations of actual linguistic expressions. To this end, we introduce a pair of non-logical rules which compact phenogrammatical terms of signs which either (a) have tectogrammatical type
and semantic type \( p \); or (b) have tectogrammatical type \( S_{q,T} \) and semantic type \( k \), where \( k \) is the type of interrogatives, defined as \( p \rightarrow p \) (see Chapter 6).

\[
\begin{align*}
(43) & \quad \text{a. } \Gamma \vdash \phi : z; S_{m,T}; \sigma : p \\
& \quad \Gamma \vdash k \phi : S; S_{m,T}; \sigma : p \quad \text{[km]} \\
& \quad \text{b. } \Gamma \vdash \phi : z; S_{q,T}; \sigma : k \\
& \quad \Gamma \vdash k \phi : S; S_{q,T}; \sigma : k \quad \text{[kq]}
\end{align*}
\]

So, when a representation of a sentence in the grammar is built up, phenogrammatically it denotes a set of \( S \)-strings. After undergoing one of these rules, it denotes a set of \( p \)-strings, where each \( p \)-string corresponds to one possible way of pronouncing that sentence. We won’t often use this rule in practice; the reader should keep in mind, however, that whenever we construct a clause of phenogrammatical type \( z \), we can always ‘linguify’ that set of \( S \)-strings into a set of \( p \)-strings, the level at which our theory technically establishes a correspondence to the empirical domain with respect to word order.
Chapter 3: Basic Word Order

3.1 Introduction

In this chapter, we analyze simple declarative Serbo-Croatian sentences consisting of an intransitive, transitive or a ditransitive verb, its object(s) and subject, and adverbial modifiers. We consider lexical noun phrases (names, pronouns and quantificational lexical noun phrases), as well as those consisting of a noun and possibly an attributive adjective, quantificational determiner or a postnominal modifier. Verbs which require complements other than noun phrases will not be considered here; instead, we return to those in later chapters.

Here, we also present a general theory of procliticization in Serbo-Croatian, since we will analyze prepositional adverbials and prepositional postnominal modifiers, and prepositions are proclitics in Serbo-Croatian. The enclitics, however, will not be considered in this chapter but in Chapter 5 which is entirely dedicated to encliticization in Serbo-Croatian.
The purpose of this chapter, other than to analyze word order in simple clauses, is to introduce basic mechanisms of combination in the grammar which are essential to the theory. As we analyze more complex constituents in the later chapters, we will retain the fundamental assumptions laid out in this chapter concerning the combination of verbs with their subjects and objects, and the construction of declarative clauses in general.

3.2 Data

3.2.1 Lexical Noun Phrases

Agreement

Lexical noun phrases in Serbo-Croatian (names and pronouns) are marked for case, number, gender and person. There are five distinct cases, two numbers, three persons for each number, and three genders. There is some syncretism in the paradigm. The examples of lexical noun phrases in Table 3.1 show some of the different combinations of these inflectional features. It is worth noting that only 3rd person noun phrases can be neuter, and that only pronouns can be 1st or 2nd person; that is, all non-pronominal noun phrases are 3rd person. Here we are

\footnote{Traditionally, Serbo-Croatian is said to have seven distinct cases. However, we will not consider vocative case-marked noun phrases as they are always extraclausal, that is, they never occur as arguments of any other expressions. Further, while traditionally dative and locative are considered functionally distinct case, there is no difference in form between dative and locative noun phrases (or nouns). Therefore, dative in our grammar subsumes the traditional dative and locative cases.}
**Table 3.1:** Examples of inflectional feature combinations on lexical noun phrases.

<table>
<thead>
<tr>
<th>CASE</th>
<th>GENDER</th>
<th>NUMBER</th>
<th>PERSON</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ane</td>
<td>genitive</td>
<td>feminine</td>
<td>sg</td>
</tr>
<tr>
<td>njemu</td>
<td>dative</td>
<td>masculine</td>
<td>sg</td>
</tr>
<tr>
<td>ono</td>
<td>nominative</td>
<td>neuter</td>
<td>sg</td>
</tr>
<tr>
<td>one</td>
<td>nominative</td>
<td>feminine</td>
<td>pl</td>
</tr>
<tr>
<td>Marka</td>
<td>genitive/accusative</td>
<td>masculine</td>
<td>sg</td>
</tr>
<tr>
<td>vi</td>
<td>nominative</td>
<td>masculine/feminine</td>
<td>pl</td>
</tr>
<tr>
<td>mmom</td>
<td>instrumental</td>
<td>masculine/feminine</td>
<td>sg</td>
</tr>
<tr>
<td>tobom</td>
<td>instrumental</td>
<td>masculine/feminine</td>
<td>sg</td>
</tr>
<tr>
<td>nama</td>
<td>dative/instrumental</td>
<td>masculine/feminine</td>
<td>pl</td>
</tr>
<tr>
<td>njih</td>
<td>genitive/accusative</td>
<td>masculine/feminine/neuter</td>
<td>pl</td>
</tr>
</tbody>
</table>

interested in the inflectional morphology in so far as it influences the distribution of noun phrases in the language, i.e. the syntactic properties of noun phrases.

Clearly, case influences the distributional properties of noun phrases since it determines whether they can be subjects, or objects of certain verbs or prepositions. For example, only nominative noun phrases can be subjects.\(^{12}\) The verb

\(^{12}\)In Serbo-Croatian, constructions with dative experiencers are pervasive. There are essentially two types of dative experiences constructions, (i) those which require a nominative argument as well, and (ii) those which do not. For example:

(i) Ani su trebali udžbenici.
   Ana_{DAT,f,sg,3} are_{3,pl} need_{ppl,m,pl} textbooks_{NOM,m,pl,3}
   ‘Ana needed textbooks’

(ii) Ani je bilo hladno.
    Ana_{DAT,f,sg,3} be_{ppl,sg,n} cold_{NOM,n,sg}
    ‘Ana was cold’

In constructions like (i), the verb agrees with the nominative argument in person, number and gender, while in constructions like (ii) with no nominative argument the verb is always neuter singular. Only nominative noun phrases induce verbal agreement.
voljeti ‘to love’ requires an accusative object, bojati se ‘to be afraid’ requires a genitive object, zadiviti ‘to impress’ requires a dative and an instrumental object. The prepositions are equally picky, with, for example, za ‘for’ requiring an accusative object, iz ‘from’ requiring a genitive object, prema ‘towards’ requiring a dative object, and sa ‘with’ requiring an instrumental object. Since differently case marked noun phrases are not interchangeable, case is syntactically significant.

The person, number and gender marking is relevant for subjects. Finite verbs\(^{13}\) agree with their subjects in number and person, but not gender. However, Serbo-Croatian has periphrastic tenses composed from a finite auxiliary and a non-finite participle. In these constructions, the auxiliary agrees with the subject in person and number, while the participle agrees with the subject in number and gender. The examples below illustrates this agreement pattern.

(44) finite verbs are gender-neutral:

Further, apart from inducing verbal agreement, nominative noun phrases are also special because only they control the interpretation of reflexives, including the pronominal reflexive sebe ‘self’ (see Chapter 4), and the subject-oriented possessive svoj. A dative experiencer cannot be coreferential with a reflexive.

So, we will consider subjects nominative noun phrases which induce verbal agreement and control the interpretation of reflexives. We will consider constructions like (ii) subjectless, while in constructions like (i) we will call the nominative argument that the verb agrees with the subject. All subjects are nominative; however, not all nominative noun phrases are subjects, cf. predicative structures, Chapter 4.

\(^{13}\)Serbo-Croatian has three non-periphrastic tenses: present, aorist and imperfect. However, of these, only the present tense is in wide-spread use, aorist and imperfect having been replaced by a periphrastic past tense. We will therefore largely exclude aorist and imperfect from our analysis, with one notable exception which we will return to in Chapter 5, namely, the aorist of biti ‘to be’. The latter are enclitic verbal forms used to form conditional mood. Keeping this one exception in mind, finite verbs will in general be used to refer to present tense verbs, since both the future tense and the most widely used past tense are periphrastic and use present tense auxiliaries.
a. Oni spavaju.  
   they_{NOM,m,pl,3rd} sleep_{pl,3rd}  
   ‘They (masculine) sleep’

b. One spavaju.  
   they_{NOM,f,pl,3rd} sleep_{pl,3rd}  
   ‘They (feminine) sleep’

c. Ona spavaju.  
   they_{NOM,n,pl,3rd} sleep_{pl,3rd}  
   ‘They (neuter) sleep’

(45) finite verbs agree with the subject in number and person:

   Marko_{NOM,m,sg,3rd} sleep_{pl,3rd}  
   [intended: ‘Marko sleeps’]

b. * Mi spavaju.  
   we_{NOM,m/f,pl,1st} sleep_{pl,3rd}  
   [intended: ‘We sleep’]

(46) participles in past tense agree with subjects in gender and number:

a. Marko je spavao.  
   Marko_{NOM,m,sg,3rd} is_{sg,3rd} sleep-PPL-m,sg  
   ‘Marko slept’

b. * Ana je spavao.  
   Ana_{NOM,f,sg,3rd} is_{sg,3rd} sleep-PPL-m,sg  
   [intended: ‘Ana slept’]

c. * Oni je spavao.  
   they_{NOM,m,pl,3rd} is_{sg,3rd} sleep-PPL-m,sg  
   [intended: ‘They (masculine) slept’]

At first glance, in simple sentences only case of object noun phrases seems to matter. In other words, If a verb or a preposition need an accusative object, any accusative noun phrase will do, regardless or gender, number or person of that noun
phrase. However, there are instances where the gender and number, in addition to case, matter even for non-nominative noun phrases.

First, it is possible to relativize on noun phrases in any grammatical case. In such relative clauses, the relative pronoun must agree in gender and number with the modified noun phrase which requires the number and gender information to be recorded on the noun phrase. Below is an example of such a relative clause, where the relative pronoun *koju* agrees with the noun phrase *Ana* in gender and number:

(47) Ana, koju znam sa fakulteta, je moja najbolja prijateljica.

Ana, who I know from college, is my best friend

Second, in object-control constructions, the person of the object matters as well. Consider the following example:

(48) a. Marko nagovara Anu da vozi. ‘Marko convinces Ana to drive’

b. Marko nagovara nas da vozimo. ‘Marko convinces us to drive’

With *nagovarati* ‘convince’ and other object-control verbs, the embedded verb has to agree in number and person with the matrix object. When an object noun phrase controls an adjective, the adjective has to agree with the object in gender and number:
Here, the accusative object *Anu* and the instrumental adjective *pametnom* must agree in gender and number. In order for this sentence to be composed, the accusative noun phrase *Anu* has to carry information about its number and gender, so that agreement with the adjective can be induced. Because of examples like these, we conclude that number, person and gender are syntactically relevant even for non-nominative noun phrases.

**Word Order**

It is uncontroversial that in Serbo-Croatian a verb, its subject and any objects can freely order with respect to one another. A sentence consisting of an intransitive verb and its subject can be pronounced two different ways:

(50) a. Vesna spava.
   Vesna*NOM* sleeps/is sleeping
   ‘Vesna sleeps/is sleeping’

b. Spava Vesna.

14 The utterances of the same sentence with different word orders are semantically (i.e. truth-conditionally) identical. It’s not clear that one could even argue for significant pragmatic differences, at least in these simple cases. For example, each of the utterances above could be offered as an answer to both *Who is sleeping?* and *What’s Vesna doing?* See Progovac (2005), Godjevac (2000) for a claim that different word orders are associated with pragmatic differences. We also direct the reader to Chapter 1 for a more extended discussion of our decision to exclude pragmatics altogether.

15 The imperfective present tense verb *spava* could be interpreted as denoting a habitual or an ongoing activity, hence the dual gloss. Present tense verbs in main clauses will in general be given in the imperfective form in the examples, and we will henceforth suppress the dual gloss of such verbs’ meaning.
A sentence consisting of a transitive verb, and its subject and object, can be pronounced six different ways (3!):

(51) a. Vesna voli Marka.
   \textit{Vesna}_{NOM} \textit{loves Marko}_{ACC}
   \textit{‘Vesna loves Marko’}
b. Vesna Marka voli.
c. Voli Vesna Marka.
d. Voli Marka Vesna.
e. Marka Vesna voli.
f. Marka voli Vesna.

Finally, a sentence consisting of a ditransitive verb, and its subject and objects can be pronounced twenty-four different ways (4!):

(52) a. Vesna predstavlja Marka Ana.
   \textit{Vesna}_{NOM} \textit{introduces Marko}_{ACC} \textit{Ana}_{DAT}
   \textit{‘Vesna introduces Marko to Ana’}
b. Vesna Marka predstavlja Ana.
c. Vesna Marka Ana predstavlja.
d. Predstavlja Marka Ana Vesna.
e. Marka Ana predstavlja Vesna.
f. Ana Vesna Marka predstavlja.
g. etc.

Therefore, the grammar must in general allow for free ordering of verbs and their noun phrase arguments.

We also note that quantificational lexical noun phrases can freely order with respect to other clausal constituents. Further, if a sentence contains two quantificational pronouns, regardless of the word order, the sentence will be ambiguous. Consider the following example:
Regardless of which of the six possible ways it’s pronounced, the sentence above remains ambiguous between the two readings, namely ‘there is some person who loves everybody’ and ‘for every person there is somebody who loves them’.

3.2.2 Phrasal Noun Phrases

Determiner-less Noun Phrases

While Serbo-Croatian has quantificational, demonstrative and possessive determiners, none of them are obligatory. Singular count nouns, bare or with modifiers, can occur as arguments of verbs or prepositions. For example:

\[(54)\]  
  \[\text{Djevojka} \text{NOM,f,sg} \text{spava.} \]
  \[\text{girl} \text{NOM,f,sg} \text{sleeps} \text{sg,3} \]
  \[‘A/The girl sleeps’\]

The meaning of the bare noun *djevojka* ‘girl’ is ambiguous between an indefinite and a definite interpretation. The same is true in cases where modifiers occur with the noun. When nouns (with or without modifiers) occur as subjects, they always induce 3rd person agreement with the verb.
Attributive Adjectives

Nouns and adjectives in Serbo-Croatian are marked for case, number and gender and they have to agree in terms of these features. The example below shows the general agreement pattern.

(55)  
  a. Dobri studenti uče.
      good\textit{NOM,m,pl} studenti\textit{NOM,m,pl} study
      ‘Good students study’
  b. * Dobar studenti uče.
      good\textit{NOM,m,sg} studenti\textit{NOM,m,pl} study
      [intended: ‘Good students study’]
  c. * Dobrih studenti uče.
      good\textit{GEN,m,pl} studenti\textit{NOM,m,pl} study
      [intended: ‘Good students study’]
  d. * Dobre studenti uče.
      good\textit{NOM,f,sg} studenti\textit{NOM,m,pl} study
      [intended: ‘Good students study’]

There does not seem to be a consensus as to the empirical facts concerning the placement of attributive adjectives in Serbo-Croatian. Consider the sentence below:

(56)  
  a. Ana kupuje novi auto.
      Ana\textit{NOM,f,sg}\,3 buys new\textit{ACC,m,sg} car\textit{ACC,m,sg}
      ‘Ana buys/is buying a new car’

In my judgment, the sentence above can be pronounced 24 different ways, i.e. all permutations of the verb, the subject NP, the noun and the attributive adjective are possible, with no change in meaning.
On the other hand, Leko (1999), officially discussing Bosnian, claims that attributive adjectives must occur immediately to the left of the noun they modify, while Zlatić (1997), officially discussing Serbian noun phrases, allows for the adjective and the noun to permute, but not for them to in general appear discontinuously in a clause. All three sets of judgments however include free permutation of a verb and its noun phrase arguments, which we discussed earlier.

**Postnominal Modifiers**

Postnominal modifiers in Serbo-Croatian include some predicative phrases\(^\text{16}\) such as certain prepositional and adjectival phrases, possessive genitive noun phrases which are not predicative but used in circumstances when a possessive determiner cannot be formed, for morphological reasons, and relative clauses. The examples below show a variety of postnominal modifiers in Serbo-Croatian. The modifiers have been enclosed in square brackets.

(57) predicative phrases:

   girl\([\text{NOM},f,sg]\) from Belgrade\([\text{GEN},m,sg]\) loves Marko\([\text{ACC},m,sg]\)
   ‘The girl from Belgrade loves Marko’

b. Ana daje poklone [vrijedne sto dollars\([\text{GEN},m,pl]\)]
   Ana\([\text{NOM},f,sg]\) gives presents\([\text{ACC},m,pl]\) worth\([\text{ACC},m,pl]\) hundred dolara].
   ‘Ana gives presents worth $100’

(58) possessive phrases:

\(^\text{16}\)By ‘predicative phrases’ we mean phrases which can be complements of the verb *biti* ‘be’. Not all predicative phrases can occur as postnominal modifiers in Serbo-Croatian.
All postnominal modifiers must occur immediately to the right of the noun they modify, which is why we refer to them as *postnominal*. For example, Serbo-Croatian in general doesn’t allow extraposition of relative clauses (Browne (1974))\(^\text{17}\).

In the following example, post nominal modifiers are enclosed in square brackets and the modified noun is underlined. They are all ungrammatical:

   d. *Djevojka dolazi [koju Ana zna sa fakulteta]

\(^{17}\)If a relative clause modifies a pronoun, then extraposition if possible, but never if it modifies a noun. This phenomenon seems syntactic in nature and doesn’t seem to correlate with whether the relative clause is restrictive or non-restrictive.
Further, all postnominal modifiers must remain contiguous, so that no main clause material, including the noun they are modifying, may break up a multi-word postnominal modifier.

Finally, we note that there is no agreement between the modified noun and a prepositional phrase or a possessive genitive noun phrase; they can modify nouns of any case, number and gender. Relative pronouns agree with the nouns modified by relative clauses in number and gender, but not case. It is possible to speculate that precisely because postnominal modifiers do not exhibit total agreement with the noun they modify, they must remain adjacent to the noun, so that the correct interpretation can be arrived at. Attributive adjectives, on the other hand, agree with the nouns they modify in number, gender and case. It is possible that because of this total agreement with the noun, they show more flexibility in terms of word order than postnominal modifiers.

Even though in my judgment an attributive adjective and a noun that it modifies can occur discontinuously in a sentence, if there is also a postnominal modifier, the attributive adjective can no longer detach. However, so long as the noun and the adjective remain contiguous and immediately to the left of the postnominal modifier, they can still permute. The examples below illustrates this pattern:

(61) a. Marko ima dobrog druga iz Beograda.
    Marko NOM, m, sg has good ACC, m, sg friend ACC, m, sg from Belgrade GEN, m, sg
    ‘Marko has a good friend from Belgrade’

b. Marko ima druga dobrog iz Beograda.

Note that both Leko (1999)’s and Zlatić (1997)’s pattern of judgments is far less complex. If they require the adjective and the noun to be contiguous anyways, presumably they must be contiguous in the presence of postnominal modifiers as well. We will sketch an analysis of all these judgments in the next section.

### Quantificational Determiners

In our judgment, just like attributive adjectives, quantificational determiners can detach from their argument noun. Unlike attributive adjectives, they can do so even when the noun has postnominal modifiers. Consider the examples below:

    'Some girl who Ana knows from college is arriving'

b. Neka dolazi djevojka koju Ana zna.
c. Djevojka koju Ana zna dolazi neka.
d. Djevojka koju Ana zna neka dolazi.
e. etc.

(63) a. Svi ljudi iz Beograda dolaze.
    'All the people from Belgrade are arriving'

b. Svi dolaze ljudi iz Beograda.
c. Ljudi iz Beograda dolaze svi.
d. Ljudi iz Beograda svi dolaze
e. etc.

In the next section we will analyze this complicated pattern of judgments, where attributive adjectives, postnominal modifiers and quantificational determiners are
all associated with different word order possibilities. Presumably, Zlatić (1997) and Leko (1999) require quantificational determiners to remain contiguous with the rest of the noun phrase, just like attributive adjectives, which we will also show how to represent in the grammar. Further, we will show how to represent Zlatić (1997)’s requirement that quantificational determiners always be left-most in the noun phrase, preceding any attributive adjectives.

3.2.3 Adverbial Modifiers

Non-Prepositional Adverbial Modifiers

Adverbial phrases can freely order with respect to the verb and its noun-phrase arguments:

(64) a. Vesna vozi brzo.
   Vesna\textsubscript{NOM} drives fast
   ‘Vesna drives fast’
   b. Vesna brzo vozi.
   c. Vozi Vesna brzo.
   d. Vozi brzo Vesna.
   e. Brzo Vesna vozi.
   f. Brzo vozi Vesna.

If the adverbial expression consists of a degree and an adverb, the degree and the adverb must remain contiguous and the degree must precede the adverb. The sequence of the degree and the adverb can, however, freely order with respect to the other clausal constituents. The following examples illustrate the relevant pattern.
(65) Vesna vozi veoma brzo.
    Vesna_{NOM} drives very fast
    ‘Vesna drives very fast’

(66) the degree and the adverb must remain contiguous:
    a. * Veoma Vesna brzo vozi.
    b. * Vozi veoma Vesna brzo.
    c. * Brzo Vesna veoma vozi.
    d. * Vozi brzo Vesna veoma.
    e. etc.

(67) the degree must precede the adverb:
    a. * Vesna vozi brzo veoma.
    b. * Vozi brzo veoma Vesna
    c. etc.

(68) the degree+adverb sequence can freely order with respect to other constituents:
    a. Vesna veoma brzo vozi.
    b. Veoma brzo Vesna vozi.
    c. Veoma brzo vozi Vesna.
    d. Vozi veoma brzo Vesna.
    e. Vozi Vesna veoma brzo.

Given this data, the grammar must have a way of both (i) allowing free reordering of constituents, and (ii) ensuring that certain multi-word phrases remain contiguous and internally ordered, while freely reordering as a unit with respect to other phrases.

**Prepositional Adverbial Modifiers**

Prepositions in Serbo-Croatian are clitics, which means that they are not phonological words (i.e., they are not associated with a lexical pitch accent; see Godjevac
(1999, 2000); also see Zec and Inkelas (1990) for a slightly different formulation to the same effect). They are proclitics, which means that they must attach to a phonological word to their right.

A more conservative set of judgments about prepositional phrases in Serbo-Croatian would be that the preposition must occur immediately to the left of its argument noun phrase, procliticizing onto the first phonological word therein, and the entire prepositional phrase must remain contiguous. The permissible order within the noun phrase that is an object of a preposition is determined by one’s judgments about the order within noun phrases in general, for example whether one believes that the adjective or the quantificational determiner must precede the noun or not. The order within the noun phrase in turn determines which phonological word the preposition procliticizes onto.

However, in my permissive judgment, whether the prepositional phrase must remain contiguous or not, depends on whether its object noun phrase must remain contiguous or not. For example, if the preposition’s noun phrase object consists of an adjective and a noun, the preposition can procliticize onto either the noun or the adjective. The two parts of the prepositional phrase can then occur discontinuously in an utterance, provided that the part that includes the preposition precedes the part that does not. The example below illustrates this pattern.

\[(69) \quad \text{a. } \text{U velikom gradu Ana živi.} \]  
\[\quad \text{in big}_{\text{DAT}} \text{ city}_{\text{DAT}} \text{ Ana}_{\text{NOM}} \text{ lives} \]  
\[\quad \text{‘Ana lives in a big city’} \]  
\[\text{b. } \text{U gradu velikom Ana živi.} \]
c. U velikom Ana živi gradu.
d. U gradu Ana živi velikom.
e. U velikom Ana gradu živi.
f. U gradu Ana velikom živi.
g. Ana u gradu živi velikom.
h. Ana u velikom živi gradu.
   etc., but:
i. * Velikom Ana živi u gradu.
j. * Gradu Ana živi u velikom.
   etc.

The same pattern is evident in my judgment if the noun phrase contains a quantificational determiner. However, since splitting a noun and its postnominal modifier is in general not possible, if such a noun phrase is the object of the preposition then the whole prepositional phrase must remain contiguous. We will sketch an analysis of both the less permissive (and easier to analyze) set of judgments whereby the entire prepositional phrase must remain contiguous, and the more permissive set of judgments whereby under certain conditions, depending on the structure of the object noun phrase, the prepositional phrase can be made discontinuous.

3.3 Analysis

3.3.1 Lexical Noun Phrases

Representation of Lexical Noun Phrases in the Grammar

To account for the different agreement properties of noun phrases, we assign them various tectogrammatical types which reflect their inflectional properties.
We start with the basic tectogrammatical types $\text{Cse}$, $\text{Gdr}$, $\text{Num}$, and $\text{Prs}$. Terms of these types represent different inflectional features. Below we list the set of terms for each of these inflectional feature types.

(70) a. $\text{Cse} = \{\text{nom, gen, dat, acc, inst}\}$
    b. $\text{Gdr} = \{m, f, n\}$
    c. $\text{Num} = \{sg, pl\}$
    d. $\text{Prs} = \{1, 2, 3\}$

The tectogrammatical type of noun phrases is then dependent on terms of $\text{Cse}$, $\text{Gdr}$, $\text{Num}$, and $\text{Prs}$, so that the grammar can distinguish between noun phrases based on their case, gender, number and person respectively. Below are a few examples of noun phrase types in the grammar:

(71) $\text{NP}_{\text{gen,f,sg,3}}$ genitive feminine singular 3rd person noun phrases
    $\text{NP}_{\text{dat,m,sg,3}}$ dative masculine singular 3rd person noun phrases
    $\text{NP}_{\text{nom,n,sg,3}}$ nominative neuter singular 3rd person noun phrases

Semantically, lexical noun phrases such as the ones we considered in this chapter denote individuals, so they are represented in the grammar as constants of type $\text{e}$. Phenogrammatically, we analyze them as denoting length one strings of languages, so they are represented as terms of type $\text{z}$. Below are full lexical entries for three different case forms of two names:

(72) a. $\vdash \text{MARKO}_z : \text{z}; \text{NP}_{\text{nom,m,sg,3}}; \text{marko} : \text{e}$
    b. $\vdash \text{MARKA}_z : \text{z}; \text{NP}_{\text{acc,m,sg,3}}; \text{marko} : \text{e}$
    c. $\vdash \text{MARKU}_z : \text{z}; \text{NP}_{\text{dat,m,sg,3}}; \text{marko} : \text{e}$
    d. $\vdash \text{VESNA}_z : \text{z}; \text{NP}_{\text{nom,f,sg,3}}; \text{vesna} : \text{e}$
    e. $\vdash \text{VESNU}_z : \text{z}; \text{NP}_{\text{acc,f,sg,3}}; \text{vesna} : \text{e}$
    f. $\vdash \text{VESNI}_z : \text{z}; \text{NP}_{\text{dat,f,sg,3}}; \text{vesna} : \text{e}$
We represent lexical noun phrases phenogrammatically as length one $S$-strings because they do not have a more intricate, lower level phenogrammatical structure, and can freely permute with verbs and their noun phrase arguments. That is, no string, low level attachment needs to be effected (see Chapter 2), and no special restrictions need to be placed on the word order of lexical noun phrases. Since they are of type $z$, they will be subject to higher level functions such as $S$-string concatenations, $\text{PER}$ and $\circ$.

**Combining Lexical Noun Phrases with Verbs**

We start with an intransitive verb, the present tense forms of $\text{spavati}$ ‘to sleep’. Recall that finite verbs require a nominative subject which agrees with them in terms of person and number, but it doesn’t care about the gender of its subject. Further, the verb has to combine itself with the subject in a way which will allow them to freely reorder with respect to one another. We associate $\text{spava}$ ‘sleeps’ with the following lexical entry:

$$(73) \quad \vdash \lambda v. \text{PER}(v \circ \text{SPAVA}) : z \rightarrow z; \prod_{x: \text{Gdr}}[\text{NP}_{\text{nom}, x, \text{sg}, 3} \rightarrow \text{S}_{m, 6}]; \text{sleep} : e \rightarrow p$$

Semantically, the verb is a function from of individuals to propositions. Phenogrammatically, the verb is looking for an argument of type $z$. Once it combines with such an argument, it creates a string of languages, and then constructs a set of all permutations of that string of languages. This allows the verb and the subject to freely order with respect to one another.$^{18}$

$^{18}$ $\text{PER}$ has to be built into the lexical entries of individual verbs because free reordering of verbs and their arguments is not always possible. For example, embedded clauses have to occur on
In very general and informal terms, the tectogrammatical type of this verb reflects the fact that, given an appropriate subject noun phrase (represented by the noun phrase parameters), it can construct a declarative sentence. Syntactically, *spava* ‘sleeps’ doesn’t care about the gender of its subject which is why it’s associated with a dependent product type. The tectogrammatical result type of *spava* is $S_{m,6}$.

Similar to noun phrase types, we also have a family of clause types, indexed by a term of type $K=\{e,m,q\}$ and $Nat=\{0,1,2,3,\ldots\}$. The first parameter, of type $K$, refers to the kind of clause in question: e(mbedded), m(ain) or q(uestion). The second parameter is a natural number which will help us enforce the order in the enclitic cluster.\(^{19}\) We will return to $K$ and $Nat$ and their role in the grammar in later chapters. For now, suffice it to say that finite non-clitic verbs build sentences associated with the tectogrammatical type $S_{m,6}$.

Now we can already construct simple sentences. Below is a derivation of one such sentence, omitting phenogrammatical and semantic types.

---

\(^{19}\)See Morrill and Gavarró (1992) for the original use of the natural number parameter to enforce the order of clitics in a clause. Also note that in practice, we only need numbers 0-6, and tectogrammatical types of signs will only make reference to those numbers.
(74)

\[ \vdash \lambda z. \text{PER}(v \circ \text{SPAVA}_z); \text{NP}_{\text{nom},m,\text{sg},3} \rightarrow S_{m,6}; \text{sleep} \quad \vdash \text{MARKO}_z; \text{NP}_{\text{nom},m,\text{sg},3}; \text{marko} \]

\[ \vdash \text{PER}(\text{MARKO}_z \circ \text{SPAVA}_z); S_{m,6}; (\text{sleep marko}) \]  

[\rightarrow \text{E}]

Since the tectogrammatical type of the object noun phrase and the argument type of the verb are the same, we can use the \([\rightarrow \text{E}]\) rule to combine them. \([\rightarrow \text{E}]\) is accompanied by function application in the phenogrammatical and semantic term calculi. The resulting sign, the conclusion of the proof above, is the representation of the sentence Marko spava ‘Marko sleeps’ in the grammar. Tectogrammatically, this sign is a declarative sentence, and semantically it denotes the proposition that Marko sleeps.  

The phenogrammatical part of the conclusion is the term \(\vdash \text{PER}(\text{MARKO}_z \circ \text{SPAVA}_z) : Z\) which denotes a set that contains exactly two strings of \(p\)-languages, \(\text{MARKO}_z \circ \text{SPAVA}_z\) and \(\text{SPAVA}_z \circ \text{MARKO}_z\). The whole sign can undergo the \([\text{km}]\) rule, which compacts the phenogrammatical term to a set of \(p\)-strings containing exactly two \(p\)-strings, namely \(\text{marko} \cdot \text{spava}\) and \(\text{spava} \cdot \text{marko}\), which correspond to the two ways this sentence could be pronounced, \(\text{Marko spava}\) and \(\text{Spava Marko}\) respectively.

Note that the grammar doesn’t treat \(\text{Marko spava}\) and \(\text{Spava Marko}\) as distinct expressions at all, but as two ways of pronouncing one and the same declarative
sentence. The grammar generates a single sign whose phenoterm, because it
denotes a set of strings of languages, specifies all the different ways that a single
sentence could be pronounced.

Lexical entries of transitive verbs are specified similarly. Below is the lexical
entry for voli ‘loves’.\(^{20}\)

\[
\lambda_{v,n,w} \text{PER}(w \circ \text{VOLI}_z \circ v) : z \rightarrow z \rightarrow z;
\prod_{x,\text{Gdr},y,\text{Num},z,\text{Prs}} [\text{NP}_{\text{acc},n,x,y} \circ \text{NP}_{\text{nom},z,\text{sg},y} \circ \text{S}_m] ;
\text{love} : e \rightarrow e \rightarrow p
\]

Phenogrammatically, this verb takes two arguments of type \(z\) (the object and the
subject noun phrase), and outputs the set of all permutations of the string obtained
by concatenating the subject, the verb and the object. Tectogrammatically, voli
needs an accusative noun phrase of any gender, number and person as its object,
and a 3rd person singular nominative noun phrase of any gender as its subject, to
construct a declarative main clause.

From the lexical entries for the verb, the subject and the object, in two steps of
\([\text{E}\circ\text{E}]\) we can obtain the following sign, which represents the sentence Marko voli Vesnu ‘Marko loves Vesna’ in the grammar:

\[
\text{PER}(\text{MARKO}_z \circ \text{VOLI}_z \circ \text{VESNU}_z) : z ; \text{S}_m ; (\text{love marko vesna}) : p
\]

\(^{20}\)Note that the \(S\)-string that the transitive verb builds and then permutes consists of the sub-
ject, followed by the verb, followed by the object. In practice, it makes no difference how we
order these arguments since they all get permuted anyways, and there is no notion of canonical
word order in our grammar (see Chapter 1). However, the reader can take our habit of writing
phenogrammatical terms for verbs as permutations of the subject-verb-object sequence as a nod to
the standard claim that Serbo-Croatian is underlyingly an SVO language; see \textit{inter alia} Godjevac
(2000); Progovac (2005).
Tectogrammatically, this sign is a declarative main clause. Semantically, it expresses the proposition that Marko loves Vesna. Phenogrammatically, it denotes a set of exactly six strings of languages. This sign can undergo the [km] rule which compacts the phenogrammatical term into a set of six $p$-strings corresponding to the six different ways of pronouncing this sentence. Therefore, the grammar predicts that all of the following are possible pronunciations: *Marko voli Vesnu, Marko Vesnu voli, Voli Marko Vesnu, Voli Vesnu Marko, Vesnu voli Marko* and *Vesnu Marko voli.*

Lexical entries for ditransitive verbs such as *predstavlja* ‘introduces’ are given in a similar fashion. Below is the lexical entry for *predstavlja* ‘introduces’, and the sign which represents the sentence *Marko predstavlja Vesnu Ana* ‘Marko introduces Vesna to Ana’ in the grammar.

(77) $\vdash \lambda_{v\chi w}.\text{PER}(w \circ \text{PREDSTAVLJA}_z \circ v \circ x) : z \rightarrow z \rightarrow z \rightarrow Z;
\text{NP}_{\text{acc},f,sg,3} \rightarrow \text{NP}_{\text{dat},f,sg,3} \rightarrow \text{NP}_{\text{nom},m,sg,3} \rightarrow S_{m,6};$
introduce : $e \rightarrow e \rightarrow e \rightarrow p$

(78) $\vdash \text{PER}(\text{MARKO}_z \circ \text{PREDSTAVLJA}_z \circ \text{VESNU}_z \circ \text{ANI}_z) : Z; S_{m,6};$
(introduce vesna ana marko) : $p$

Phenogrammatically, the sentence *Marko predstavlja Vesnu Ana* ‘Marko introduces Vesna to Ana’ is represented as the term $\vdash \text{PER}(\text{MARKO}_z \circ \text{PREDSTAVLJA}_z \circ \text{VESNU}_z \circ \text{ANI}_z) : Z.$ After undergoing [km], the phenogrammatical term denotes a set of exactly twenty-four strings which correspond to the twenty-four different ways of pronouncing this sentence.
Quantificational Lexical Noun Phrases

Recall that quantificational lexical noun phrases can freely order with respect to other clausal constituents, just like any other noun phrases, so the sentence below can be pronounced six different ways, without a change in meaning.

(79) Ana \(\text{voli }\) svakoga.  
\(\text{Ana}^{\text{NOM},f,\text{sg,3 }}\text{love}_{\text{pl,3 }}\text{everybody}^{\text{ACC,m,sg,3}}\)  
‘Ana loves everybody’

Our general strategy with respect to quantification will be to use quantifier lowering along the lines of Oehrle (1994), whereby quantificational noun phrases combine with ‘gappy’ sentences, i.e. sentences missing a noun phrase, and scope over such constituents, which are semantically properties of individuals.\(^{21}\) Phenogrammatically, quantificational noun phrases lower themselves into and take the place of the ‘gap’, hence the name quantifier lowering. Below are the lexical entries required to construct the sentence above.

(80) a. \(\vdash \text{ANA}_z : z; \text{NP}_{\text{nom,f,sg,3 ; ana}} : e\)

b. \(\vdash \lambda w. \text{PER}(w \circ \text{VOLI}_z \circ v) : z \rightarrow z \rightarrow z; \text{NP}_{\text{acc,m,sg,3} \rightarrow \text{NP}_{\text{nom,f,sg,3}} \rightarrow }\)  
\(S_{m,6}; \text{love} : e \rightarrow e \rightarrow p\)

c. \(\vdash \lambda f. (F \text{SVAKOGA}_z) : (z \rightarrow z) \rightarrow z; (\text{NP}_{\text{acc,m,sg,3} \rightarrow S_{m,6}} \rightarrow S_{m,6});\)
\(\text{everyone} : (e \rightarrow p) \rightarrow p\)

\(^{21}\)We direct the reader to Chapter 7 for a more extended, though by no means exhaustive, discussion of quantification in Serbo-Croatian.
In the lexical entry for svakoga, the semantic term $\vdash \text{everyone} \rightarrow \text{everyone} : (\exists \rightarrow \text{p}) \rightarrow \text{p}$ is an abbreviation for $\vdash (\text{every person}) : (\exists \rightarrow \text{p}) \rightarrow \text{p}$. The hyperintensional generalized quantifier $\vdash \text{every} : (\exists \rightarrow \text{p}) \rightarrow (\exists \rightarrow \text{p}) \rightarrow \text{p}$ is related to its extensional counterpart via the following meaning postulates:

\begin{align*}
\forall_{PQ}[ (\forall PQ)@w = \forall_x[ (P \forall x)@w \rightarrow (Q \forall x)@w ]
\end{align*}

Below we show the step-by-step derivation of the sentence above. We omit phenogrammatical and semantic types, as well as the non-case NP parameters for type-setting reasons.

\begin{align*}
(82) \quad \vdash \lambda_{\forall x}. \text{PER}(w \circ \text{VOLI}_z \circ x); \text{NP}_{\text{acc}} \rightarrow \text{NP}_{\text{nom}} \rightarrow \text{S}_{m,6}; \text{love} & \quad \vdash \lambda_{\forall x}. \text{PER}(w \circ \text{VOLI}_z \circ x); \text{NP}_{\text{acc}} \rightarrow \text{NP}_{\text{nom}} \rightarrow \text{S}_{m,6}; (\text{love } x)
\end{align*}

The first step is to introduce a hypothesis or a trace via [Ax]. The trace is of the same type as the verb’s first argument—tectogrammatically, an accusative noun phrase; phenogrammatically, a string of languages; and, semantically, an individual. Then the verb and the hypothesis combine via [$\rightarrow \text{E}$], but the hypothesis is still kept track of in the context, to the left of the turnstile. Now we can proceed and combine the verb phrase with its subject:

\begin{align*}
(83) \quad \vdash \lambda_{\forall x}. \text{PER}(w \circ \text{VOLI}_z \circ x); \text{NP}_{\text{acc}} \rightarrow \text{NP}_{\text{nom}} \rightarrow \text{S}_{m,6}; (\text{love } x) & \quad \vdash \lambda_{x}. \text{PER}(\text{ANA}_z \circ \text{VOLI}_z \circ x); \text{NP}_{\text{acc}} \rightarrow \text{S}_{m,6}; (\text{love } x \text{ ana})
\end{align*}

After combining the verb phrase with the subject, we withdrew the hypothesis,
i.e. bound the trace via $[\neg \ominus I]$. This means that the hypothesis no longer appears in the context, and all occurrences of variables that originated with the hypothesis have been bound in the succedent. Now the quantificational noun phrase can combine with this ‘gappy’ sentence.

\[(84)\]
\[\vdash \lambda x. (F \text{SVAKOGA}_z); \text{NP}_{\text{acc}} \rightarrow S_{m,6}; \text{everyone} \quad \vdash \lambda x. \text{PER}(\text{ANA}_z \circ \text{VOLI}_z \circ x); \text{NP}_{\text{acc}} \rightarrow S_{m,6}; \lambda x(\text{love } x \text{ ana}) \quad [\neg \ominus E] \]

Once the quantificational noun phrase combines with the ‘gappy’ sentence, the term $\vdash \text{SVAKOGA}_z : z$ takes the place of variable $\vdash x : z$. That sign can undergo [km]. The resulting phonogrammatical term denotes a set of six $p$-strings, corresponding to the six different pronunciations of this sentence. Semantically, the sentence is analyzed as expressing the expected universally quantified proposition, and tectogrammatically it is an ordinary declarative sentence.

3.3.2 Phrasal Noun Phrases

Representing Lexical Nouns in the Grammar

Just as with noun phrase types, we have a family of noun types indexed by gender, case and number parameters. No person parameters are necessary since nouns do not participate in verbal agreement for which the person parameter is required; only noun phrases agree with verbs. Below we give lexical entries for a few noun phrases, to illustrate the tectogrammatical noun types.

\[(85)\]  

a. $\vdash \text{DJEVOJKA}_z : z; \text{N}_{\text{acc,f,sg}}; \text{girl} : e \rightarrow p$
b. ⊢ DJEVOJKE_{z} : z; N_{gen,f,sg}; girl : e → p  
c. ⊢ DJEVOJKM_{z} : z; N_{inst,f,sg}; girl : e → p  
d. ⊢ STUDENT_{z} : z; N_{nom,m,sg}; student : e → p  
e. ⊢ STUDENTA_{z} : z; N_{gen,m,sg}; student : e → p  
f. ⊢ STUDENTOM_{z} : z; N_{inst,m,sg}; student : e → p

Phenogrammatically, we treat nouns as strings of languages (type $z$), while semantically we analyze them as expressing functions from individuals to propositions (type $e → p$).

**Quantifying Nouns in Absence of Determiners**

Recall that Serbo-Croatian doesn’t have obligatory determiners. Consider the example below.

(86) Djevojka spava.  
girl_{NOM,f,sg,3} sleeps_{sg,3}  
‘A/The girl sleeps’

Here, the bare noun *djevojka* ‘girl’ occurs as the subject of the 3rd person singular verb *spava* ‘sleeps’ and can mean either ‘a girl’ or ‘the girl’. Therefore, we have to have a general schema that will convert nouns into quantificational noun phrases.

We convert nouns into quantificational noun phrases in two steps. First, we state a rule schema that converts nouns of a given case, gender and number into noun phrases of the same case, gender and number. This will be a purely tectogrammatical schema, in the sense that the phenogrammatical and semantic terms and types will not be affected by it at all. Next, we state a rule schema
that converts signs which are tectogrammatically noun phrases, but semantically denote properties of individuals, into quantificational noun phrases.

The reason we are doing this in two steps has to do with enforcing ordering restrictions imposed on the attributive adjective in the presence of a postnominal modifier, as we will see later.

Below is a rule schema, [NC], that converts nouns into noun phrases. It allows any noun to be treated, tectogrammatically, as a 3rd person noun phrase of the same case, gender and number, since only pronouns can be of 2nd or 1st person. The phenogrammatical and the semantic portion of the sign remain unchanged.

(87)

\[
\begin{align*}
\vdash & \phi : z; N_{t':Cse,t''':Gdr,t'''':Num}; \sigma : e \rightarrow p \\
\vdash & \phi : z; NP_{t':Cse,t''':Gdr,t'''':Num,3}; \sigma : e \rightarrow p \quad [NC]
\end{align*}
\]

Below is the rule that turns nouns whose tectogrammatical type is noun phrase, but which have the semantic type \(e \rightarrow p\), to be converted into full blown quantificational noun phrases.

(88)

\[
\begin{align*}
\vdash & \phi : z; NP_{t':Cse,t''':Gdr,t'''':Num,3}; \sigma : e \rightarrow p \\
\vdash & \lambda_F. (F \phi) : (z \rightarrow z); (NP_{t':Cse,t''':Gdr,t'''':Num,3} \circ S_{m,6}) \circ S_{m,6}; \\
& (\exists \sigma) : (e \rightarrow p) \rightarrow p \quad [\text{Quant}]
\end{align*}
\]

The hyperintensional constant \(\vdash \exists \sigma : (e \rightarrow p) \rightarrow (e \rightarrow p) \rightarrow p\) is related to its extensional counterpart via the following meaning postulate:

(89) \(\forall_{PQw}[\exists x((Px)@w \land (Qx)@w)]\)
As mentioned before, determiner-less noun phrases in Serbo-Croatian can also have definite meaning, in addition to the indefinite meaning which [Quant] introduces. Deriving such definite interpretations of noun phrases requires positing another rule similar to [Quant]. We will not pursue that here, instead focusing on the indefinite interpretation only. Below we show how to apply [NC] and [Quant] to turn the noun djevojka ‘girl’ into a quantificational noun phrase. We omit phenogrammatical and semantic types.

\[
\frac{\text{DJEVOJKA}_{z}; \text{N}_{\text{nom},f,sg}; \text{girl}}{\text{DJEVOJKA}_{z}; \text{NP}_{\text{nom},f,sg,3}; \text{girl}} \quad [\text{NC}]
\]

\[
\frac{\lambda F. (\text{DJEVOJKA}_{z}) \circ (\text{NP}_{\text{nom},f,sg,3} \rightarrow S_{m,6}) \rightarrow S_{m,6}; (\text{exists girl})}{[\text{Quant}]}
\]

A sentence such as Djevojka spava ‘A girl sleeps’ would be represented in the grammar by the following sign:

\[
\frac{\text{PER}(\text{DJEVOJKA}_{z} \circ \text{SPAVA}_{z}) : \text{z}; S_{m,6}; (\text{exists girl}) \text{ sleep} : \text{p}}
\]

This sentence expresses an existentially quantified proposition that there exists a girl who sleeps, and it can be pronounced two ways, Djevojka spava or Spava djevojka.

**Attributive Adjectives**

In this section, we show how to represent in the framework different generalizations about nouns and attributive adjectives which were presented earlier in this chapter.
First, we analyze the most permissive empirical generalization concerning attributive adjectives, namely, that they can be detached from the noun they are modifying so that the noun and the adjective can individually freely reorder with respect to other clausal constituents.

Tectogrammatically, adjectives combine with nouns of a certain gender, case and number, and output a sign with the same tectogrammatical type. Semantically, their type is \((e \rightarrow p) \rightarrow (e \rightarrow p)\).

Phenogrammatically, we assume that attributive adjectives are of type \(z \rightarrow z\) and combine with nouns via ordinary string of languages concatenation. Therefore, the result of combining an attributive adjective with a noun is a length two string of languages. This will allow the adjective and the noun to individually freely reorder with respect to other constituents in the sentence.

Suppose we are trying to generate the sentence we mentioned earlier in the chapter, *Ana kupuje novi auto* ‘Ana is buying a new car’. Below are the required lexical entries.

(92) a. \(\vdash \text{ANA}_z : z; \text{NP}_{\text{nom},f,sg,3}; \text{ana} : e\)
   b. \(\vdash \lambda_{v,w} \text{PER}(v \circ \text{KUPUJE}_z \circ w) : z \rightarrow z \rightarrow z; \text{NP}_{\text{acc},m,sg,3} \circ \circ \text{NP}_{\text{nom},f,sg,3} \circ \circ \text{Sm,6}; \text{buy} : e \rightarrow e \rightarrow p\)
   c. \(\vdash \text{AUTO}_z : z; \text{N}_{\text{acc},m,sg}; \text{car} : e \rightarrow p\)
   d. \(\vdash \lambda_{x} \text{NOVI}_z \circ x : z \rightarrow z; \text{N}_{\text{acc},m,sg} \rightarrow \circ \text{N}_{\text{acc},m,sg}; \lambda_{\text{Py}}(Py) \) and (new \(y) : (e \rightarrow p) \rightarrow (e \rightarrow p)\)

We can combine the noun and the adjective (93a) and then apply [NC] and [Quant] to the resulting sign to obtain (93b).
The verb combines with an introduced hypothesis that is the object noun phrase trace, then combines with the subject Ana. Finally, when that hypothesis is withdrawn via \([-\circ I]\) (i.e., when the accusative trace is bound), the quantificational noun phrase \textit{novi auto} ‘a new car’ can combine with it. The result is given below:

\[
\Gamma \vdash \text{PER} (\text{ANA}_z \circ \text{KUPUJE}_z \circ \text{NOVI}_z \circ \text{AUTO}_z) : z; S_{m,6};
\exists (\lambda_y (\text{car } y) \land \text{(new } y)) ((\lambda_x (\text{buy } x \text{ ana})) : \text{p})
\]

The phenoterm of this sign denotes a set of twenty-four strings of languages, corresponding to the twenty-four possible pronunciations of this sentence, according to the most permissive empirical generalization which we are currently considering.

For the intermediate case, where the adjective and the noun are allowed to freely reorder with respect to one another, but must stay contiguous in the sentence, we give the following lexical entry for \textit{novi} ‘new’.

\[
\Gamma \vdash \lambda_x. \text{toZ}(k(\text{PER}(\text{NOVI}_z \circ x))) : z \rightarrow z; N_{\text{acc,m,sg}} \circ N_{\text{acc,m,sg}}; \\
\lambda_{p_y} (P y) \land \text{(new } y) : (e \rightarrow p) \rightarrow (e \rightarrow p)
\]

Semantically and tectogrammatically, everything is the same. The difference is entirely phenogrammatical. While the phenogrammatical type of the adjective is the same, instead of merely concatenating itself with the noun and creating a length two S-string, as in the most permissive grammar, here the adjective permutes itself with the noun via \textbf{PER}, then compacts the resulting set of strings of
languages into a set of strings via $k$. Finally, that set of strings is turned into a length one string of languages via $toZ$. This ensures that while the adjective and noun can reorder with respect to one another, they cannot be made discontinuous in a clause.

In this grammar, the sentence *Ana kupuje novi auto* ‘Ana is buying a new car’ is represented by the following sign:

\[
\text{PER(ANA}_z \circ \text{KUPUJE}_z \circ toZ(k(\text{PER(NOVI}_z \circ \text{AUTO}_z)))) : Z; S_{m,6};
\]

\[
\exists (\lambda_y.(\text{car } y) \text{ and } (\text{new } y))(\lambda_x.(\text{buy } x \text{ ana})) : p
\]

Looking at the phenoterm in more detail, $k$ compacts $\text{PER(NOVI}_z \circ \text{AUTO}_z)$ into a set of strings (type $S$) which contains exactly two strings, $\text{novi}_s \circ \text{auto}_s$ and $\text{auto}_s \circ \text{novi}_s$. That set of strings is then turned into a length one string of languages (type $z$). Therefore, while the length one string of languages $toZ(k(\text{PER(NOVI}_z \circ \text{AUTO}_z)))$ can freely permute with respect to $\text{ANA}_z$ and $\text{KUPUJE}_z$, the adjective and the noun cannot be made discontinuous.

For the most restrictive option, where the adjective and the noun must remain contiguous and the adjective must precede the noun, we give the following lexical entry for *novi* ‘new’.

\[
\text{\lambda}_x.toZ(L(\text{NOVI}_x \circ x)) : z \rightarrow z; N_{\text{acc,m,sg}} \rightarrow N_{\text{acc,m,sg}};
\]

\[
\lambda_{P_y}.(P y) \text{ and } (\text{new } y) : (e \rightarrow p) \rightarrow (e \rightarrow p)
\]

The adjective now concatenates itself with the noun, and then immediately ‘linguifies’ that string of languages into a set of strings via $L. toZ$ then turns that set
of strings into a length one string of languages. In this grammar, the sentence Ana kupuje novi auto ‘Ana is buying a new car’ is represented by the following sign:

\[(98) \vdash \text{PER}(\text{ANA}_z \circ \text{KUPUJE}_z \circ \text{toZ}(\text{L(NOVI}_z \circ \text{AUTO}_z))) : z; S_m,6; \exists \lambda y. (\text{car } y) \text{ and (new } y) (\lambda x. (\text{buy } x \text{ ana})) : p\]

In the phenoterm of this sign, \(\text{L(NOVI}_z \circ \text{AUTO}_z)\) denotes a set of strings with exactly one member, \(\text{novi}_s \circ \text{auto}_s\). So, the adjective must precede the noun, and while the length one string of languages \(\text{toZ}(\text{L(NOVI}_z \circ \text{AUTO}_z))\) can freely reorder with respect to other length one strings of languages in the sentence, it cannot be made discontinuous.

**Postnominal Modifiers**

Recall that it is uncontroversial that postnominal modifiers must remain contiguous and occur immediately to the right of the noun they are modifying. In this section, we analyze postnominal modifiers that are prepositional phrases, thereby simultaneously giving our theory of procliticization in Serbo-Croatian.

As in the previous section, we start with the most permissive generalization, building on the set of judgments according to which, under ordinary circumstances, an attributive adjective and a noun can detach and freely reorder with respect to other clausal constituents. In this permissive grammar, a sequence of a noun and an attributive adjective is phenogrammatically represented as a length two \(S\)-string, which allows them to freely permute.
In the presence of postnominal modifiers, however, a noun and an attributive adjective can still reorder with respect to one another, but must remain contiguous and occur immediately to the left of the postnominal modifier. The example below illustrates the judgments patterns that we are currently considering.

(99)  

MarkoNOM,m,sg,3 hassg,3 goodACC,m,sg friendACC,m,sg from Beograda.  
Marko has a good friend from Belgrade 

b. Marko ima druga dobrog iz Beograda.  
c. Marko dobrog druga iz Beograda ima.  
d. Marko druga dobrog iz Beograda ima.  
e. * Marko dobrog ima druga iz Beograda.  
g. * Marko ima iz Beograda dobrog druga.  
h. etc.

Note that last two examples, where the prepositional phrase and the adjective+noun sequence occur discontinuously are grammatical, but not on the relevant interpretation. That is, they are compatible with the interpretation on which Marko and his good friend met in Belgrade (the prepositional phrase being interpreted as adverbial). However, the target interpretation is the one on which the friend is from Belgrade (the prepositional phrase being interpreted as modifying the noun).

The first challenge we have to address is the fact that a postnominal modifier restricts the word order possibilities for the noun and the adjective. Informally, it’s as if the adjective has to ‘know’ whether there is also a postnominal modifier
or not, and then ‘behave’ accordingly. We account for this by demanding that any attributive adjectives combine with the noun *before* any postnominal modifiers do so. Then, when the postnominal modifier appears, it ‘freezes’ the entire phrase, preventing adjectives from escaping.

More formally, we accomplish this via tectogrammatical typing. Whereas attributive adjectives require an argument whose type is in the \( N \) family, we analyze postnominal modifiers as requiring an argument whose type is in the \( NP \) family. In other words, we require that the target of postnominal modification be nouns that have undergone the \([NC]\) rule, which changes their tectogrammatical type from a noun type to a corresponding noun phrase type, while leaving its phenogrammar and semantics unchanged, so that they still denote strings of languages, and their semantic type is still \( e \rightarrow p \).

For example, the type of *druga* ‘friend’ is \( N_{acc,m,sg} \), as is the type of *dobrog druga* ‘good friend’. That phrasal noun can then undergo the rule \([NC]\), which changes its tectogrammatical type to \( NP_{acc,m,sg,3} \). At this stage, a postnominal modifier can apply and output something of the same noun phrase type, so that *dobrog druga iz Beograda* ‘good friend from Belgrade’ would be of type \( NP_{acc,m,sg,3} \). Since attributive adjectives require arguments of a certain noun, not noun phrase, type, no more attributive adjectives could apply because of a type mismatch. In this way, the grammar forces all attributive adjectives to combine with the noun before any postnominal modifiers do.
It may seem unusual to allow mismatches between tectogrammatical and semantic types, and particularly, to allow expressions of a noun phrase type to denote properties of individuals semantically. Further, given our analysis, quantificational determiners, which we return to later in this chapter, will require an argument which is tectogrammatically a noun phrase, not a noun.

We could alternatively posit three tectogrammatical types, one for noun phrases, one for nouns with no postnominal modifiers, and one for nouns with postnominal modifiers (reminiscent of N'), the latter two denoting properties of individuals. This would allow us to maintain a functional mapping from tectogrammatical to semantic types. However, we see no reason to insist on such a functional mapping. Once we eschew that requirement, we don’t have to introduce new types because we can use the existing tectogrammatical and semantic types to effect the needed three-way distinction.

While it may seem counterintuitive to have noun phrases which denote properties of individuals, it doesn’t have any undesirable consequences for our theory from a practical standpoint. Should we discover an independent reason to maintain functional interfaces, the analysis given here would easily translate into such an architecture, requiring only the introduction of a third tectogrammatical type, in addition to the type of nouns and noun phrases.
Since postnominal modifiers do not agree with the nouns they modify,\textsuperscript{22} they will in general be associated with the following tectogrammatical type:

\begin{equation}
\prod_{x:Cse,y:Gdr,z:Num,3}[\text{NP}_{x:Cse,y:Gdr,z:Num,3} \rightarrow \text{NP}_{x:Cse,y:Gdr,z:Num,3}]
\end{equation}

Intuitively, this type just captures the fact that whatever the case, gender or number of its argument, the postnominal modifier will combine with it and then output something with the same agreement features.

Let’s consider the preposition iz ‘from’ in the context of postnominal modifiers. It has to combine with its argument noun phrase and procliticize onto some word in its argument. Then, it combines with a noun that it modifies, and must ensure that all chunks of that noun occur immediately to the left of the prepositional phrase.

Below is the sign representing iz ‘from’ as it occurs in the sentence Marko ima dobrog druga iz Beograda ‘Marko has a good friend from Belgrade’ which we examined above.

\begin{equation}
\vdash \lambda_{\nu\omega,\text{toZ}}[\text{k(PER w)} \bullet (\lambda_{z,\exists_{\epsilon}[(\text{k(PER v)} \epsilon)} \wedge \\
\quad s = (iz\#(\text{fst}_p \epsilon))_s \cdot (\text{rst}_s \epsilon)])] : z \rightarrow z \rightarrow z; \\
\text{NP}_{\text{gen,m,sg,3}} \rightarrow \text{NP}_{\text{acc,m,sg,3}} \rightarrow \text{NP}_{\text{acc,m,sg,3}}; \\
\lambda_{x\rho y}((\text{from x y}) \text{ and (P y)} : e \rightarrow (e \rightarrow p) \rightarrow (e \rightarrow p))
\end{equation}

Semantically, the preposition needs an argument of type e (its argument noun phrase), then an argument of type e \rightarrow p (the noun that the prepositional phrase

\textsuperscript{22}Relative pronouns do agree with the modified noun in gender and number; an analysis of relative clauses is unfortunately beyond the scope of this thesis.
will modify). The constant $\vdash \text{from} : e \to e \to p$ expresses a binary relation on individuals.

Now we examine the phenogrammatical term in more detail. Intuitively, the variable $v$ stands for the noun phrase argument of the preposition, while $w$ stands for the noun that the prepositional phrase modifies.

The preposition permutes the noun it modifies, then compacts it into a string of languages which is expressed by the subterm $k(\text{PER} \ w)$. The reason it permutes the noun is that, given the set of judgments we are considering now, the adjective and the noun can freely order with respect to each other before the postnominal modifier. So, if the noun is "dobrog druga ‘good friend’, it predicts that both "dobrog druga and "druga dobrog are possible.

As for its argument noun phrase, the preposition permutes it as well, for the same reasons as in the case of the modified noun. Then it compacts the resulting set of strings of languages into a set of strings. $t$ is one string in that set. This is all expressed in the subterm $(k(\text{PER} \ v) \ t)$.

The constant $\vdash \# : c \to p \to p$ takes a clitic and and phonological word and procliticizes the clitic onto that phonological word resulting in another phonological word. The preposition $iz$ in this way procliticizes onto the first word of its argument noun phrase, which is expressed in the subterm $(iz#(fst_p \ t))$. This new phonological word is then turned into a length one string and concatenated with the rest of the noun phrase, expressed in the subterm $(iz#(fst_p \ t)) s \cdot (rst_s \ t)$.
We then fuse the set of all strings constructed in this way, by procliticizing the preposition onto the first phonological word of the noun phrase, and then putting it together with the rest of the noun phrase, with the set of strings we obtained by compacting the permutations of the modified noun. The result is a set of strings in which the postnominal modifier, with the appropriately placed clitic, occurs to the right of the modified noun.

This $p$-language (of type $S$) is then converted into a length one $S$-string via $\text{toZ}$, which ensures that the entire phrasal noun, with the postnominal modifier and perhaps some attributive adjectives as well, remains contiguous and that it is impervious to any permutations the verb may require of its arguments.

More concretely, we show how to construct the noun phrase $\text{dobrog druga iz Beograda}$ ‘a good friend from Belgrade’. First we combine the preposition and its argument noun phrase, which results in the following sign:

$$
\lambda_w.\text{toZ}\left(k(\text{PER }w) \cdot (\lambda_s.\exists_z\left([k(\text{PER BEOGRADA}_z) \cdot t])_s \cdot (\text{rst}_s t))\right)_s : z \rightarrow z; \\
\text{NP}_{\text{acc,m,sg,3}} \rightarrow \text{NP}_{\text{acc,m,sg,3}}; \lambda_{P_y}.(\text{from belgrade }y) \text{ and } (P_y)
$$

In this case the object of the preposition is a one-word noun phrase, so $iz$ procliticizes onto $\text{beograda}$, which ultimately results in the string $(iz\#\text{beograda})_s$. That string is then concatenated with $e_p$, since $e_p$ is the identity for string concatenation. So, phenogrammatically, the second argument of $\bullet$ in the sign above simply denotes the set of strings $\{(iz\#\text{beograda})_s\}$. 

99
Then, we combine the adjective with the noun, and apply the rule [NC] which results in the following sign:

\[ \vdash \text{DOBROG}_z \circ \text{DRUGA}_z : z; \text{NP}_{\text{acc},m,\text{sg},3}; (\text{good friend}) : e \rightarrow p \]

Now \textit{iz Beograda} can combine with \textit{dobrog druga}, resulting in the following sign:

\[ \vdash \text{toZ}\left[k(\text{PER} (\text{DOBROG}_z \circ \text{DRUGA}_z)) \cdot (\lambda_s.\exists_t ((k(\text{PER \ BEOGRADA}_z) t) \land s = (iz\#(\text{fst}_p t)))_s \cdot (\text{rst}_s t))] : z; \text{NP}_{\text{acc},m,\text{sg},3}; \lambda_y. (\text{from belgrade y}) \text{ and (good friend y)} : e \rightarrow p \]

The phenogrammatical term denotes a length \( S \)-string constructed out of the language that contains exactly two \( p \)-strings, \( \text{dobrog}_x \cdot \text{druga}_s \cdot (iz\#\text{beograda})_s \), and \( \text{druga}_s \cdot \text{dobrog}_s \cdot (iz\#\text{beograda})_s \). Now we can apply the [Quant] rule, and turn this sign into a quantificational noun phrase which can then combine with a sentence with an accusative ‘gap’. Even though verbs allow free reordering of themselves and their noun phrase arguments, this noun phrase has to stay intact because it is a length one string of languages and the verb can’t take it apart. The entire noun phrase can however permute with respect to the subject and the verb.

For the less permissive sets of judgments with respect to the ordering of attributive adjectives and nouns, we can simplify the lexical entry given above for the preposition \textit{iz} ‘from’ and just give the following one:

\[ \vdash \lambda_{xy}.\text{toZ}\left[(\text{L} \circ y) \cdot (\lambda_s.\exists_t ((\text{L} \lor t) \land s = (iz\#(\text{fst}_p t)))_s \cdot (\text{rst}_s t))] : z \rightarrow z \rightarrow z; \text{NP}_{\text{gen},m,\text{sg},3} \rightarrow \text{NP}_{\text{acc},m,\text{sg},3} \rightarrow \text{NP}_{\text{acc},m,\text{sg},3}; \lambda_{xy}.(\text{from x y}) \text{ and (P y)} : e \rightarrow (e \rightarrow p) \rightarrow (e \rightarrow p) \]
This lexical entry works both for the grammar where the adjective and noun can reorder with respect to one another but must remain contiguous, and for the grammar where the adjective must precede the noun. Here, the modified noun and the argument noun phrase of the preposition are not required to permute. Everything else is the same as in the lexical entry for the most permissive grammar.

**Quantificational Determiners**

Recall that in the most permissive grammar quantificational determiners can in general be detached from their argument noun, even if that noun contains a postnominal modifier, in contrast to attributive adjectives which may not do so. In the less permissive case, the quantificational determiners have to immediately precede the argument noun (Zlatić, 1997).

For the more permissive grammar, we add the following lexical entry for the quantificational determiner *svaka* ‘every’.

\[(106) \vdash \lambda_{vF}.(F(z \text{ svaka} v)) : z \rightarrow (z \rightarrow z) \rightarrow z;\]
\[\text{NP}_{\text{nom,f,sg,3}} \rightarrow (\text{NP}_{\text{nom,f,sg,3}} \rightarrow \text{S}_{m,6}) \rightarrow \text{S}_{m,6};\]
\[\lambda_{xp}.(\text{every } x) P : (e \rightarrow p) \rightarrow (e \rightarrow p) \rightarrow p\]

This determiner has to pick up an argument of an appropriate noun phrase type first, but otherwise it works just like the quantificational pronouns we considered earlier in the chapter. Phenogrammatically, it concatenates itself with its first argument, and the resulting string of languages is simply lowered into the ‘gap’ of its second argument, the finite verb phrase.
Since the verbs in general permute themselves with their arguments and in this case the noun phrase which contains svaka ‘every’ is of length greater than one, the quantificational determiner can be detached from the rest of the noun phrase and freely reorder with respect to other constituents. This is the case no matter what the internal structure of its first argument, i.e. whether its first argument contains postnominal modifiers, or attributive adjectives, or not. So we predict that, in the permissive grammar, a sentence such as Svaka djevojka spava ‘Every girl sleeps’ can be pronounced six different way, and a sentence such as Svaka djevojka iz Beograda spava ‘Every girl from Belgrade sleeps’ can be pronounced also six different ways, since svaka can detach from the rest of the noun phrase material but djevojka iz Beograda, because of how we analyzed postnominal modification, must remain contiguous.

For the less permissive grammar which requires that the quantificational determiner occur immediately to the left of its argument, we give the following lexical entry:

\[
(107) \quad \lambda_{xP}.(\text{every } x) \to \text{P} : (e \to p) \to (e \to p) \to p
\]

The only difference is in the phenogrammatical term. In this case, the quantificational determiner concatenates itself with its first argument. The resulting string is turned into a length one S-string. This ensures that once the quantificational noun phrase combines with the verb phrase, the noun phrase cannot be made
discontinuous and the determiner must occur on its left periphery. In this grammar, a sentence such as *Svaka djevojka spava* ‘Every girl sleeps’ and a sentence such as *Svaka djevojka iz Beograda spava* ‘Every girl from Belgrade sleeps’ are predicted to be pronounceable two different ways, since all the noun phrase material must remain contiguous.

### 3.3.3 Adverbial Modifiers

#### Single-Word Adverbs

Tectogrammatically, we analyze adverbial expressions as verb phrase modifiers. This means that adverbs need a verb phrase argument, and output something of the same type. However, in our grammar there is strictly speaking no verb phrase type, since various finite verb phrases are tectogrammatically distinguished in terms of the kind of subject required to form a sentence. In other words, they differ in terms of number, gender and person parameters of their subject noun phrase. Adverbs need to be able to combine with a verb phrase requiring a subject of such-and-such gender, number and person, and output a modified verb phrase that retains those same subject requirements, so that the verb/subject agreement is retained.

We define $\text{Adv}$ as an abbreviation for the following type:

23Since adverbs can occur not just in main declarative clauses, but also in interrogative and embedded clauses, and can modify participial and infinitival verb phrases as well, we have to generalize their tectogrammatical types not just with respect to the subject noun phrase parameters but also with respect to the resulting clause type parameters. In this chapter, we abstract away from that and only deal with adverbs in main declarative clauses.
The dependent product type above ensures that the gender, number and person parameters of the subject required by the non-modified argument verb phrase are also required by the resulting, modified verb phrase.

Now we can give the following lexical entry for the adverb brzo ‘fast’:

\[
\lambda_{F,v,w}. \exists_x [(F \land (\text{BRZO}_z \land w)) : (z \rightarrow z) \rightarrow z \rightarrow z; \text{Adv}; \text{fast} : (e \rightarrow p) \rightarrow e \rightarrow p]
\]

Examining the phenogrammatical term of this sign in more detail, we see that the adverb first combines with the argument of type \( z \rightarrow z \) (\( F \), the verb phrase). The variable \( v \) is a placeholder for the subject. Recall that verbs build a set of all permutations of the string of languages consisting of itself, the subject and any objects, so the subterm \( F \land v \) above denotes the set of strings obtained by combining the verb phrase with its subject. \( x \) is one \( S \)-string in that set. A sentence that contains the adverb denotes a set of strings of languages \( w \), where \( w \) is any string obtained by shuffling the adverb into \( x \).

To give a concrete example, we show how to construct a representation of the sentence Marko vozi brzo ‘Marko drives fast’. Below are the lexical entries for the tectogrammatically appropriate versions of vozi and the adverb.

\[
\lambda_{F,v,w}. \exists_x [(F \land (\text{VOZI}_z \land w)) : (z \rightarrow Z) ; NP_{\text{nom,m,sg,3}} : S_{m,6}; \text{drive} : e \rightarrow p]
\]

\[
\lambda_{F,v,w}. \exists_x [(F \land (\text{BRZO}_z \land w)) : (z \rightarrow z) \rightarrow z \rightarrow z; \text{Adv}_{m,sg,3}; \text{fast} : (e \rightarrow p) \rightarrow e \rightarrow p]
\]
From these two signs and the lexical entry for Marko we can construct the following sign:

\[(112) \quad \vdash \lambda w. \exists x \left[ \left( \text{PER}(\text{MARKO}_z \circ \text{VOZI}_z) x \right) \land \left( \text{BRZO}_z \circ x \circ w \right) \right]; S_{m,6}; \]

\[\text{(fast drive marko)}: p\]

The subterm \(\text{PER}(\text{MARKO}_z \circ \text{VOZI}_z)\) denotes a set which contains exactly two strings of languages, \(\text{MARKO}_z \circ \text{VOZI}_z\) and \(\text{VOZI}_z \circ \text{MARKO}_z\). By shuffling \(\text{BRZO}_z\) into each of these two strings, we obtain the following set of strings of languages, which is precisely what the whole phenoterm of the sign above denotes:

\[(113) \quad \{ \text{MARKO}_z \circ \text{VOZI}_z \circ \text{BRZO}_z, \]

\[\text{MARKO}_z \circ \text{BRZO}_z \circ \text{VOZI}_z, \]

\[\text{VOZI}_z \circ \text{MARKO}_z \circ \text{BRZO}_z, \]

\[\text{VOZI}_z \circ \text{BRZO}_z \circ \text{MARKO}_z, \]

\[\text{BRZO}_z \circ \text{VOZI}_z \circ \text{MARKO}_z, \]

\[\text{BRZO}_z \circ \text{MARKO}_z \circ \text{VOZI}_z \}\]

After undergoing \([km]\), this sign phenoterm denotes six \(p\)-strings which correspond exactly to the six possible pronunciations of the sentence \(\text{Marko vozi brzo}\) ‘Marko drives fast’.

At this point the reader may be wondering why we are shuffling in the adverb by using \(\circ\), instead of letting it permute with the constituents in the sentence by using \(\text{PER}\). It doesn’t matter which of those functions we choose for the simple sentences we are considering here in which all verbal arguments are lexical noun phrases. However, when we extend the grammar to deal with more complex constituents such as phrasal noun phrases and sentential complements, if the adverb were to introduce \(\text{PER}\) in its phenogrammatical term, it could wreck
any pre-existing structures and islands of inflexible order already established in the verb phrase. Shuffling it in allows enough flexibility to get the possible word orders, without interfering too much with the existing word order in the verb phrase.

Adverbial Degrees

Next we turn to cases where an adverbial expression occurs with a degree. Recall that the degree must immediately precede the adverb it modifies, but the entire degree+adverb sequence can be freely ordered with respect to other clausal constituents. Semantically, we must analyze adverbial degrees as adverbial modifiers, of type \((\mathbf{e} \rightarrow \mathbf{p}) \rightarrow \mathbf{e} \rightarrow \mathbf{p}) \rightarrow \mathbf{e} \rightarrow \mathbf{p}\). To preserve verb/subject agreement, we define \(\text{Deg}\) to be the following tectogrammatical type:

\[
\text{Deg} = \text{def } \prod_{\text{nom},w: \text{Gdr},x:N,y: \text{Prs}}[(\text{NP}_{\text{nom},w,x,y} \rightarrow \text{S}_{m,6}) \rightarrow \text{NP}_{\text{nom},w,x,y} \rightarrow \text{S}_{m,6}) \rightarrow (\text{NP}_{\text{nom},w,x,y} \rightarrow \text{S}_{m,6}) \rightarrow \text{NP}_{\text{nom},w,x,y} \rightarrow \text{S}_{m,6}]
\]

Below we focus on the phenogrammatical part of the lexical entry for the adverbial degree \(\text{veoma} \text{ 'very'}\). Note that in the phenoterm below, \(\vdash F : z \rightarrow z\) and \(\vdash G : (z \rightarrow z) \rightarrow z \rightarrow z\).

\[
\vdash \lambda_{G_{Fvw}} \exists_{xy}[F \vee x] \land (G (\lambda_z.1_s) \circ y) \land w = t o Z(L(\text{VEoma}_z \circ y)) \circ x] : ((z \rightarrow z) \rightarrow z \rightarrow z) \rightarrow (z \rightarrow z) \rightarrow z \rightarrow z
\]

So, the degree first combines with the adverb (\(G\)). The resulting sign then combines with the verb phrase (\(F\)), and finally the subject (\(\vee\)). The subterm (\(F \vee\))
stands for the verb phrase combined with the subject. It denotes a set of that contains all permutations of the string consisting of the subject and the verb phrase. The subterm \((F \lor x)\) means that \(x\) is a string in \((F \lor)\).

As for the adverb, \(G\), the degree essentially ‘destroys’ all the argument slots in \(G\), by feeding it the empty \(S\)-language, \(1_S\), and then the empty \(S\)-string, \(e_S\). In the case of the adverb \(brzo\) ‘fast’, \((G(\lambda_x.1_S) e_S)\) would amount to the set of strings of languages that contains exactly one string of languages, namely, \(BRZO_S\). Call that string \(y\).

The degree then concatenates itself with \(y\), the adverb, and then ‘linguifies’ the resulting string via \(L\) thereby creating a set of strings. Finally, that set of strings is turned into a length-one string of languages via \(toZ\), which is then shuffled into \(x\), a string in the set of all permutations of the verb phrase and the subject.

More concretely, the sign representing the sentence \(Marko\ vozi\ veoma\ brzo\ ‘Marko drives very fast’\) has the following phenterm:

\[
\lambda_w.\exists_{xy}[P(E(R(MARKO_S \circ VOZI_S) x) \land \exists_{x'}[[1_S x'] \land (BRZO_S \circ x' y)]] \land w = toZ(L(VEOMA_S \circ y)) \circ x] : Z
\]

The fact that the \(S\)-string consisting of the degree and the adverb (of length two) is ‘linguified’ into the set \(\{vEOMA_S \cdot brzo_S\}\) which is then turned into length-one string of \(S\)-languages, ensures that the degree and the adverb remain contiguous and exactly in that order. Since \(\circ\) (and also \(PER\)) is a function on \(S\)-strings, it cannot pull apart the degree+adverb unit, since it has been turned into a length one
S-string. \( \odot \) cannot ‘see’ the internal structure of the range of that length one string of languages, i.e. it doesn’t have access to the set of strings \{veoma\_s \cdot brzo\_s\}.

**Prepositional Adverbial Modifiers**

Recall that in the less permissive grammar, prepositional phrases must remain contiguous. The preposition must occur immediately to the left of its argument noun phrase, and no discontinuities in the noun phrase are allowed either.

In the more permissive grammar, discontinuity is allowed only if the chunk of the noun phrase that the preposition procliticized onto precedes in the sentence the chunk of the noun phrase that the preposition did not procliticize onto.

We start with the less permissive set of judgments first, because the phenogrammatical part of the lexical entry for the preposition is very similar to the one we gave in the case of prepositional postnominal modifiers.

Suppose we are trying to generate the sentence Ana živi u velikom gradu ‘Ana lives in a big city’. We give the following lexical entry for the preposition that builds a verb phrase modifier:

\[
\lambda_{F_{Fv_{w}}} \exists_{x_{zt}}[(F \lor x) \land (L \lor t) \land z = \text{toZ}(\lambda_s.s = (u\#(f_{st_p}.t))_s \cdot (r_{st_s}.t))
\land ((z \odot x) w) : z \rightarrow (z \rightarrow Z) \rightarrow z \rightarrow Z; NP_{dat, m, sg, 3} \rightarrow \text{Adv};
\lambda_{y_{px} \cdot ((in y) P) x} : e \rightarrow (e \rightarrow p) \rightarrow e \rightarrow p
\]
Phenogrammatically, this sign is very similar to the sign for the preposition which builds a postnominal modifier. The preposition procliticizes onto the first phonological word in its argument noun phrase. The grammar treats the whole prepositional phrase as a length one S-string. Therefore, when such an adverbial prepositional phrase is shuffled into the sentence, it cannot be made discontinuous.

For the more permissive grammar, we give the following lexical entry for the preposition that builds a verb phrase modifier:

\[
(118) \quad \lambda_{vFw} \exists_{xzt} [(F \vee x) \wedge (k \text{ PER } y t) \wedge z = \text{toZ}(\lambda_{s.} s = (u\#(fst_{p} t)) s \circ \text{toZ}(rst_{s} t)) \wedge ((z \odot x) w)]: z \rightarrow (z \rightarrow z) \rightarrow z \rightarrow z;
\]

\[
\text{NP}_{\text{dat, m, sg, 3}} \rightarrow \text{Adv}; \lambda_{yPw} (((\in y) P) x): e \rightarrow (e \rightarrow p) \rightarrow e \rightarrow p
\]

Phenogrammatically, the difference between this lexical entry and the analogous entry in the less permissive grammar, is that the prepositional phrase is treated as an S-string of length two, not one. The first S-string is constructed out of the preposition and its host, and the second string is the remainder of the preposition’s object noun phrase. Therefore, when the adverbial prepositional phrase is shuffled into the sentence, the two strings can appear discontinuously, but the first string, built out of the proclitic and its host, must always precede the second string.
3.4 Conclusion

In this chapter we have analyzed some simple Serbo-Croatian sentences, but more importantly, we have illustrated how the grammar works, and seen the basic effects of some essential phenogrammatical functions such as \textbf{PER}, \textbf{G}, \textbf{L}, \textbf{k} and \textbf{toZ}, as well as the cliticization function \#, which will continue to play an important role in our theory of Serbo-Croatian word order. We have also made several generalizations about the representations of Serbo-Croatian expressions in the grammar which are summarized below.

All agreement features of nouns and noun phrases are built into the tectogrammatical types in the \textbf{N} and \textbf{NP} family. In order to account for the different word order possibilities, nouns and noun phrases are assigned to different combinations of tectogrammatical, phenogrammatical and semantic types, summarized in Table 3.2. These different combinations of semantic and tectogrammatical types are possible because we do not require a functional mapping between tectogrammatical and semantic types; we could have alternatively posited an additional tectogrammatical type, intermediate between nouns and noun phrases.

The first type in Table 3.2 corresponds to lexical nouns. Signs of that type are arguments and results of adjectival modification. Finally, signs of that type can undergo the [NC] rule which changes their tectogrammatical type from some \textbf{N} type to the corresponding \textbf{NP} type.
<table>
<thead>
<tr>
<th>TECTOGRAMMATICALLY TYPE</th>
<th>SEMANTIC TYPE</th>
<th>PHENOGRAMMATICALLY TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 N family</td>
<td>e → p</td>
<td>z</td>
</tr>
<tr>
<td>2 NP family</td>
<td>e → p</td>
<td>z</td>
</tr>
<tr>
<td>3 NP family</td>
<td>e</td>
<td>z</td>
</tr>
<tr>
<td>4 (NP → S) → S</td>
<td>(e → p) → p</td>
<td>(z → z) → z</td>
</tr>
</tbody>
</table>

Table 3.2: Summary of noun and noun phrase types.

The signs that tectogrammatically have NP types, but otherwise behave just like nouns, are the result of the application of the [NC] rule. They are also arguments and results of postnominal modification. Finally, they can be arguments of determiners as well as undergo the [Quant] rule.

The signs that tectogrammatically have NP types but semantically denote individuals are non-quantificational lexical noun phrases such as proper names.

Finally, the fourth type in the table above consists of quantificational noun phrases. This includes lexical quantificational noun phrases, the signs that result from the application of the [Quant] rule, as well as noun phrases that contain quantificational determiners.

Intransitive, transitive and ditransitive verbs combine with either lexical noun phrases which denote individuals, or are eventually picked up by quantificational noun phrases as arguments. Phenogrammatically, such verbs combine with arguments of type z and via PER construct sets of all permutations of themselves and
their arguments (type \( z \)), to account for the free ordering of verbs, and their objects and subjects. Quantificational noun phrases lower themselves into the ‘gap’, and therefore also participate in the free ordering of verbs and their arguments.

Adverbial phrases combine with verb phrases and output modified verb phrases with the same subject requirements. Phenogrammatically, they shuffle themselves into various permutations of the verb and its arguments via \( \diamond \). Adverbial degrees construct length-one \( S \)-strings from the degree+adverb unit via \( L \) and \( t o Z \), which ensures that the degree always immediately precedes the adverb as they are shuffled into the sentence.

Finally, prepositions, whether in the case of noun or verb phrase modification, procliticize onto the first word of their object noun phrase. Depending on the prepositional phrase in question, and how permissive one’s grammar is, the prepositional phrase is either turned into a length one \( S \)-string which requires it to remain intact, or it is turned into a length two \( S \)-string, which allows it to occur discontinuously in the sentence.
4.1 Introduction

Whereas in the previous chapter we considered only verbs whose arguments are noun phrases, in this chapter we turn our attention to verbs with more complex arguments. In particular, we analyze embedded declarative clauses, subject and object control structures and predicative complements.

It is necessary to first provide our general theory of embedding, control and predication, before we analyze enclitics, because (i) embedded clauses and controlled finite verb phrases are domains for enclitic placement, and (ii) the set of enclitics in Serbo-Croatian includes predicative and control verbs. In this chapter, we focus only on non-clitic predicative and control verbs, but we build our analysis of clitic verbs on the basic assumptions laid out here. In Chapter 7 we further refine the theory of embedding, control and predication presented here.
4.2 Embedded Declarative Clauses

4.2.1 Data

Embedded declarative clauses in Serbo-Croatian must occur with a complementizer. There are different complementizers and the two most common declarative complementizers are \( \text{da}^{24} \) and \( \text{što} \). Clause embedding expressions require an embedded clause with a certain complementizer; i.e. not all types of embedded clauses are compatible with all embedding expressions. Although here we mainly focus on clauses headed by \( \text{da} \), the obligatoriness of a particular kind of complementizer is illustrated in the examples below.

\[(119)\]

a. Ana misli da Marko spava.
   Ana\(_{NOM,f,sg,3}\) think\(_{sg,3}\) DA Marko\(_{NOM,m,sg,3}\) sleep\(_{sg,3}\)
   ‘Ana thinks that Marko is sleeping’

b. * Ana misli što Marko spava.

c. * Ana misli Marko spava.

\[(120)\]

a. Ani smeta što Marko stalno spava.
   Ani\(_{DAT,f,sg,3}\) bother\(_{sg,3}\) ŠTO Marko\(_{NOM,m,sg,3}\) always sleep\(_{sg,3}\)
   ‘It bothers Ana that Marko is always sleeping’

b. * Ani smeta da Marko stalno spava.

c. * Ani smeta Marko stalno spava.

\(^{24}\)It has been argued (see, for example, Zec (1987) and references therein) that there are two distinct expressions, \( \text{da}_1 \) and \( \text{da}_2 \), with \( \text{da}_2 \) having certain modal properties and restricting the tense of its complement sentence to perfective present. We do not postulate two distinct types of the complementizer, and we largely ignore modal differences between embedded declaratives, since a theory of modality is beyond the scope of this thesis. In Chapter 7 we will address the distinction between so called S-verbs (subjunctive like) and I-verbs (indicative like; see Progovac (2005)) with respect to clitic climbing in controlled complements.
The kinds of constituents which can freely order inside a main declarative clause can also freely reorder inside an embedded declarative clause. The complementizer, however, must be leftmost in the embedded clause.

(121) a. Ana misli da Marko voli Vesnu.
   Ana_{NOM,f,sg,3} think_{sg,3} DA Marko_{NOM,m,sg,3} love_{sg,3} Vesna_{ACC,f,sg,3}
   ‘Ana thinks that Marko loves Vesna’

   b. Ana misli da Marko Vesnu voli.
   c. Ana misli da voli Vesnu Marko.
   d. Ana misli da Vesnu Marko voli.
   e. * Ana misli voli da Marko Vesnu.

Embedded clauses cannot be made discontinuous. That is, main clause material cannot occur inside the embedded clause:

(122) a. Ana misli da Marko spava.
   Ana_{NOM,f,sg,3} think_{sg,3} DA Marko_{NOM,m,sg,3} sleep_{sg,3}
   ‘Ana thinks that Marko is sleeping’

   b. * Ana da Marko misli spava.
   d. etc.

Embedded clauses also have to occur on the right edge of the matrix clause and cannot freely reorder with respect to the verb and the subject:

(123) a. Ana misli da Marko spava.
   Ana_{NOM,f,sg,3} think_{sg,3} DA Marko_{NOM,m,sg,3} sleep_{sg,3}
   ‘Ana thinks that Marko is sleeping’

   b. * Ana da Marko spava misli.
   d. etc.
The only exception to this are matrix clause adverbial expressions, which may occur immediately to the right of the embedded clause, though they still can’t occur inside of the embedded clause. Below we consider a sentence in which the matrix verb is compatible with an adverbial prepositional phrase with a dative complement (denoting a location), but the embedded verb is not.

(124) a. Ana je rekla da će Marko doći na sastanku, ali meni je kasnije rekla da neće doći.
   AnaNOM,f,sg,3 issg,3 sayppl,f,sg DA will3,sg MarkoNOM,m,sg,3 comeinf
   na sastanku, ali meni je kasnije rekla da neće doći.
tatmeetingDAT,m,sg butIDAT,s,sg after sayppl,f,sg DA notwillsg,3 comeinf

‘Ana said at the meeting that Marko would/will come, but later she told me he wouldn’t/won’t

# ‘Ana said that Marko would/will come to the meeting, but later she told me he wouldn’t/won’t’

b. Ana je rekla na sastanku da će Marko doći.

c. Ana je na sastanku rekla da će Marko doći.

d. Na sastanku je Ana rekla da će Marko doći.

e. * Ana je rekla da će na sastanku Marko doći.

The adverbial in the (a) sentence cannot be modifying the embedded clause, that is, it cannot mean ‘to the meeting’, so it must be a matrix adverbial. That adverbial can, just like adverbials in general, freely reorder with respect to other clausal constituents (b-d), but cannot occur inside of the embedded clause (e).

If we pick an adverbial that is compatible with both the matrix and the embedded verb, and place that adverbial on the right edge of the sentence, ambiguity will arise:
The sentence (a) is ambiguous with respect to the adverbial interpretation. In sentences (b) and (c) we provide disambiguating context to draw out each possible interpretation of the adverbial.

The preceding examples were just supposed to show that main clause adverbials are special in comparison with other main clause constituents in that they can occur to the right of an embedded clause, which can give rise to ambiguities.

4.2.2 Analysis

So far we’ve only been concerned with clauses whose $K$ parameter is $m$, i.e. main declarative clauses, since the ultimate result type of finite verbs is $S_{m,6}$. We analyze complementizers such as $da$ as expressions that turn main declarative clauses into embedded declarative clauses, whose type is $S_{e,6}$. 

117
Suppose we are trying to generate the sentence Ana misli da Marko spava ‘Ana thinks that Marko is sleeping’. Below are the lexical entries for the complementizer *da* and the sentence embedding verb *misli*:

\[(126) \vdash \lambda_{Xw} \exists_v [(X v) \land w = (DA_z \circ v)] : Z \rightarrow z; S_{m,6} \rightarrow S_{e,6}; \lambda_{q,q} : p \rightarrow p\]

\[(127) \vdash \lambda_{Xvw} \exists_y [(PER(v \circ MISLI_z) y) \land w = y \circ toZ(k x)] : Z \rightarrow z \rightarrow z; S_{e,6} \rightarrow NP_{nom,f,sg,3} \rightarrow S_{m,6}; \text{think} : p \rightarrow e \rightarrow p\]

Tectogrammatically, the complementizer picks up a main declarative clause and outputs an embedded declarative clause. Semantically, it’s an identity function on propositions, meaning that it does not affect the meaning of its complement clause. Phenogrammatically, its lexical entry ensures that the complementizer occurs on the left edge of the embedded clause, with the rest of the embedded clause material reordering insofar as the embedded verb allows such reordering of itself and its arguments.

The sentence embedding verb *misli* ‘thinks’ needs an embedded clause argument and a subject argument to make a main declarative sentence. Semantically, it expresses a relation between an individual and a proposition. Phenogrammatically, such a verb allows free reordering of itself with its subject. However, it turns its sentential complement into a length one string of languages which ensures that no embedded clause material can escape into the matrix clause and that no matrix clause material may occur inside the embedded clause. It then requires the embedded clause to occur on the right edge of the main clause, after some permutation of itself and its subject.
Below we show how to construct the sentence Ana misli da Marko spava ‘Ana thinks that Marko is sleeping’. First, the complementizer combines with the declarative main clause Marko spava ‘Marko sleeps’ and turns it into an embedded clause with the same meaning.

\[(128) \vdash \lambda_{xw}\exists_v[(x \circ v) \land w = (DA_z \circ v)]; S_{m,6} \rightarrow S_{e,6}; \lambda_q q\]

Next, the sentence embedding verb combines with its sentential complement, resulting in the following sign:

\[(129) \vdash \lambda_{xy}\exists_z[(PER(MARKO_z \circ SPAVA_z) z) \land y = z \circ toZ(k (\lambda_{xw}\exists_v[(PER(MARKO_z \circ SPAVA_z) v) \land w = (DA_z \circ v)]))]; z \rightarrow Z; NP_{nom,f,sg,3} \rightarrow S_{m,6}; \text{think(sleep marko)} : e \rightarrow p\]

Finally, this verb phrase can combine with the subject, resulting in the following sign:

\[(130) \vdash \lambda_y\exists_z[(PER(ANA_z \circ MISLI_z) z) \land y = z \circ toZ(k (\lambda_{xw}\exists_v[(PER(MARKO_z \circ SPAVA_z) v) \land w = (DA_z \circ v)]))]; Z; S_{m,6}; \text{think(sleep marko)ana} : p\]

Tectogrammatically, this is a main declarative clause. Semantically it expresses the proposition that Ana thinks that Marko sleeps. Phenogrammatically, it denotes a set of S-strings, each of which consists of some permutation of the verb and the subject, followed by the length one S-string constructed out of the embedded clause in which the complementizer is always initial.
Since we analyze adverbial expressions as shuffling into the sentence via $\odot$, we predict that matrix adverbials will be able to occur anywhere in the main clause, including at its right edge after the embedded clause. Since the verb turns the embedded clause into a length one string of languages, the matrix adverb will not, however, be able to occur inside the embedded clause.

4.3 Predicatives

4.3.1 Data

In this section, we are concerned with complements of the verb *biti* ‘be’. It has a perfective and an imperfective present tense paradigm, the latter consisting of full and enclitic forms. Both paradigms are presented in Table 4.1.

This verb can take a variety of complements, including predicative adjectives, predicative prepositional phrases, noun phrases, passive participles and past participle, the latter being used in the periphrastic past tense construction.
Predicative prepositional phrases do not agree with the matrix subject at all. Adjectives and passive participles, which have the same morphology as adjectives, must occur in the nominative case in predicative contexts. Predicative adjectives, passive participles and past participles must agree with the subject in number and gender.

Predicative noun phrases also must occur in the nominative case, but do not necessarily have to agree with the subject in gender and number. For example, in addition to (c) above, where the subject and the predicative noun phrase agree in gender and number, the following are also possible:

(132)  a. Ona je student.
       she_{NOM,f,sg,3} is_{sg,3} student_{NOM,m,sg,3}
       ‘She is a student’
b. Huligani su veliki problem u našem društvu.
   hooligan NOM, m, pl, 3 are pl, 3 big NOM, m, sgl problem NOM, m, sgl in our DAT, n, sgl society DAT, n, sgl
   ‘Hooligans are a big problem in our society’

In the next chapter, we will analyze the clitic forms of the verb *biti*. In this chapter we will abstract away from that complicating factor and only consider the full forms *biti*. Here we only mention, but do not analyze, the conditional mood construction which consists of the aorist of *biti* and a past participle, since there are no non-clitic forms of the aorist of *biti*; we return to them in Chapter 5. Below, *biste* is glossed as ‘would’ but it is really an aorist clitic of *biti*.

(133) Vi biste kupili to.
   you NOM, pl, 2 would pl, 2 buy ppl, m, pl that ACC, n, sgl
   ‘You would buy that’

Considering only predicative structures that contain non-clitic forms of the copula, the word order is largely unrestricted. For example, a predicative adjective, or a predicative noun phrase can freely order with respect to the verb and the subject:

(134) a. Igor jeste pametan.
   Igor NOM, m, sgl is sgl, 3 smart NOM, m, sgl
   ‘Igor is smart’
   b. Pametan jeste Igor.
   c. Pametan Igor jeste.
   d. Jeste Igor pametan.
   e. Jeste Pametan Igor.
   f. Igor pametan jeste.
Verbal predicative complements, that is passive and past participles, can also freely order with respect to other clausal constituents. If these participles have arguments of their own, those arguments can also be freely ordered in the sentence.

However, if the copula combines with a control verb participle which has as its argument an infinitival or finite controlled complement, word order becomes more
complex and the controlled complement tends to remain contiguous and occur at the right edge of the main clause. We return to the analysis of such structures in Chapter 7.

As for prepositional phrases, as in the case of adverbial prepositional phrases, we will entertain two sets of judgments. A less permissive set of judgments requires that the entire predicative prepositional phrase remain contiguous, but freely order with respect to other clausal constituents. A more permissive set of judgments accepts discontinuities in the prepositional phrase so long as the chunk which contains the preposition precedes the chunk of the prepositional phrase which does not.

Below we present an analysis of these predicative structures, and also try establish connections with the remainder of the grammar by exploring the relationships between predicative complements and the counterparts of those expressions that occur as verbal or nominal modifiers.

4.3.2 Analysis

We introduce a family of tectogrammatical types \( \text{Prd}_x \) which are the result type of predicative phrases. This type is parametrized by terms of type D (that is, \( x \) in \( \text{Prd}_x \) is of type D), which represent the kind of predicative complement in question, namely, \( \text{ps}, \text{pl}, \text{n}, \text{a} \) and \( \text{pp} \) for passive participles, past participles, noun phrases, adjectives and prepositional phrases respectively.
Past and Passive Participles

Recall that past and passive participles agree with the subject in gender and number. In addition, passive participles must occur in nominative case predicatively.

Suppose we are trying to construct a representation of the sentence Marko jeste spavao ‘Marko slept’. Below we give the lexical entry for the past participle spavao:

\[(138) \vdash \text{SPAVAO}_z : z; \prod_{p:PN}[\text{NP}_{\text{nom},m,sg,p} \rightarrow \text{Prd}_{pl}]; \lambda x.\text{PST}(\text{sleep } x) : e \rightarrow p]\]

The participle spavao requires of its subject that it be a masculine singular nominative noun phrase, but it does not care about the person of the subject, i.e. \(Ja \text{ jesam spavao} ‘I slept’\) and \(Ti \text{ jesi spavao} ‘You slept’\) are both possible, in addition to many similar sentences with a 3rd person subject. For this reason, its tectogrammatical type is a dependent product.

The non-clitic 3rd person singular form jeste ‘is’ is represented by the following sign:

\[(139) \vdash \lambda_{w,v,\text{PER}}(w \circ \text{JESTE}_z \circ v) : z \rightarrow z \rightarrow z;\]

\[\prod_{g:Gdr,d:D}[\text{(NP}_{\text{nom},g,sg,3} \rightarrow \text{Prd}_{d}) \rightarrow \text{NP}_{\text{nom},g,sg,3} \rightarrow \text{S}_{m,6}];\]

\[\lambda_{\text{Px}}(\text{Px}) : (e \rightarrow p) \rightarrow e \rightarrow p\]

The auxiliary jeste requires of its subject that it be a singular 3rd person singular nominative noun phrase but it doesn’t care about its gender. Also, it will take any

\[25\text{We abstract away from a tense analysis and simply assume that there is a propositional operator } \vdash \text{PST} : p \rightarrow p \text{ which contributes the correct temporal interpretation.}\]
predicative phrase as its complement. This is why its tectogrammatical type is a
dependent product. Below is this same sign with the gender parameter instan-
tiated as masculine, i.e. looking for a masculine subject as in Marko jeste spavao
‘Marko slept’:

(140) \[ \lambda_{v_{w}.PE}(w \circ JESTE_{z} \circ v) : z \rightarrow z ; \]
(Prd_{pl}) \rightarrow NP_{nom,m,sg,3} \rightarrow NP_{nom,m,sg,3} \rightarrow S_{m,6} ;
\lambda_{Px,PX} : (e \rightarrow p) \rightarrow e \rightarrow p

Semantically, jeste predicates its complement’s meaning of its subject’s meaning.
Phenogrammatically, it permutes itself with its arguments, resulting in a set of
S-strings. When we combine the copula with the participle we get the following
sign:

(141) \[ \lambda_{w.PE}(w \circ JESTE_{z} \circ SAVA_{z}) : z \rightarrow Z ; \]
NP_{nom,m,sg,3} \rightarrow S_{m,6} ; \lambda_{x.PST} : (e \rightarrow p) \rightarrow e \rightarrow p

Because of the tectogrammatical typing of the copula and the past participle it is
impossible to introduce a subject with an inappropriate case, or gender, person or
number features.

As another example, below we give a lexical entry for a past participle of a
transitive verb kupiti ‘to buy’ and the sign that represents the sentence Ana jeste
kupila knjigu ‘Ana bought a book’.

(142) \[ \lambda_{v.KUPILA_{z} \circ v} : z \rightarrow z ; \]
\prod_{Gd,n;Num,p,p'}:Prs[NP_{acc,g,n,p'} \rightarrow NP_{nom,f,sg,p} \rightarrow Prd_{pl}] ;
\lambda_{yx.PST} : (e \rightarrow y) \rightarrow e \rightarrow p

(143) \[ \lambda_{x.JESTE_{z} \circ KUPILA_{z} \circ KNJIGU_{z} : z ; S_{m,6} ;} \]
exists (book) (\lambda_{x.PST} : (e \rightarrow y) : p)
Recall that for passive participles we introduced another constant of type D, namely ps. Below we give a lexical entry for the passive participle pročitana ‘read’.

(144) \( \vdash \text{PROČITANA}_z : z; \prod_{p: \text{Prs}}[\text{NP}_{\text{nom},f,sg,p} \rightarrow \text{Prd}_{ps}]; \lambda_x.\exists(\text{person})(\lambda_y.\text{read } x y) : e \rightarrow p \)

This passive participle requires of the subject that it be a nominative feminine singular noun phrase, of whatever person. Below is the result of instantiating the D parameter in the lexical entry of jeste ‘is’ as ps instead of pl, so it can combine with a passive participle.

(145) \( \vdash \lambda_v.\text{PER}(w \circ \text{JESTE}_z \circ v) : z \rightarrow z \rightarrow z; \prod_{g: \text{Gdr}}[(\text{NP}_{\text{nom},g,sg,3} \rightarrow \text{Prd}_{ps}) \rightarrow (\text{NP}_{\text{nom},g,sg,3} \circ \text{Sm},6)]; \lambda_x.(P_x) : (e \rightarrow p) \rightarrow e \rightarrow p \)

Now we can represent the sentence Knjiga je pročitana ‘A book is read’ as follows:

(146) \( \vdash \text{PER}(\text{KNJIGA}_z \circ \text{JESTE}_z \circ \text{PROČITANA}_z) : z; \text{Sm},6; \exists(\text{book})(\lambda_x.\exists(\text{person})(\lambda_y.\text{read } x y)) : p \)

As another example, below we give a lexical entry for a neuter singular passive participle of a ditransitive verb, and the representation of the sentence Pismo jeste poslano Ani ‘A letter is/has been sent to Ana’.

(147) \( \vdash \lambda_v.\text{POSLANO}_z \circ v : z \rightarrow z; \prod_{g: \text{Gdr},n: \text{Num},p': \text{Prs}}[\text{NP}_{\text{dat},g,n,p'} \circ (\text{NP}_{\text{nom},n,sg,p} \circ \text{Prd}_{ps})]; \lambda_x.\exists(\text{person})(\lambda_y.\text{send } x z y) : e \rightarrow e \rightarrow p \)

(148) \( \vdash \text{PER}(\text{PISMO}_z \circ \text{JESTE}_z \circ \text{POSLANO}_z \circ \text{ANI}_z) : z; \text{Sm},6; \exists(\text{letter})(\lambda_x.\exists(\text{person})(\lambda_y.\text{send } x \text{ ana } y)) : p \)
**Predicative Noun Phrases and Adjectives**

Recall that predicative noun phrases have to be nominative but in general don’t have to agree with subjects in gender or number. Consider the sentence *Ana jeste student* ‘Ana is a student’.

We give the following lexical entry for the predicative version of *student*.

\[(149) \vdash \text{STUDENT} \circ z; \prod_{Gd} \circ \text{Num},n,p : \text{Prs}[\text{NP}_{\text{n}\text{om},g,n,p} \rightarrow \text{Prd}_n]; \text{student} : e \rightarrow p\]

Of course, if we wanted to enforce gender or number agreement between a predicative noun phrase and the subject, in case we believed that the two have to agree, we could appropriately instantiate the relevant parameters. Also note that we are assuming that the semantic type of a predicative noun phrase is \( e \rightarrow p \). We will return to this later in the chapter.

Below is the tectogrammatical version of the copula looking for a predicative noun phrase complement.

\[(150) \vdash \lambda_{vw,\text{PER}}(w \circ \text{JESTE} \circ v) : z \rightarrow z \rightarrow z; \]
\[\prod_{Gd}[\text{NP}_{\text{n}\text{om},g,3} \rightarrow \text{Prd}_n] \rightarrow \text{NP}_{\text{n}\text{om},g,3} \rightarrow \text{Sm}_6]; \lambda_{Px}(Px) : (e \rightarrow p) \rightarrow e \rightarrow p\]

The sentence *Ana jeste student* ‘Ana is a student’ is represented in the grammar by the following sign:

\[(151) \vdash \text{PER}(\text{ANA} \circ \text{JESTE} \circ \text{STUDENT}) : z; \text{Sm}_6; (\text{student ana}) : p\]

For predicative adjectives, we must enforce gender and number agreement between the adjective and the subject. Consider the sentence *Marko jeste pametan* ‘Marko is smart’. We give the following lexical entry for the predicative adjective.
The sentence *Marko jeste pametan* ‘Marko is smart’ is represented in the grammar by the following sign.

\[ \vdash \text{PER}(\text{MARKO} \circ \text{JESTE} \circ \text{PAMETAN}) : \text{z}; \text{S}_{m,6}; (\text{smart marko}) : \text{p} \]

**Predicative Prepositional Phrases**

A predicative prepositional phrase does not care about the subject’s agreement features, except that it be nominative. Recall that prepositions in Serbo-Croatian are proclitics. Phenogrammatically, we analyze prepositions in predicative prepositional phrases similar to postnominal prepositional phrases. That is, the preposition permutes its complement noun phrase, then procliticizes onto the first word of its complement. The entire prepositional phrase is then turned into a length one string of languages, ensuring that it remains contiguous.

Consider the sentence *Marko jeste iz Beograda* ‘Marko is from Belgrade’. We give the following lexical entry for the preposition *iz* ‘from’ which takes a 3rd person masculine singular genitive noun phrase complement and builds a predicative prepositional phrase.

\[ \vdash \lambda_{xy}.(\text{from } x y) : \text{e} \rightarrow \text{e} \rightarrow \text{p} \]

This lexical entry accounts for the less permissive set of judgments because it does not allow any discontinuities in the prepositional phrase. According to the more permissive set of judgments, the predicative prepositional phrase, as in the case
of adverbial prepositional phrases, may be split into two chunks, one consisting of the preposition procliticized onto some word of its complement noun phrase, and the other of the remainder of the complement noun phrase, with the condition that the chunk containing the preposition precede the other chunk in the sentence.

To account for this set of judgments, we analyze a predicative prepositional phrase as an expression which takes as an argument a finite sentence missing a predicative prepositional phrase. This allows the prepositional phrase to split itself into two chunks and then shuffle into the sentence. The phenogrammatical term of this sign is more like the permissive lexical entry for an adverbial prepositional phrase, and not like the one for a postnominal modifier prepositional phrase.

\[
\lambda_{yFw.}\exists_{xyzt}[(F \in x) \land (k(\text{PER} y)t) \land z = \text{toZ}(\lambda_{s.s} = (iz#(\text{fst}_s t)))_s \\
\circ \text{toZ}(_{rst_s t}) \land ((z \odot x) w) : z \rightarrow (z \rightarrow Z) \rightarrow Z; \\
\text{NP}_{\text{gen,m,sg,3}} \circ \prod_{g: \text{Gdr},r: \text{Num},p:\text{Prd}[((\text{NP}_{\text{nom,g,n,p}} \circ \text{Prd}_{pp}) \circ \text{S}_{m,6}) \circ \text{S}_{m,6}]; \\
\lambda_{yp}(P(\text{from } y)) : e \rightarrow ((e \rightarrow p) \rightarrow p) \rightarrow p]
\]

**Predication and Nominal Modification**

In this section, we explore connections between predicative phrases and noun modifiers, in order to streamline the grammar as much as possible.
As for past participles, they cannot be used to modify nouns at all, and in fact only appear in the periphrastic past tense and the conditional mood construction, the discussion of the latter having been deferred until Chapter 5.26

All passive phrases, by which we mean a passive participle and any of its arguments excluding the subject, can be used as postnominal modifiers. However, recall that in predicative uses, passive participles must occur in nominative case. As postnominal modifiers, they must agree with the noun they are modifying in case (and number and gender, just like adjectives). For example:

(156) a. Ana voli tursku kafu skuhanu
Ana NOM,f,sg,3 love,sg,3 Turkish ACC,f,sg coffee ACC,f,sg cooked pass,ACC,f,sg sa puno šećera.
with lots sugar GEN,m,sg
'Ana likes Turkish coffee cooked with lots of sugar'

b. Marko ne voli hranu pripremljenu sa puno začina.
Marko NOM,m,sg,3 not-love,sg,3 food ACC,f,sg prepared pass,ACC,f,sg with puno začina.
lots spices GEN,m,pl
'Marko doesn’t like food prepared with lots of spices'

While we can write a non-logical rule that maps predicative passive phrases to postnominal modifiers, since all predicative passive participles are nominative, we would only generate a small subset of passive postnominal modifiers, namely, only those that modify nominative nouns. For the (many) other cases, we would have to directly add lexical entries for such passive phrases.

26Past participles also occur without the copula in some kind of not very productive optative like construction, for example Živjeli ppl ‘May we live on!’ (used as ‘Cheers!’), or, in an old Chernobyl joke Tražila ppl te majka gaigerovim brojačem! ‘May your mother look for you with a Geiger counter!’ In this use, the past participles also occur in many profane expressions.
For example, here is a non-logical rule, call it [psNP] that takes a predicative passive phrase and outputs a postnominal modifier of nominative nouns:

\[
\vdash \phi : z; \text{NP}_{\text{nom},t,t',3} \rightarrow \text{Prd}_{\text{ps}}; \sigma : e \rightarrow p
\]

\[
\vdash \lambda_v. \text{toZ}(k(\text{PER } v) \bullet (k(\text{PER } \phi))) : z \rightarrow z; \text{NP}_{\text{nom},t,t',3} \rightarrow \text{NP}_{\text{nom},t,t',3};
\]

\[
\lambda_{\text{Py}.}(P y) \text{and}(\sigma y) : (e \rightarrow p) \rightarrow e \rightarrow p
\]

Recall that postnominal modifiers combine with nouns whose tectogrammatical types have been converted to the corresponding noun phrase types. Since all such phrases are 3rd person, the passive phrase must have its person parameter instantiated as 3, before it can undergo this rule. Here is an example of an output of this rule:

\[
\vdash \lambda_v. \text{toZ}(k(\text{PER } v) \bullet (k(\text{PER } \phi))) : z \rightarrow z;
\]

\[
\text{NP}_{\text{nom},n,sg,3} \rightarrow \text{NP}_{\text{nom},n,sg,3};
\]

\[
\lambda_{\text{Py}.}\exists(\text{person})(\lambda_y.\text{send } z \text{ ana } y) \text{ and } (P z) : (e \rightarrow p) \rightarrow e \rightarrow p
\]

If we want an appropriately case marked version of this passive phrase that can modify, say, instrumental nouns, we must introduce a new lexical entry for the passive participle itself, like so:

\[
\vdash \lambda_{w_v}. \text{toZ}(k(\text{PER } v) \bullet (k(\text{PER } \phi))) : z \rightarrow z; \text{NP}_{\text{nom},n,sg,3} \rightarrow \text{NP}_{\text{inst},n,sg,3};
\]

\[
\lambda_{x_P z}.\exists(\text{person})(\lambda_y.\text{send } z x y) \text{ and } (P z) : (e \rightarrow p) \rightarrow e \rightarrow p
\]

Further, all passive participles (just participles, not passive phrases) can be used as attributive adjectives. However, again the issue is that in attributive uses passive participles agree with nouns they modify in case (and number and gender).
We will just assume that we have to explicitly assert lexical entries for passive participles which behave like attributive adjectives.

As for predicative adjectives, not all attributive adjectives can be used predicatively, for example *navodni* ‘alleged’. Assuming that all predicative adjectives can be used attributively (and we are not aware of counterexamples to this generalization), we can give the following non-logical rule, call it [aN], which maps predicative adjectives to their attributive counterparts.

(160)

\[
\begin{align*}
\vdash \phi : z; NP_{\text{nom}, \tau, \tau'} & \rightarrow o Prd_{\tau}; \sigma : e \rightarrow p \\
\vdash \lambda_{\nu.\phi \circ \nu} : z \rightarrow z; N_{\text{nom}, \tau, \tau'} & \rightarrow o N_{\text{nom}, \tau, \tau'}; \\
\lambda_{P_{\gamma}(P \ y) \text{and}(\sigma \ y)} : (e \rightarrow p) & \rightarrow e \rightarrow p
\end{align*}
\]

This rule converts a predicative adjective into an attributive adjective which behaves permissively, i.e. it’s detachable from the remainder of the noun. We leave it to the reader to formulate an appropriate version of the rule that will output attributive adjectives which behave in accordance with more restrictive judgments concerning the ordering of noun phrase material discussed in the previous chapter.

As in the case of passive phrases, the conversion via this rule only works for nominative adjectives, since all predicative adjectives are nominative. We have to independently introduce lexical entries for differently case marked versions of such adjectives. We also have to introduce lexical entries of attributive adjectives, nominative and otherwise, which do not have predicative counterparts.
We analyze predicative noun phrases as derived from signs which are targets of postnominal modification, that is, possibly phrasal nouns which are semantically of type $e \to p$ but tectogrammatically have a noun phrase type. We independently motivated this mismatch between the tectogrammatical and semantic typing in Chapter 4, to account for certain facts concerning the word order within phrasal noun phrases which contain both attributive and postnominal modifiers. Now, we exploit the fact that we already have such signs in the grammar to generate the set of predicative noun phrases.

The following rule, call it [NPn], maps signs which are tectogrammatically nominative case marked noun phrases but semantically of type $e \to p$ into predicative phrases.

(161)  
\[
\begin{align*}
\vdash \phi : z; \NP_{\text{nom}, r', \beta', \beta}; \sigma : e \to p \\
\vdash \phi : z; \prod g; G\text{dr, n}; \text{Num, p}; \text{Prs}[\NP_{\text{nom}, g, n, p} \to \text{Prdn}]; \sigma : e \to p \quad \text{[NPn]}
\end{align*}
\]

This rule has an additional advantage in that it allows modification of nouns to proceed as usual. Once any attributive adjectives and postnominal modifiers have combined with a nominative case marked noun, it can undergo this rule and become a predicative phrase. This version of the rule produces predicative phrases which do not agree with the subject in number or gender, in accordance with the empirical generalization presented earlier in this chapter. We leave it to the reader.
to formulate less permissive versions of this rule which would impose more stringent agreement requirements on the predicative phrase which is its output.

Finally, all predicative prepositional phrases can be used as postnominal modifiers. The non-logical rule below, call it \([\text{ppN}]\), maps predicative prepositional phrases into postnominal modifiers. Recall that we entertained two very different lexical entries for predicative prepositional phrases, one which allows discontinuities in the prepositional phrase and the other one which doesn’t. However, all the grammars converge on not allowing discontinuities in the noun+postnominal modifier sequence. We will accordingly give two formulations of the rule though both rules have to output the same kind of thing, a postnominal modifier which must remain contiguous and occur immediately to the right of the noun it modifies. We start with the version of the rule that would be added to the less permissive grammar where discontinuities in predicative prepositional phrases are not allowed.

(162) version 1 - less permissive grammar

\[ \vdash \phi : z; \prod_{\text{Gdr},n: \text{Num},p: \text{Pns}} [\text{NP}_{\text{nom},g,n,p} \rightarrow \text{Prd}_{pp}]; \]

\[ \sigma : e \rightarrow p \]

\[ \vdash \lambda \text{L}. \text{toZ}((\text{L} \nu) \bullet (\text{L} \phi)) : z \rightarrow z; \prod_{\text{Cse},g: \text{Gdr},n: \text{Num}} [\text{NP}_{c,g,n,3} \rightarrow \text{NP}_{c,g,n,3}]; \]

\[ \lambda \text{Px} (P x) \text{and} (\sigma x) : (e \rightarrow p) \rightarrow e \rightarrow p \]

The output of this rule is a postnominal modifier which can combine with nouns (tectogrammatically associated with a noun phrase type) of any case, gender and number. The phenomenological term ensures that the prepositional phrase occur immediately to the right of the noun it modifies. The prepositional phrase and
the noun form a length one string of languages which ensures that they remain continuous and in exactly that order.

Recall that in order for a predicative prepositional phrase to occur discontinuously in a sentence, we had to analyze it as a functor over sentences missing a predicative prepositional phrase. Now we give a non-logical rule that can be added to the more permissive grammar in which predicative prepositional phrases are allowed to occur discontinuously.

\[ (163) \text{ version 2 - more permissive grammar} \]

This complicated rule essentially outputs the same kind of signs as its counterpart in the less permissive grammar, by outputting prepositional phrases which are postnominal modifiers, and requiring them to occur contiguously and immediately to the right of the noun they modify.

### 4.4 Subject and Object Control

#### 4.4.1 Data

Here, we consider verbs which are object or subject control verbs. By *subject control* verb we mean a verb which has a finite verb phrase complement (and possibly a noun phrase object as well), and whose subject, which occurs in the matrix clause, is also interpreted as the embedded subject. By *object control* verb we mean
a verb which has a finite verb phrase complement and a noun phrase object, and whose object is interpreted as the embedded subject. We will perhaps abuse terminology and talk about the matrix subject (object) controlling not the non-existent embedded subject but controlling the finite verb phrase. We will also consider verbs whose objects control a noun phrase or an adjective.

As we will see below, it is more instructive to think of these controlled embedded verb phrases as embedded sentences with a subject (nominative) gap, since they must occur with the complementizer, just like full embedded clauses.

**Subject Control**

One important subject control verb in Serbo-Croatian is *htjeti* ‘want, will’, which can take a variety of complements and also participates in the subject controlled future tense construction. There are enclitic and non-clitic present tense forms of *htjeti*. We show the full paradigm in Table 4.2, but we return to its clitic forms in the next chapter and here only consider the full forms.

<table>
<thead>
<tr>
<th></th>
<th>sg</th>
<th>pl</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>hoću/ću</td>
<td>hoćemo/ćemo</td>
</tr>
<tr>
<td>2</td>
<td>hoćeš/češ</td>
<td>hoćete/čeete</td>
</tr>
<tr>
<td>3</td>
<td>hoće/če</td>
<td>hoće/če</td>
</tr>
</tbody>
</table>

Table 4.2: The verb *htjeti* ‘want, will’ paradigm.
With the meaning ‘want’, *htjeti* can take a noun phrase complement (a), a full embedded sentence (b) or an embedded sentence with a subject gap (c), the latter being an instance of subject control.

(164)  

a. Ana hoče pivo.
Ana_{NOM,f,sg,3} want_{sg,3} beer_{ACC,n,sg,3}
‘Ana wants a beer’
b. Ana hoče da Marko ode.
Ana_{NOM,f,sg,3} want_{sg,3} DA Marko_{NOM,m,sg,3} leave_{sg,3}
‘Ana wants Marko to leave’
c. Ana hoče da ode.
Ana_{NOM,f,sg,3} want_{sg,3} DA leave_{sg,3}
‘Ana wants to leave’

Future tense has the same structure as the subject control sentence (c) above. While typically, only the clitic forms participate in future tense formation, the full forms can have the future interpretation as well when used contrastively. The result is that the subject control versions of *htjeti* are ambiguous between the future meaning and the ordinary ‘want’ meaning. In the right context, the (c) sentence above can mean ‘Ana will leave’ and not ‘Ana wants to leave’. Below we show an example where the full form of *htjeti* under contrastive focus expresses future tense.

(165)  

a. A: Nikad nećeš da dobiješ taj posao!
   never not-will_{sg,2} DA get_{sg,2} that_{ACC,m,sg} job_{ACC,m,sg}
   ‘You’ll never get that job’
b. B: Sigurno hoću!
   definitely will_{sg,1}
   ‘I definitely will’
(166) a. Možda ti nećeš, ali ja sigurno hoću da maybe you\textit{NOM,sg,2} not-will\textit{sg,2}, but I\textit{NOM,sg,1} definitely will\textit{sg,1} DA
dobijem \textit{taj} posao.
\textit{get}_{sg,1} \textit{that}_{ACC,m,sg} \textit{job}_{ACC,m,sg}
‘Maybe you won’t, but I definitely will get that job’

Modal verbs such as \textit{smjeti} ‘may’, \textit{trebati} ‘need, should’, \textit{morati} ‘must’ and \textit{moći} ‘be able to’ are also subject control verbs, as well as \textit{znati} ‘know’ with the meaning
‘know how/ be able to do something’. Below are some examples.

(167) a. Ana \textit{mora} da ode.
\textit{Ana}_{NOM,f,sg,3} \textit{must}_{sg,3} DA \textit{leave}_{sg,3}
‘Ana must leave’

b. Mi \textit{smijemo} da odemo.
\textit{we}_{NOM,pl,1} \textit{may}_{pl,1} DA \textit{leave}_{pl,1}
‘We may leave’

c. Oni \textit{trebaju} da odu.
\textit{they}_{NOM,pl,3} \textit{should}_{pl,3} DA \textit{leave}_{pl,3}
‘They should/ need to leave’

d. Ona \textit{zna} da vozi.
\textit{she}_{NOM,f,sg,3} \textit{know}_{sg,3} DA \textit{drive}_{sg,3}
‘She can/ knows how to drive’

The set of subject control verbs in Serbo-Croatian includes also \textit{pokušati} ‘try’, \textit{obećati} ‘promise’ and \textit{voljeti} ‘love’, the latter in the sense of ‘like to do something’. Below are some examples.

\textit{Marko}_{NOM,m,sg,3} \textit{is}_{sg,3} \textit{promise}_{ppl,m,sg} DA \textit{will}_{sg,3} DA \textit{drive}_{sg,3}
‘Marko promised that he will/ would drive’

b. Marko \textit{voli} da vozi.
\textit{Marko}_{NOM,m,sg,3} \textit{love}_{sg,3} DA \textit{drive}_{sg,3}
‘Marko likes to drive’
Note that in subject control constructions the matrix subject has to agree in number and person with both the matrix and the embedded verb, as shown below:

   Marko\text{NOM, m, sg, 3} must\text{sg, 3} DA leave\text{sg, 3}
   ‘Marko must leave’

   b. * Marko mora da odemo\text{pl, 1}

   c. * Marko mora da odete\text{pl, 2}

   d. * Marko mora da odem\text{sg, 1}

Further, the gender information of the subject controller must be accessible as well. While gender marked verbal forms cannot occur in controlled verb phrases (i.e. past tense or conditional forms), the controlled verb phrase may be predicative and then the gender of the subject controller does matter. Consider the following examples:  

   Marko\text{NOM, m, sg, 3} must\text{sg, 3} DA is\text{sg, 3} cautious\text{NOM, m, sg}
   ‘Marko must be cautious’

   b. * Marko mora da bude oprezna\text{f, sg}

   c. Ana mora da bude oprezan.
   Ana\text{NOM, f, sg, 3} must\text{sg, 3} DA is\text{sg, 3} cautious\text{NOM, f, sg}
   ‘Ana must be cautious’

   d. * Ana mora da bude oprezan\text{m, sg}

These examples show that gender agreement between the matrix subject and a predicative adjective in a controlled verb phrase must be maintained. Therefore,

\textsuperscript{27}The form \textit{bude} glossed as ‘is’ is a perfective present form of \textit{biti} ‘be’, whereas \textit{je} which we’ve also been glossing as ‘is’ is an imperfective present form of the same verb. There are restrictions on the tense and aspect of the verb in certain embedded environments, including controlled verb phrases, the details of which are unfortunately beyond the scope of this thesis.
not only the number and person but also the gender of the subject controller matters.

In addition to finite embedded clauses with a subject gap, all subject control verbs can also take an infinitival verb phrase complement. In this case, there is no person or number agreement between the infinitive and the matrix subject, as the infinitive does not carry any agreement information. The examples below show some combinations of various subject controllers with controlled infinitival verb phrases.

(171)  a. Mi moramo otići.
        weNOM,pl,1 mustpl,1 leaveinf
        ‘We must leave’
   b. Marko hoče dobiti posao.
        MarkoNOM,m,sg,3 willsg,3 getinf jobACC,m,sg,3
        ‘Marko wants to/will get a job’
   c. Ja znam voziti bicikl.
        inOM,sg,1 knowsg,1 driveinf bicycleACC,m,sg,3
        ‘I can/know how to ride a bicycle’

However, if the controlled infinitival verb phrase contains a predicative adjective, gender of the subject controller does matter:

(172)  a. Marko mora biti oprezan.
        MarkoNOM,m,sg,3 mustsg,3 beinf cautiousNOM,m,sg
        ‘Marko must be cautious’
   b. * Marko mora biti opreznaNOM,f,sg
   c. Ana mora biti oprezna.
        AnaNOM,f,sg,3 mustsg,3 beinf,3 cautiousNOM,f,sg
        ‘Ana must be cautious’
   d. * Ana mora da bude oprezanNOM,m,sg
Further, the controlled infinitive may itself be a subject control verb, and if it embeds a finite controlled verb phrase, the person and number of the matrix subject controlled again matter:

(173)  a. Ana mora pokušati da dođe.
   Ana_{NOM,sg,f,3} must_{sg,3} try_{inf} DA come_{sg,3}
   ‘Ana must try to come’
  b. * Ana mora pokušati da dođete_{pl,2}

In sum, regardless of whether the controlled verb phrase is finite of infinitival, the grammar must keep track of number, gender and person requirements of that verb phrase so the agreement with the matrix subject controller can be established.

As for word order in subject control sentences, the generalizations are largely the same as with embedded clauses which we discussed earlier in the chapter. The complementizer must occur leftmost in the embedded clause, and the rest of the embedded clause material may reorder freely, depending on the embedded verb. In the matrix clause, the embedded clause must occur rightmost, with the exception of adverbials, while the rest of the matrix clause material may freely reorder.

Inside infinitival verb phrase complements, constituents can freely reorder, for example:

(174)  a. Oni moraju kupiti Ani poklon.
   they_{NOM,pl,3} must_{pl,3} buy_{inf} Ana_{DAT,f,sg,3} present_{ACC,m,sg,3}
   ‘They have to buy Ana a present’
  b. Oni moraju Ani kupiti poklon.
  c. Oni moraju poklon kupiti Ani.
d. Oni moraju Ani poklon kupiti.
e. etc.

In fact, in our judgment, the infinitive and its complements may freely reorder with respect to the matrix clause material, in contradistinction to embedded finite clauses:

(175) a. Oni moraju kupiti Ani poklon.
    they\NOM,pl,3 must\pl,3 buy\inf\ Ana\DAT,f,sg,3 present\ACC,m,sg,3
    ‘They have to buy Ana a present’
b. Oni Ani moraju kupiti poklon.
c. Poklon oni moraju Ani kupiti.
d. Kupiti poklon oni Ani moraju.
e. etc.

However, infinitival verb phrases also have the option of behaving in a more clause-like manner, occurring contiguously at the right edge of the clause, with the infinitive even hosting its own enclitics. We return to this pattern in Chapter 7, and here only consider free reordering of infinitival verb phrase constituents with main clause constituents, as in the example above.

**Object Control**

In contrast to subject control, an infinitival verb phrase is never a possible complement of an object control verb. Instead, such verbs have to combine with a finite embedded clause with a subject gap.
Object control verbs in Serbo-Croatian include *zamoliti* ‘ask, request’, *natjerati* ‘force, make’, *nagovoriti* ‘persuade’ and *ponuditi* ‘offer’. Below we show some examples of object control structures:

(176) a. Mi smo zamolili Anu da donese pivo.
    ‘We asked Ana to bring beer’

    b. Marko je nagovorio Anu da dođe u Ameriku.
    ‘Marko persuaded Ana to come to America’

Note that the embedded verb has to agree in person and number with the matrix object, which is one of the reasons we are keeping track of number on non-nominative noun phrases.

In addition to person and number, the gender of the object controller matters as well. First, as in the case of subject control, the controlled verb phrase may be predicative:

(177) a. Marko je zamolio Anu da bude pristojna.
    ‘Marko asked Ana to be polite’

b. * Marko je zamolio Anu da bude pristojan

Second, objects of some verbs control predicative complements are not verb phrases. If the controlled complement is an adjective, it has to agree with the matrix object in gender, as shown in the example below.

(178) a. Marko je smatrao Anu pristojnom.
‘Marko considered Ana polite’

b.  * Marko je smatralo Anu pristojnim

Therefore, as in the case of subject control, the object controlled constituent has to agree with the controller in number, person and gender.

As for word order in object control structures, the generalizations are the same as for subject control, whereby the complementizer must come first in the embedded clause and the embedded clause must remain contiguous and occur rightmost in the matrix clause. Other matrix clause material, including the object controller, may freely reorder with respect to the matrix verb and subject. The generalizations concerning the placement of adverbs are the same as well.

4.4.2 Analysis

Subject Control

We first analyze subject controlled finite verb phrases, then the infinitival ones. We analyze controlled finite verb phrases as embedded declarative clauses with a bound subject trace. For example, a controlled verb phrase *da ode* in the sentence *Ana hoće da ode* ‘Ana wants to leave’ is built up as follows. First, an appropriate subject trace is introduced and a declarative sentence is constructed:

(179)
Second, the complementizer *da* combines with the sentence, turning it into an embedded sentence, and then the subject trace is bound, i.e. the hypothesis is withdrawn:

(180)

\[ \vdash \lambda x.\exists y [(x \cdot y) \land w = (DAz \circ v)]; S_{m,6} \rightarrow S_{e,6}; \lambda p.q \]

\[ \vdash x; NP_{nom,sg,3}; \{x_1; \text{PER}(x \cdot ODEz); S_{m,6}; \text{(leave x)} \} \]

The conclusion of the proof above is the kind of sign that can be a complement of a subject control verb: an embedded sentence with a bound subject trace. Note that the relevant agreement features of the missing embedded subject are recorded in the tectogrammatical type. This will ensure that the matrix subject and the embedded predicate agree in number, person, and gender. We give the following lexical entry for the control verb *hoče* ‘wants’.

(181) \[ \vdash \lambda F_{xy}, \exists y [(F \cdot e \circ x) \land (\text{PER}(y \cdot HOCEz) \cdot y) \land w = y \cdot \text{toZ}(L \cdot x)]: (z \rightarrow z) \rightarrow z \rightarrow z; \prod_{G \cdot \text{Gr}}[(NP_{nom,sg,3}, e, S_{e,6}) \rightarrow NP_{nom,sg,3,6} \rightarrow S_{m,6}]; \lambda Fx.(\text{want}(F \cdot x) \cdot e \rightarrow p) \rightarrow e \rightarrow p \]

This finite verb requires a third person singular subject, so it is only compatible with verb phrase complements which also require such subjects. It is not, however, inflected for gender, which is why it has a dependent product type. Whatever the gender feature of its complement verb phrase, *hoče* will require a subject of that gender. To construct the sentence *Ana hoče da ode* ‘Ana wants to leave’
we need a feminine version of the verb. Combining that verb with the embedded clause with a subject gap, and then with the matrix subject Ana we get the following sign:

\[ (182) \quad \lambda_{w, x, y, v} (\text{PER}(\text{ODE}_z) \cdot v) \land x = (\text{DA}_z \circ v) \land (\text{PER}(\text{ANA}_z \circ \text{HOĆE}_z) \cdot y) \land \\
\quad w = y \circ \text{toZ}(L x) : Z, S_{m, o}, (\text{want (leave ana) ana}) : p \]

Semantically, we predict that this sentence expresses the proposition that Ana wants that she (Ana) leave. Syntactically, it is a main declarative clause. Phenogrammatically, it denotes a set of \( S \)-string each of which consists of some permutation of the matrix verb and the subject, followed by a length one \( S \)-string constructed out of the embedded sentence with a subject gap, just as in the case of ordinary embedded sentences. So we predict that this sentence can be pronounced exactly two ways, \( Ana \ hoće \ da \ ode \) and \( Hoće \ Ana \ da \ ode \).

Recall that infinitives are not inflected for gender, number and person. However, the infinitive could itself be a control verb, with a finite verb phrase complement, or a predicative verb, in which case the gender, number and person features matter to establish agreement with the matrix subject controller.

A subject control verb taking an infinitival complement cannot know in advance whether its complement is itself a control or a predicative verb (in which case agreement matters) or is not a control or predicative verb (in which case it doesn’t). For example, in \( Ana \ mora \ otići \) ‘Ana must leave’ the agreement features do not matter. But in \( Ana \ mora \ biti \ pristojna \) ‘Ana must be polite’, the adjective \( pristojna \) and the matrix subject must agree in number and gender, while in \( Ana \)
mora pokušati da dođe ‘Ana must try to come’, the embedded finite verb dođe and the matrix subject must agree in number and person. So a subject control verb like mora ‘must’, has to be prepared to deal with all these different complements and, if needed, make sure that the matrix subject is an appropriate controller.

For this reason, the agreement information must be recorded on the infinitival verb phrases in general, regardless of whether it’s a control verb or not. In other words, a control verb taking an infinitival complement has to err on the side of too much information, just in case its infinitival complement is a control or predicative verb.

Suppose we’re trying to construct a representation of the sentence Ana hoće voziti bicikl ‘Ana wants to ride a bike’ where the infinitive is not a control verb. We give the following lexical entry for the infinitive:

\[
\lambda_x. \text{VOZITI}_z \circ v : z \rightarrow z; \\
\prod_{x : \text{Gdr,n:Num,p:Prs}} [\text{NP}_{\text{acc,m,sg,3}} \circ \text{NP}_{\text{nom,g,n,p}} \circ \text{S}_{\text{inf,6}}]; \\
\lambda_{xy}. (\text{ride } x \ y) : e \rightarrow e \rightarrow p
\]

Here, inf is a term of type K. It is introduced especially to distinguish infinitival verb phrases from declarative and interrogative clauses. Note that there is a mismatch between the phenogrammatical typing on the one hand, and the semantic and tectogrammatical typing on the other. While semantically voziti needs two individual arguments, and tectogrammatically two noun phrase arguments (the object and the subject), phenogrammatically, it only needs one argument of type z. That is, phenogrammatically, voziti is expecting only one argument (the object
noun phrase). This is to ensure that no non-sentences consisting of a nominative noun phrase and an infinitival verb phrase are ever generated by the grammar. The tectogrammatical and semantic typing ensures that the appropriate agreement information is available and that the correct interpretation is generated. The phenogrammatical typing ensures that there are no signs of type $S_{inf,6}$ in the grammar, as they do not correspond to any actual sentences of Serbo-Croatian.

Below we give the lexical entry for *hoće* ‘wants’ which allows free reordering of the infinitive and its arguments with matrix clause constituents. The full range of word order possibilities in control will be explored in Chapter 7.

\[
\begin{align*}
\lambda_{xy} . \text{PER}(y \circ \text{HOĆE}_{z} \circ x) : z & \rightarrow z \rightarrow Z; \\
\Pi_{\text{Gdr}}[(\text{NP}_{\text{nom},g,sg,3} \rightarrow S_{inf,6}) \rightarrow \text{NP}_{\text{nom},g,sg,3} \rightarrow S_{m,6}]; \\
\lambda_{Fx} . (\text{want} (Fx) x) : (e \rightarrow p) \rightarrow e \rightarrow p
\end{align*}
\]

This lexical entry allows free reordering of all constituents in the infinitival verb phrase with matrix constituents, so that the sentence *Ana hoće voziti bicikl* ‘Ana wants to ride a bike’ is predicted to be pronounceable 24 different ways. The sign below represents this sentence in the grammar.

\[
\begin{align*}
\lambda_{xy} . \text{PER}(\text{ANA}_{z} \circ \text{HOĆE}_{z} \circ \text{VOZITI}_{z} \circ \text{BIČIKL}_{z}) : Z; S_{m,6}; \\
\text{exists(bicycle)} \lambda_{y} . (\text{want (ride y ana) ana}) : p
\end{align*}
\]

We’ve simply assumed that we have two different lexical entries for each subject control verb, one looking for a finite verb phrase, the other for an infinitival verb phrase complements. Below we state a non-logical rule [SF], which maps each subject control verb looking for an infinitival complement to its finite verb
phrase seeking counterpart. In the rule, $\phi$ is a metavariable over phenogrammatical terms of type $z \rightarrow z \rightarrow Z$, and $F$ is a variable of type $z \rightarrow Z$. For typographical reasons, we suppress the semantic component, as it remains unchanged.

(186)

\[
\vdash \phi; \Pi_{g:Gdr}[(\text{NP}_{\text{nom},g,n,p} \rightarrow \text{S}_{\text{inf},6}) \rightarrow \text{NP}_{\text{nom},g,n,p} \rightarrow \text{S}_{\text{e},6}]
\]

\[\vdash \lambda_{Fxy}.\exists_{vw}[(\phi \in S \times w) \land (F \in S v) \land y = w \circ \text{toZ}(L v)];
\]

\[
\Pi_{g:Gdr}[(\text{NP}_{\text{nom},g,n,p} \rightarrow \text{S}_{\text{e},6}) \rightarrow \text{NP}_{\text{nom},g,n,p} \rightarrow \text{S}_{\text{e},6}]
\]

This main impact of this rule is in the phenogrammatical term transformation, since instead of seeking a term of type $z$ (infinitival verb phrase), the subject control verb now has to combine with a term of type $z \rightarrow Z$ (embedded clause with a subject gap). Tectogrammatically, the subject control verb’s argument type has changed from $\text{NP}_{\text{nom},g,n,p} \rightarrow \text{S}_{\text{inf},6}$ to $\text{NP}_{\text{nom},g,n,p} \rightarrow \text{S}_{\text{e},6}$.

Object Control

Object control verbs cannot combine with infinitival verb phrases but instead only with embedded clauses with a subject gap. Suppose we are trying to construct a representation of the sentence Marko nagovara Anu da dođe ‘Marko is persuading Ana to come’. The embedded sentence with the subject gap is constructed in exactly the same way as in the case of subject control. We give the following lexical entry for nagovara ‘persuades’:

(187) \[\vdash \lambda_{z_{FWV},e_{xy}}[(\text{PER}(v \circ \text{NAGOVARA}_z \circ z) y) \land w = y \circ \text{toZ}(L x)];
\]

\[z \rightarrow (z \rightarrow Z) \rightarrow z \rightarrow Z;
\]

\[
\Pi_{g,g':Gdr,n:Num,p:Prs}[\text{NP}_{\text{acc},g,n,p} \rightarrow \text{NP}_{\text{nom},g,n,p} \rightarrow \text{S}_{\text{e},6}) \rightarrow \text{NP}_{\text{nom},g',sg,3} \rightarrow \text{S}_{\text{m},6}];\lambda_{yFx}.(\text{persuade } y (F y) x) : e \rightarrow (e \rightarrow p) \rightarrow e \rightarrow p
\]
Phenogrammatically, this object control verb works much like a subject control verb: it allows free reordering of itself, its subject and its accusative object, and forces the embedded clause with a subject gap, transformed into a length one S-string, to occur on the right edge of the matrix sentence. Semantically, we analyze the verb as expressing a relation between an individual, a proposition and another individual.

Tectogrammatically, \textit{nahovara} ‘persuades’ requires a nominative 3rd person subject of any gender. The \(g\)’ variable stands for that ‘missing’ Gdr parameter of its subject. In addition, its object must be an accusative noun phrase, but its Gdr, Num and Prs parameters are not specified. Whatever those parameters turn out to be, however, \textit{nahovara} will then require the controlled verb phrase to be missing a nominative noun phrase with the same Gdr, Num and Prs parameters as its object’s. For example, the sentence Marko \textit{nahovara} Ana da dođe ‘Marko is persuading Ana to come’ requires the following tectogrammatical version of \textit{nahovara}, where the object and the ‘missing’ subject must be 3rd person singular feminine noun phrases, and the matrix subject must be masculine:

\begin{verbatim}
(188) \(\lambda_{xFw.}\exists_{xy}[\lambda(FeSx) \land (\text{PER}(v \circ \text{NAGOVARA}_z \circ z) y) \land w = y \circ \text{toZ}(L x): z \rightarrow (z \rightarrow z) \rightarrow z \rightarrow z; \ NP_{\text{acc}, f, sg, 3} \rightarrow (\NP_{\text{nom}, f, sg, 3} \rightarrow S_{e, 6}) \rightarrow \NP_{\text{nom}, m, sg, 3} \rightarrow S_{m, 6}; \ \lambda_{yFx.}(\text{persuade } y (F y) x): e \rightarrow (e \rightarrow p) \rightarrow e \rightarrow p]
\end{verbatim}

The whole sentence is represented in the grammar as follows:

\begin{verbatim}
(189) \(\lambda_{w.}\exists_{xyv}[(\text{PER}(\text{DOD}_z \circ v) v) \land x = (DA_z \circ v) \land (\text{PER}(\text{MARKO}_z \circ \text{NAGOVARA}_z \circ \text{ANU}_z) y) \land w = y \circ \text{toZ}(L x): z; S_{m, 6}; (\text{persuade } \text{ana} (\text{come } \text{ana}) \text{marko}) : p]
\end{verbatim}
Tectogrammatically, this is a declarative main clause. Semantically, it expresses the expected proposition. Phenogrammatically, the sign denotes a set of $S$-strings consisting of some permutation of the verb, the subject and the object, followed by the embedded clause with the nominative gap which has been transformed into a length one string of languages. The grammar predicts that this sentence can be pronounced 6 different ways, which is correct.

Finally, we analyze verbs whose object controls a predicative adjective or a predicative noun phrase. We focus on controlled predicative adjectives, as in *Ana smatra Marka pristojnim* ‘Ana considers Marko polite’. The adjective must occur in instrumental case, but it agrees with the accusative object in number and gender.

The verb *smatrati* ‘consider’ can also combine with a finite clause, as in *Ana smatra da je Marko pristojan*, which expresses the same meaning as its object control counterpart. Therefore, we analyze the meaning of this verb as a relation between an individual and a proposition.

Recall that adjectives which are complements of the copula must be nominative. It was to these kinds of adjectives that we assigned the tectogrammatical result type $Prd_a$. So that we don’t get a case mismatch between a predicative adjective and the subject in such predicative structures with the copula, we will analyze this instrumental adjective that occurs as an argument of *smatrati* ‘consider’ as an ordinary attributive instrumental adjective. So, in the sentence *Ana
smatra Marka pristojnim ‘Ana considers Marko polite’ the adjective is associated with the following lexical entry:28

\[
(190) \vdash \lambda_{v}.\text{PRISTOJNIM}_z \circ v : z \rightarrow z; N_{\text{inst},m,sg} \rightarrow N_{\text{inst},m,sg}; \text{polite} : e \rightarrow p
\]

Below is the lexical entry for the verb smatra.

\[
(191) \vdash \lambda_{Fw}.\text{PER}(w \circ \text{SMATRA}_z \circ v \circ (F \circ e_s)) : z \rightarrow (z \rightarrow z) \rightarrow z \rightarrow z;
\]

\[
\Pi_{g,g'}:Gdr,n:Num,p:Prs[NP_{\text{acc},g,n,p} \rightarrow (N_{\text{inst},g,n} \rightarrow N_{\text{inst},g,n}) \rightarrow NP_{\text{nom},g',sg,3} \rightarrow S_{m,6}] ; \lambda_{xPy}.(\text{consider}(P \times y)) : e \rightarrow (e \rightarrow p) \rightarrow e \rightarrow p
\]

This verb requires an accusative object, an instrumental adjective and a nominative singular 3rd person subject. Its tectogrammatical type ensures that its object and the adjective that its object controls have the same gender and number. The phenogrammatical term of this verb allows free reordering of all the constituents. The sentence Ana smatra Marka pristojnim ‘Ana considers Marko polite’ is represented in the grammar as follows.

\[
(192) \vdash \text{PER}(\text{ANA}_z \circ \text{SMATRA}_z \circ \text{MARKA}_z \circ \text{PRISTOJNIM}_z) : z; S_{m,6};
\]

(consider(polite marko) ana) : p

4.5 Conclusion

In this chapter, we analyzed embedded declarative clauses, and predicative and control structures. We introduced a new value of the K parameter for infinitival phrases. Following the observation that all and only subject control verbs can combine with infinitival verb phrases in addition to finite verb phrases, we gave a

28 Here we consider only the most permissive version of the attributive adjective, of all the versions entertained in Chapter 3.
rule [FS] which takes a subject control verb subcategorized for an infinitival complement and outputs its counterpart which takes a finite complement.

Finally, we incorporated predicative complements into the grammar, by introducing a family of predicative types. Doing so enabled us to give a single lexical entry for a given form of the copula, which can take as its complement whichever kind of predicative phrase. We also explored the connection between predicative complements and postnominal modifiers and, where appropriate, devised non-logical rules which establish the connection between the two incarnations of certain classes of expressions.

We analyzed controlled finite verb phrases, in both subject and object control structures, as embedded declarative clauses with a subject gap. However, nothing in our grammar guarantees that the missing subject in such an embedded declarative clause be the closest subject. If the embedded clause itself contains an embedded clause, and the downstairs embedded clause is missing its subject, our grammar in fact allows for this sort of expression to be an argument of a control verb. In other words, our grammar currently allows long distance control, which is in reality impossible.

The purpose of this chapter was to sketch a preliminary theory of embedding, control and predication, so that we can give a theory of encliticization in Serbo-Croatian because embedded clauses are domains for clitic placement, and because all clitic verbs in Serbo-Croatian are either subject control or predicative verbs. We
will refine our theory of embedding, control and predication in Chapter 7, and constrain our grammar to prevent long distance control. In that chapter, we will also address different word order possibilities concerning participles of control verbs and controlled infinitival complements, as well as the difference between control I-verbs and control S-verbs with respect to clitic climbing out of their controlled complement.
Chapter 5: Enclitics

5.1 Introduction

Serbo-Croatian enclitics have attracted a lot of attention in the literature over the decades (see Browne (1974); Halpern (1995); Schütze (1994); Progovac (1996, 2005); Radanović-Kocić (1996); Penn (1999a); Bošković (2001, 2004) inter alia). In contrast to proclitics (such as prepositions which we examined in the previous chapter), enclitics attach to a phonological word to their left, and together with their host form a new phonological word.

The enclitics, which include pronouns, reflexives and auxiliaries, all cluster together in a certain not entirely predictable order. The resulting enclitic cluster is extremely limited in terms of possible placement within the sentence. The ordering within the clitic cluster and the placement of the cluster in the sentence are the biggest challenges when it comes to constructing an analysis. In this chapter we first examine the empirical facts concerning the order and the placement of the clitic cluster, and then present our analysis.
5.2 Data

5.2.1 Order

Enclitics include dative, genitive and accusative pronominal clitics, the accusative reflexive pronoun and the inherent reflexive se, the interrogative complementizer li which occurs in polar interrogatives, and several different kinds of auxiliaries. The auxiliaries include the clitic forms of htjeti ‘want, will’ used in future tense formation, and two different sets of clitics of the verb biti ‘be’. The imperfective present tense clitics of biti ‘be’ are used as copula and for past tense formation, while its aorist clitics are used to construct conditional forms of verbs. Of all these enclitics, only li and the aorist of biti do not have corresponding non-clitic forms.

The enclitics cluster together and are strictly ordered as follows:

\[ \text{li} < \text{auxiliaries} - \{\text{je}\} < \text{dative} < \text{accusative} < \text{genitive} < \text{se} < \text{je} \]

Here, je ‘is’, is the 3rd person singular present tense clitic of biti ‘be’ and occurs in a different slot in the clitic cluster than all other auxiliary clitics. Any violations of this order in the clitic cluster, or any attempt to make the cluster discontinuous, result in sharp ungrammaticality (Browne, 1974; Progovac, 1996, 2005; Radanović-Kocić, 1996; Franks and King, 2000).

While at first glance it may seem that the clitics are ordered by their syntactic categories, this cannot be maintained without much stipulation. First, the reflexive se is accusative but occurs in a different slot than other accusative clitics. In
addition to *se* being a possible argument of a verb needing an accusative object, it is sometimes simply a part of a lexical verb:

(193)  

a. Ana \( \text{NOM} \) vidi Maju \( \text{ACC} \) u ogledalu.  
\hspace{1cm} Ana\( \text{NOM} \) sees Maja\( \text{ACC} \) in mirror  
\hspace{1cm} ‘Ana sees Maja in the mirror’

b. Ana \( \text{NOM} \) se vidi u ogledalu.  
\hspace{1cm} Ana\( \text{NOM} \) REFL\( \text{ACC} \) sees in mirror  
\hspace{1cm} ‘Ana sees herself in the mirror’

(194)  

a. Ana \( \text{NOM} \) se bavi lingvistikom.  
\hspace{1cm} Ana\( \text{NOM} \) does (professionally) linguistics\( \text{INST} \)  
\hspace{1cm} ‘Ana does linguistics (professionally)’

b. Ana \( \text{NOM} \) se boji Maje.  
\hspace{1cm} Ana\( \text{NOM} \) is-afraid-of Maja\( \text{GEN} \)  
\hspace{1cm} ‘Ana is afraid of Maja’

While in (193b) *se* simply occurs as an argument of *vidi* ‘sees’, in (194a) and (194b) it is just a part of the verb and cannot be left out of the sentence or replaced by some other accusative expression, as if the verb and *se* constituted a phrasal idiom. I will call the *se* that occurs in (194) an *inherent* reflexive. Both kinds of *se*, however, occur in the same slot in the clitic cluster, although they are tectogrammatically different.

Also note that *je*, 3rd person singular present tense *biti* ‘be’ (which functions as a copula and as an auxiliary in periphrastic past tense) is idiosyncratic in that it occurs at the end of the clitic cluster and not where all the other auxiliaries occur.

In sum, we will maintain that the order of enclitics in the cluster must be stipulated and cannot be derived from any general syntactic properties of the clitics. This is in contrast to popular MGG views (Bošković, 2001, 2004; Progovac, 1996, 158)
2005), which consider the order in the cluster as determined by underlying syntactic principles, and therefore either require special mechanisms for placing je, as well as the two kinds of reflexive se, or do not account for their placement at all.

5.2.2 Placement

The even trickier issue is the placement of the enclitic cluster within the clause. Below we argue that its placement cannot be accounted for either purely syntactically or purely prosodically. This fact alone undermines the MGG analyses that I’m familiar with, since they all essentially assume either purely syntactically or purely prosodically conditioned placement.

Radanović-Kocić (1996) attempts to give a prosodic account of enclitic placement and states that the clitics need to come right after the first phonological phrase in their intonational phrase (Bošković (2001, 2004) accepts this generalization and makes use of it in his analysis). She however offers no definition of what constitutes a phonological phrase in Serbo-Croatian other than “one or more phonological words” (p. 441). She suggests that degemination may be prohibited across a phonological phrase boundary, but has no phonetic evidence to substantiate this claim.

A phonological phrase is supposed to be a prosodic constituent larger than a phonological word but smaller than an intonational phrase. It seems that Radanović-Kocić (1996) wants syntactic constituents in general to correspond to phonological
phrases, and that a focused (presumably contrastively focused, given her examples) word within a constituent may constitute its own phonological phrase.

Note that both of the following are possible:

(195)  
\begin{align*}
\text{a. } & \text{Moja je sestra došla.} \\
& \text{my} _{\text{NOM}} \text{is sister} _{\text{NOM}} \text{arrived} \\
& \text{`My sister arrived’}
\end{align*}

\begin{align*}
\text{b. } & \text{Moja sestra je došla.}
\end{align*}

Example (195a) shows the so-called 1W clitic cluster placement (after the first word), and example (195b) 1C clitic cluster placement (after the first constituent).

According to Radanović-Kocić (1996), \textit{moja sestra} in (186b) forms a phonological phrase and so the example obeys her generalization. (195a), she claims, is only possible if the possessive is focused and forms its own phonological phrase. This allows her to maintain that in (195a) the clitic does come after the first phonological phrase in its intonational phrase, since the possessive is stipulated to constitute its own phonological phrase. However, the production study in Yu (2009) directly contradicts Radanović-Kocić (1996) in that no evidence for a prosodic break after \textit{moja} in sentences like (195a) was found. It also doesn’t seem to be the case that the clitic host in 1W examples has to be contrastively or otherwise focused.

Further, Godjevac (2000) in her extensive study of Serbo-Croatian intonation found no phonetic evidence for an intermediate prosodic constituent, even when the utterance is very long. While in very long utterances there is a periodic pitch reset (an effect she calls pleating), the points at which the pitch is reset do not seem
to reliably correspond to syntactic constituent boundaries, and she analyzes such utterances as consisting of single intonational phrases.

So the idea that there are phonological phrases in Serbo-Croatian doesn’t appear to be empirically grounded. Moreover, even if phonological phrases turn out to be in the Serbo-Croatian prosodic inventory, there doesn’t seem to be a prosodic boundary where Radanović-Kocić (1996)’s theory predicts it to occur.

Clearly, a prosodic generalization on clitic cluster placement cannot be stated in terms of phonological words alone since the cluster can attach to the right edge of the first or the second phonological word in an intonational phrase (as we saw above), or be delayed by many phonological words, e.g.:

(196)  

a. Roditelji uspešnih studenata **su se razišli.**  
parents\textit{NOM} successful\textit{GEN} students\textit{GEN} are dispersed  
‘The parents of successful students dispersed’  
[adapted from Progovac (1996)]

b. Ona moja sestra koja je u Sarajevu **vas se sjeća.**  
that\textit{NOM} my\textit{NOM} sister\textit{NOM} who\textit{NOM} is in Sarajevo you\textit{GEN}  
remembers  
‘That sister of mine who is in Sarajevo remembers you’  
[adapted from Radanović-Kocić (1996)]

So, the generalization concerning the placement of the clitic cluster cannot be stated over prosodic constituents. However, their placement is clearly sensitive to prosodic factors since they cannot occur immediately to the right of proclitics, or be the first elements in an intonational phrase, because they need a host that is a phonological word.

161
Progovac (1996, 2005) attempts to give a syntactic generalization concerning the placement of the clitic cluster. She maintains that only expressions that are independently ‘moveable’ around the clause can host clitics (so does Bošković (2004), see p.12). So, for example, Progovac would claim that an attributive adjective can host the clitics because it can in general be detached from the rest of the noun phrase, but since postnominal modifiers cannot be detached from the noun they are modifying, clitics cannot come between a noun and its postnominal modifier.

We believe this generalization to be largely correct. However, in some cases enclitics can in fact split constituents that otherwise must remain contiguous. While the examples may be somewhat degraded, it is possible to place the clitic cluster between a noun and a post-nominal modifier (Browne (1974); Halpern (1995), Aaron Halpern p.c. to Progovac (2005)).

(197) ? Sestra će moje prijateljice doći.
   sisterNOM will myGEN friendGEN arriveINF
   ‘My sister’s friend will arrive’

The clitics can also occur right after the first conjunct, although co-ordinate structures in general cannot appear discontinuously in a clause in Serbo-Croatian, as shown below:

(198) a. ? Knjige ću i teke sutra kupiti.
    books will and notebooks tomorrow buyINF
    ‘I’ll buy books and notebooks tomorrow’
 b. Knjige i teke ću sutra kupiti.
 c. * Knjige sutra i teke ću kupiti.
In our analysis, we will first account for Progovac’s generalization, namely that the set of potential clitic hosts coincides with the set of ‘moveable’ expressions; i.e. expressions that can freely reorder with other clausal constituents. These expressions may be one phonological word long (1W placement) or many phonological words long (1C placement). In our grammar, they will turn out to correspond exactly to phenogrammatical terms that are length one $s$-strings.

Then, we will add a rule that allows clitics to be placed in such a way that they break up otherwise unbreakable constituents, e.g. between a noun and a postnominal modifier. Depending on how permissive one’s judgments are, this rule can be added to the grammar or not.

Either way, within our theory, the generalization concerning the clitic cluster placement can be stated as follows: the cluster encliticizes to the last phonological word ‘inside’ the first length one string of languages in the phenogrammatical denotation of a sentence.

5.3 Analysis

5.3.1 Clitics in Categorial Grammar

While both Kraak (1998) and Morrill and Gavarró (1992) give categorial analyses of Romance clitics (French and Catalan respectively), it is not at all obvious how such analyses could be extended to Serbo-Croatian enclitics. Here we briefly discuss the gist of their approach and point out the difficulties posed by
Serbo-Croatian clitics. We are not claiming that an adequate analysis of Serbo-
Croatian clitics cannot be given in a mainstream multimodal categorial grammar
which does not have a designated phenogrammatical component, such as those
assumed by Kraak (1998) and Morrill and Gavarró (1992). We are simply claim-
ing that we do not see a straightforward extension of their approach to the case of
Serbo-Croatian enclitics.

Romance clitics are significantly different from Serbo-Croatian clitics, although
in both cases a clitic cluster is formed. First, Romance clitic clusters consist of pro-
clitics, Serbo-Croatian of enclitics. Second, Romance clitic clusters always procliti-
cize onto the finite verb. In contrast, Serbo-Croatian enclitics can be hosted by a
variety of constituents. Third, the position of the Romance clitics in a clause is de-
termined by the finite verb onto which they are procliticized. In Serbo-Croatian,
however, the clitic cluster has to occur 2nd, regardless what is 1st or 3rd in the
clause; so they are truly 2P clitics and cannot be anchored to a specific kind of
constituent.

For both Kraak (1998) and Morrill and Gavarró (1992) essentially analyze cli-
tics as combining with verb phrases missing them, and they use a natural number
parameter on syntactic types to enforce their ordering. We borrow from them both
the idea to treat clitics as functors over expressions missing them and to use the
natural number parameter to enforce their ordering. However, we cannot adopt
their analysis in general because we cannot anchor Serbo-Croatian clitics to verbs,
or verb phrases, or any specific kind of syntactic constituent at all. Further, recall that Serbo-Croatian clitics may break up otherwise unbreakable constituents. Worse, some of Serbo-Croatian clitics are verbs themselves.

In sum, we think that Serbo-Croatian clitics strongly call for an architecture which allows explicit representation of the phenogammatical structure. A generalization concerning the placement of the clitic cluster in Serbo-Croatian clauses cannot be easily stated at the syntactic level of representation. Even if it could, we would like to suggest that it would be conceptually misleading to do so. Serbo-Croatian clitics are, in our opinion, truly situated at the syntax/word order interface, sensitive to structure at both those levels of representation.

In our theory, we will be able to state the generalization concerning the placement of the clitic cluster entirely at the phenogammatical level of representation. However, phenogammarm and tectogrammarm have to work in parallel to ensure that we derive the appropriate phenogammatical representation of the sentence so that the clitics can be placed correctly. Therefore, regardless of whether it is in principle impossible to give a theory of encliticization in Serbo-Croatian in mainstream categorial grammar or not, we think that our theory is conceptually satisfying because it treats 2P clitics as an interface phenomenon, sensitive to both phenogammarm and tectogrammarm.
5.3.2 Preliminaries

First, we make a case for the enclitic interrogative complementizer *li* which occurs in polar questions not being treated as an independent lexical item. *li* always comes encliticized either onto a finite verb or the complementizer *da*. The verb+*li* or *da+li* must occur question-initially. So we can give a lexical entry for *da li* which forms polar interrogatives out of finite clauses and requires *da li* to occur on the left edge of the question. Similarly, we give interrogative versions of finite verbs onto which *li* is already encliticized (see Chapter 6).

We are going to treat enclitics as functions over sentences with bound traces, similar to how we analyzed quantificational noun phrases. So, a clitic will target a sentence missing an expression that’s exactly like that clitic, tectogrammatically and semantically, which is essentially standard in categorial grammar (see Nishida (1996), Morrill and Gavarró (1992) and Kraak (1998)).

Table 5.1 shows the required order of clitics in the enclitic cluster (excluding *li* as explained at the beginning of this section). We’re going to enforce the order in the cluster via tectogrammatical types in the *S* family. As discussed in Chapter 3

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUX except <em>je</em></td>
<td>dative clitics</td>
<td>accusative clitics</td>
<td>genitive clitics</td>
<td><em>se</em></td>
<td><em>je</em></td>
</tr>
</tbody>
</table>

Table 5.1: The order of enclitics in the cluster.
and 4, the S family of types is indexed by natural numbers and $\mathbf{K} = \{m, e, q, \text{inf} \}$ parameter reflects the kind of sentence we’re dealing with—a main declarative clause, an embedded declarative clause, a question (main or embedded), or an infinitival clause.

Intuitively, for each sign whose tecto type is $S_{k,n}$, $n$ refers to the maximal number of clitics that could be placed in that sentence. For example, if a sentence is associated with the type $S_{k,6}$, it means that no clitics have been placed inside that sentence, and up to six clitics could still occur inside of it. If a sentence has the type $S_{k,0}$ it means that no more clitics can be placed inside of it. Recall that finite verbs in general build sentences of type $S_{m,6}$ since they can’t know in advance how many clitics may wind up placed inside of that sentence at the end.

The enclitics combine with ‘gappy’ sentences in the left-to-right order of their appearance in the clitic cluster. As they do so, they systematically reduce the number parameter of the sentence type, thereby preventing clitics that must occur to their left to apply after them. This way the order in the cluster is enforced.

The idea to use natural number parameters on sentence types to order clitics comes from Morrill and Gavarró (1992), however, our implementation is somewhat different in that the number parameter order reverses the order of various clitic slots in the enclitic cluster. Table 5.2 lists, for each clitic slot, the type of sentence that the clitic can combine with, and the type of sentence that results after a given clitic combines with that sentence.
<table>
<thead>
<tr>
<th>ARGUMENT’S RESULT TYPE</th>
<th>$S_{n&gt;5}$</th>
<th>$S_{n&gt;4}$</th>
<th>$S_{n&gt;3}$</th>
<th>$S_{n&gt;2}$</th>
<th>$S_{n&gt;1}$</th>
<th>$S_{n&gt;0}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESULT TYPE</td>
<td>$S_5$</td>
<td>$S_4$</td>
<td>$S_3$</td>
<td>$S_2$</td>
<td>$S_1$</td>
<td>$S_0$</td>
</tr>
<tr>
<td>CLITICS</td>
<td>AUX except je</td>
<td>dative clitics</td>
<td>accusative clitics</td>
<td>genitive clitics</td>
<td>se</td>
<td>je</td>
</tr>
</tbody>
</table>

Table 5.2: Tectogrammatical types of enclitics.

So, for example, a dative clitic can combine with a sentence whose number parameter is greater than 4, that is, inside of which at least 5 more enclitics can be placed. It outputs a sentence whose number parameter is exactly 4, meaning that after the dative clitic is placed in the sentence, exactly 4 more enclitics can also occur in that sentence.

Next, we introduce a couple of abbreviations to streamline the presentation of the lexical entries for clitics. Recall the function $\vdash \#_{ec} : p \rightarrow c \rightarrow p$, which encliticizes a clitic onto its phonological word host. We define a higher level function out of $\#_{ec}$ as follows:

\[
(199) \quad \begin{align*}
\text{a. } & \vdash +_{ec} : c_p \rightarrow c \rightarrow s \\
\text{b. } & +_{ec} =_{\text{def}} \lambda c.a.(j\ snc_p(tsr_p\ a)(lst\ a)\#_{ec}c)
\end{align*}
\]

In practice, we omit the subscripts on $+$ and $\#$. Intuitively, this function ‘looks inside’ a $p$-string and finds a suitable host for the clitic. In particular, it takes a non-null $p$-string, finds the right-most phonological word in that non-null string by applying $lst_p$ to it, and attaches the clitic to that phonological word. It then
assembles a non-null string, with the clitic attached to its last word and, via the right projection \( j \), constructs a \( \mathbf{p} \)-string out of it. Because \( \mathbf{lst}_p \) is defined on non-null \( \mathbf{p} \)-strings, the enclitics are guaranteed to find an actual host, and cannot be stranded.

This function now ensures that the clitic finds a phonological word host. However, we still need to make sure that the clitic’s host occurs in the right position in the clause. Given that in our grammar clauses are of phenogrammatical type \( \mathbf{z} \), and clitic attachment occurs at the level of phonological words, the clitic has to ‘look inside’ a very high level phenogrammatical object, and somehow ‘zoom in’ until it finds a suitable host. The following function, called \( \text{wac} \) for Wackernagel, accomplishes just that.

\[
\text{(200) a. } \Downarrow \text{wac} : \mathbf{c} \to (\mathbf{z} \to \mathbf{z}) \to \mathbf{z}
\]

\[
\text{b. } \text{wac} =_\text{def} \lambda_{cFw} \exists_{vt} \left[ (F(i \in \mathbf{s})(j \mathbf{v}')) \land (\mathbf{fst}_\mathbf{s} \mathbf{v}(j \mathbf{t})) \land w = (\text{toZ} \lambda_{s,s} = t + c) \circ (\mathbf{rst}_\mathbf{s} \mathbf{v}) \right]
\]

Intuitively, here is how \( \text{wac} \) works. First, it takes a clitic (\( \mathbf{c} \)), and a ‘gappy’ sentence (\( F \)) as its arguments and gets rid of the gap. Then, it looks at each \( \mathbf{S} \)-string in the resulting \( \mathbf{S} \)-language, finds its initial length one \( \mathbf{S} \)-string, and looks for the \( \mathbf{p} \)-string out of which this initial length one \( \mathbf{S} \)-string is constructed:
\( (F(e_S ) (j v)) \) the gap in the sentence is plugged by \( e_S \);
\( v \) is a non-null \( S \)-string in the resulting \( S \)-language;

\( (\text{fst}_S v (j t)) \quad t \) is a non-null \( p \)-string in the \( p \)-language
out of which the initial length one \( S \)-string \( v \) is constructed.

Then it attaches the clitic to \( t \) and ultimately outputs a set of \( S \)-strings, each of
which \((w)\) is constructed as follows:

\[ w = \]
\[ (\text{toZ} \lambda_s . s = t + c) \quad \text{the clitic is attached to } t, \text{ and the result} \]
\[ \text{is turned into a length one } S \text{-string} \]
\[ \circ(\text{rst}_S v) \quad \text{...which is concatenated with the rest of } v. \]

Recall that a sentence denotes a set of \( S \)-strings. For each of those \( S \)-strings, an
enclitic attaches to the last phonological word in the \( p \)-string which the initial
length one \( S \)-string is constructed out of. So, correctly placing enclitics in a clause
for us amounts to specifying what can count as the initial length one \( S \)-string in the
phenogrammatical denotation of a sentence. Within our theory, the Wackernagel
position is analyzed as the right edge of the first \( S \)-string.

Having presented these formal definitions, as before, we will stop notating
canonical injections for typographical reasons.

### 5.3.3 Pronominal Clitics

Suppose we’re trying to generate the sentence \( \text{Ana mu daje knjigu} \) ‘Ana gives
him a book’, where \( \text{mu} \) ‘him’ is a dative enclitic. The non-clitic version of the same
pronoun is *njemu* ‘him’. Below we give the lexical entry for the non-clitic and the clitic version of this pronoun.\(^\text{29}\)

\[ \begin{align*}
(201) & \vdash \text{NJEMU}_z : z; \text{NP}_{\text{dat, m, sg, 3}}; x : e \\
(202) & \vdash (\text{wac } \mu) : (z \rightarrow z) \rightarrow z; (\text{NP}_{\text{dat, m, sg, 3}} \rightarrow \text{S}_{m,n>4}) \rightarrow \text{S}_{m,4}; \\
& \lambda_G(G \times) : (e \rightarrow p) \rightarrow p
\end{align*} \]

Tectogrammatically, the dative clitic is looking for a sentence in which at least 5 more clitics can be placed, and which is missing a noun phrase of exactly the same type as the non-clitic version of this dative pronoun, \(\text{NP}_{\text{dat, m, sg, 3}}\).

Phenogrammatically, the clitic works as explained in the previous section: It first feeds the ‘gappy’ sentence the empty string, \(e_S\), which results in a set of \(S\)-strings, corresponding to the possible pronunciations of the sentence without the dative argument. For each of those \(S\)-strings, the clitic ‘looks’ inside its initial length one \(S\)-string, finds the last phonological word that it’s built out of and encliticizes onto it.

Below we show how to construct the sentence *Ana mu daje knjigu* ‘Ana gives him a book’. First, we give the lexical entry for the verb.

\[ \begin{align*}
(203) & \vdash \lambda_{\text{vxw}, \text{PER}}(\omega \circ \text{DAJE}_z \circ v \circ x) : z \rightarrow z \rightarrow z \rightarrow z; \\
& \text{NP}_{\text{acc, f, sg, 3}} \rightarrow \text{NP}_{\text{dat, m, sg, 3}} \rightarrow \text{NP}_{\text{nom, f, sg, 3}} \rightarrow \text{S}_{m,6}; \\
& \text{give} : e \rightarrow e \rightarrow e \rightarrow p
\end{align*} \]

The accusative object is a bare noun which has to undergo \([\text{NC}]\) and \([\text{Quant}]\) to become a full-fledged quantificational noun phrase, which will then scope over a

\(^{29}\text{We are semantically representing pronominal clitics as variables of type } e. \text{ We recognize that this is inadequate, however, the details about pronominal meanings and binding are unfortunately beyond the scope of this dissertation.}\)
sentence with the appropriate gap. So we first introduce an accusative trace and combine it with the verb.

Similarly, since the dative argument is a clitic, we introduce a dative trace as well and proceed with the sentence construction. Finally, the verb can combine with the subject, but both the dative and the accusative trace are kept track of in the context. The result is the following sign.

\[
\begin{align*}
(204) \quad & x : z; NP_{dat,m,sg,3}; x : e, \\
& y : z; NP_{acc,f,sg,3}; y : e \vdash \text{PER}(\text{ANA}_{z} \circ \text{DAJE}_{x} \circ y \circ x) : z; S_{m,6}; (\text{give } y \times \text{ana}) : p
\end{align*}
\]

We can bind the traces in either order. However, if we first bind the dative trace and place the clitic in the sentence, we won’t be able to then use the quantificational version of *knjigu* ‘a book’. This is because quantificational noun phrases are looking for ‘gappy’ sentences whose number parameter is 6. But after the dative clitic combines with the sentence, its number parameter is reduced to 4. Therefore, the tectogrammatical typing requires us to first bind the accusative trace and introduce *knjigu* ‘a book’ into the derivation, and then deal with the clitic. In general, quantificational noun phrases will have to be dealt with before any clitics, since those quantificational noun phrases may wind up being clitic hosts.

After we bind the accusative trace and scope *knjigu* ‘a book’, we get the following sign:

\[
\begin{align*}
(205) \quad & x : z; NP_{dat,m,sg,3}; x : e \vdash \text{PER}(\text{ANA}_{z} \circ \text{DAJE}_{x} \circ \text{KNJIGU}_{z} \circ x) : z; S_{m,6}; \\
& (\text{exists book})(\lambda y. (\text{give } y \times \text{ana})) : p
\end{align*}
\]
The dative trace is still in the context. Now we can bind it and finally introduce the dative clitic. The result is the following sign, shown in abbreviated form with \textit{wac}.

\begin{equation}
\vdash (\text{wac }\mu)(\lambda_x.\text{PER}(\text{ANA}_z \circ \text{DAJE}_z \circ \text{KNJIGU}_z \circ x)) : Z; S_{m,4}; \\
(\exists \text{book} )(\lambda_y.(\text{give } y \times \text{ana})) : p
\end{equation}

Tectogrammatically, because of the number parameter, no more dative clitics, and no auxiliary clitics can be placed in this sentence. Phenogrammatically, the sign denotes a set of \textit{S}-strings in each of which the clitic has now been encliticized onto the last phonological word in its first length one \textit{S}-string. Below is the expanded phenogrammatical term, as well as a listing of some of the actual \textit{S}-strings that the term denotes:

\begin{equation}
\vdash \lambda_w.\exists v t .[((\text{PER}(\text{ANA}_z \circ \text{DAJE}_z \circ \text{KNJIGU}_z) v) \land (\text{fst}_z v t)) \land \\
w = (\text{toZ}(\lambda_s.s = t \times \text{mu}) \circ (\text{rst}_z v)) : Z;
\end{equation}

\begin{enumerate}
\item \text{toZ}(\lambda_s.s = \text{ana}_z + \text{mu}) \circ \text{DAJE}_z \circ \text{KNJIGU}_z
\item \text{toZ}(\lambda_s.s = \text{daje}_z + \text{mu}) \circ \text{ANA}_z \circ \text{KNJIGU}_z
\item \text{toZ}(\lambda_s.s = \text{knjigu}_z + \text{mu}) \circ \text{DAJE}_z \circ \text{ANA}_z
\item etc.
\end{enumerate}

Now we show how to generate the sentence \textit{Ana mu ga daje} ‘Ana gives it to him’, which contains two clitics that have to occur in this order. This sentence can be pronounced only two ways, \textit{Ana mu ga daje} or \textit{Daje mu ga Ana}. Below we give the lexical entry for the accusative clitic.

\begin{equation}
\vdash (\text{wac }\gamma) : (z \rightarrow Z) \rightarrow Z; (\text{NP}_{\text{acc,m,sg,3}} \circ S_{m,n>3}) \circ S_{m,3}; \\
\lambda_G.(G \times) : (e \rightarrow p) \rightarrow p
\end{equation}
Phenogrammatically, this accusative clitic works the same way as the dative one we considered earlier. Tectogrammatically, the difference lies in the type of gap its argument can have, and the sentence number parameter. Because the accusative clitic reduces the number parameter of the sentence to 3, it is impossible to first place the accusative clitic and then the dative clitic. So we first have to bind the dative trace and place the dative clitic, and then bind the accusative trace and place the accusative clitic. The result is the following sign:

(210)  \[ \vdash (\text{wac ga})(\lambda y.(\text{wac mu})(\lambda x.\text{PER}(\text{ANA}_x \circ x \circ y \circ \text{DAJE}_z))): Z; S_{m,3}; \\
\text{(give } x \ y \ \text{ana)}: p \]

The dative clitic takes the first length one S-string and encliticizes onto the last phonological word that string is constructed out of. Then, that same length one S-string is rebuilt, except now it contains the clitic on its right edge. The accusative clitic does the same thing, and therefore, has no choice but to encliticize onto the phonological word that, at this point, already contains the dative clitic. The phenogrammatical term above denotes a set which contains exactly the following two strings of languages:

(211)  a. toZ(\lambda x. s = \text{ana}_x + \text{mu} + \text{ga}) \circ \text{DAJE}_z \\
b. toZ(\lambda x. s = \text{daje}_x + \text{mu} + \text{ga}) \circ \text{ANA}_x \\

We analyze genitive clitics similarly.
5.3.4 The Inherent Reflexive se

In this section, we analyze the inherent reflexive se, which occurs in the same slot in the clitic cluster as the true reflexive se. While the true reflexive se has a corresponding full form sebe, the inherent reflexive does not.

Recall that certain verbs require the occurrence of the inherent reflexive se, which contributes nothing in terms of meaning. We analyze such verbs as introducing a hypothesis via their lexical entries. For example, consider the sentence Boji ga se Ana ‘Ana is afraid of him’, where the relevant verb is bojati se ‘be afraid’.

We give the following lexical entries for boji, ga and the inherent reflexive se:

\[
(212) \quad x : z; SE; x : e \vdash \lambda_{uv}.PER((v \circ BOJI_{z} \circ u \circ x) : z \rightarrow z \rightarrow z; NP_{gen,m,s,g,3} \rightarrow NP_{nom,f,s,g,3} \rightarrow S_{m,6}; \lambda_{uv}(is-afraid \ u \ v) : e \rightarrow e \rightarrow p
\]

\[
(213) \quad \vdash (wac ga) : (z \rightarrow z) \rightarrow z; (NP_{gen,m,s,g,3} \rightarrow S_{m,n,>2}) \rightarrow S_{m,2};
\]

\[
\lambda_{G}(G \ y) : (e \rightarrow p) \rightarrow p
\]

\[
(214) \quad \vdash (wac se) : (z \rightarrow z) \rightarrow z; (SE \rightarrow S_{m,n,>1}) \rightarrow S_{m,1};
\]

\[
\lambda_{G}(G \ x) : (e \rightarrow p) \rightarrow p
\]

The tectogrammatical type SE is a special type introduced just for analyzing the inherent reflexive. In constructing the sentence, we combine the verb with the genitive trace and the subject, then withdraw the genitive hypothesis and place the genitive clitic, which results in the following sign:

\[
(215) \quad x : z; SE; x : e \vdash (wac ga)(\lambda_{y}.PER(ANA_{z} \circ BOJI_{z} \circ y \circ x)) : z;
\]

\[
S_{m,2}; (is-afraid \ y \ ana) : p
\]

While the verb introduces a hypothesis in its lexical entry, which is still in the context in the sign above, since se doesn’t contribute anything semantically the
semantic variable in the hypothesis does not occur anywhere in the verb’s semantic term. When the hypothesis is withdrawn, the semantic term is vacuously abstracted on:

\[ \vdash \lambda_x . \text{(wac ga)}(\lambda_y . \text{PER} (\text{ANA}_z \circ \text{BOJI}_z \circ y \circ x)) : z \rightarrow z; SE \rightarrow S_{m,2}; \lambda_x . \text{(is-afraid y ana)} : e \rightarrow p \]

The clitic se simply gets rid of this vacuous abstraction. Below is the sign that represents in the grammar the sentence Ana ga se boji ‘Ana is afraid of him’:

\[ \vdash (\text{wac se})(\lambda_x . \text{(wac ga)}(\lambda_y . \text{PER} (\text{ANA}_z \circ \text{BOJI}_z \circ y \circ x))) : z; S_{m,2}; \text{(is-afraid y ana)} : p \]

The phenogrammatical term of this sign denotes a set that contains the following two S-strings:

\[ \begin{align*}
\text{a. } & \text{toZ}(\lambda_z . s = \text{ana}_z + \text{ga} + \text{se}) \circ \text{BOJI}_z \\
\text{b. } & \text{toZ}(\lambda_z . s = \text{boji}_z + \text{ga} + \text{se}) \circ \text{ANA}_z
\end{align*} \]

5.3.5 The True Reflexive se

First we will analyze the non-clitic reflexive sebe, then extend that analysis to the clitic reflexive.

The Non-Clitic Reflexive

The reflexive sebe is an accusative case marked noun phrase, but has no gender or number features. In simple sentences, it is interpreted as coreferential with the subject. For example:
If *sebe* occurs in a subject controlled complement, it is interpreted as coreferential with the matrix subject.

(220) subject control

a. Ana\(_i\) hoće da voli sebe\(_i\).
   Ana\(_{NOM,f,sg,3}\) love\(_{sg,3}\) sebe\(_{ACC}\)
   ‘Ana wants to love herself’

b. Marko\(_i\) mora voljeti sebe\(_i\).
   Marko\(_{NOM,m,sg,3}\) must\(_{sg,3}\) love\(_{inf}\) sebe\(_{ACC}\)
   ‘Marko must love himself’

If *sebe* occurs in an object controlled complement, it is interpreted as coreferential with the matrix object; if *sebe* is the object controller, then it is interpreted as coreferential with the matrix subject.

(221) object control

a. Ana\(_i\) nagovara Marko\(_i\) da nominira sebe\(_i/\ast j\).
   Ana\(_{NOM,f,sg,3}\) persuade\(_{sg,3}\) Marko\(_{ACC,m,sg,3}\) DA nominate\(_{sg,3}\) sebe\(_{ACC}\)
   ‘Ana is persuading Marko to nominate himself’

b. Ana\(_i\) nagovara sebe\(_i\) da nominira Marko.
   Ana\(_{NOM,f,sg,3}\) persuade\(_{sg,3}\) sebe\(_{ACC}\) DA nominate\(_{sg,3}\) Marko\(_{ACC,m,sg,3}\)
   ‘Ana is persuading herself to nominate Marko’

We give the following lexical entry for the reflexive *sebe*:
(222) \[ \lambda_{Gw}. (G \ SEBE_z \ w) : (z \to z \to z) \to z \to Z; \prod_{g,g'}^{\text{Prs}} Gdr, n, n' ; \text{Num, p, p'} ; \text{Num, p, p'} ; \text{Prs} \]
\[ \left[ (\text{NP}_{\text{acc, f, s, g, n, p}} \to \text{NP}_{\text{nom, f, s, g, n', p'}} \to S_{m, 6}) \to \text{NP}_{\text{nom, f, s, g, n', p'}} \to S_{m, 6} \right] ; \]
\[ \lambda_{Fz}. F z z : (e \to e \to p) \to e \to p \]

So, *sebe* takes as its argument a finite verb phrase missing an accusative object and alters its meaning so that by the time the whole verb phrase combines with the subject the appropriate interpretation is obtained, and the reflexive is coreferential with the matrix subject (see Szabolsci (1987)). Phenogrammatically, the reflexive just lowers itself into its argument, much like a quantificational noun phrase, ensuring that it freely permutes with other clausal constituents, just like noun phrases in general.

Suppose we’re trying to analyze the sentence *Ana voli sebe* ‘Ana loves herself’. Below is the version of the lexical entry for *sebe* needed to generate the sentence *Ana voli sebe* ‘Ana loves herself’, and the result of its combination with the verb *voli*.

(223) \[ \lambda_{Gw}. (G \ SEBE_z \ w) : (z \to z \to z) \to z \to Z; \]
\[ (\text{NP}_{\text{acc, f, s, g, n, p}} \to \text{NP}_{\text{nom, f, s, g, n', p'}} \to S_{m, 6}) \to \text{NP}_{\text{nom, f, s, g, n', p'}} \to S_{m, 6} ; \]
\[ \lambda_{Fz}. F z z : (e \to e \to p) \to e \to p \]

(224) \[ \lambda_{w, \text{PER}} (w \circ \text{VOLI}_z \circ \text{SEBE}_z) : z \to Z; \text{NP}_{\text{nom, f, s, g, n, p}} \to S_{m, 6} ; \]
\[ \lambda_z. \text{love} z z : e \to p \]

At this point, the verb phrase can combine with the subject resulting in the following sign, whose semantic term guarantees the correct interpretation whereby the object and the subject of the verb are coreferential.

(225) \[ \text{PER (ANA}_z \circ \text{VOLI}_z \circ \text{SEBE}_z) : Z; S_{m, 6} ; \text{(love ana ana)} : p \]
Now we show an example of the reflexive occurring in subject controlled verb phrases. Suppose we are trying to construct a representation of *Ana hoče da voli sebe* ‘Ana wants to love herself’. We first combine the reflexive with the verb as in the example above, and then introduce a subject trace.

(226)

$$\vdash \lambda_w.\text{PER}(w \circ \text{VOLI}_z \circ \text{SEBE}_z);$$

$$\text{NP}_{\text{nom,f,sg,3}} \rightarrow \text{S}_m,6; \lambda_z.(\text{love } z z) \quad \text{NP}_{\text{nom,f,sg,3}}; x \vdash \text{NP}_{\text{nom,f,sg,3}}; x \rightarrow \text{PER}(x \circ \text{VOLI}_z \circ \text{SEBE}_z); \text{S}_m,6; \text{love } x x$$

We then proceed to construct an embedded clause with a subject gap, as we do for subject control sentences in general.

(227)

$$\vdash \lambda_x.\exists_y([x \circ \text{VOLI}_z \circ \text{SEBE}_z] \wedge w = (\text{DA}_z \circ v)]; \quad \text{NP}_{\text{nom,f,sg,3}}; x \vdash \text{PER}(x \circ \text{VOLI}_z \circ \text{SEBE}_z);$$

$$\text{S}_m,6 \rightarrow \text{S}_e,6; \lambda_q.\text{q} \quad \text{NP}_{\text{nom,f,sg,3}}; x \vdash \lambda_y.\exists_z([\text{PER}(x \circ \text{VOLI}_z \circ \text{SEBE}_z) \wedge w = (\text{DA}_z \circ v)];$$

$$\text{NP}_{\text{nom,f,sg,3}}; S_{m,6}; (\text{love } x x) \rightarrow \lambda_x.\exists_y([\text{PER}(x \circ \text{VOLI}_z \circ \text{SEBE}_z) \wedge w = (\text{DA}_z \circ v)];$$

The conclusion of the proof above can be an argument of a subject control verb.

Next, *hoče* ‘wants’ combines with this embedded clause with a subject gap, and then the whole verb phrase combines with the subject:

(228)

$$\vdash \lambda_w.\exists_{zyv}([\text{PER}(x \circ \text{HOČE}_z) \wedge w = y \circ \text{toZ}(L v)] : z \rightarrow z;$$

$$\text{NP}_{\text{nom,f,sg,3}} \rightarrow \text{S}_m,6; \lambda_z.\text{want } (\text{love } z z) ; e \rightarrow p$$

(229)

$$\vdash \lambda_w.\exists_{zyv}([\text{PER}(x \circ \text{HOČE}_z) \wedge w = y \circ \text{toZ}(L v)] : z; \text{S}_m,6; \text{want } (\text{love } z z) ; e \rightarrow p$$

179
Next, we look at an example where the reflexive occurs in the object controlled complement. Suppose we are trying to represent the sentence *Ana nagovara Marka da nominira sebe* ‘Ana is persuading Marko to nominate himself’. We combine the reflexive with the embedded verb, then construct an embedded sentence with a subject gap, resulting in the following sign.

$$\Gamma \Delta xFw_\forall y[(\textit{PER} \circ \textit{NOMINIRA}_z \circ \textit{SEBE}_z) v) \land w = (\textit{DA}_z \circ v)] : z \rightarrow z;$$

$$\textit{NP}_{\text{nom, f, sg, 3}} \circ S_{e, 6}; \lambda_x.(\textit{nominate} \; x \; x) : e \rightarrow p$$

Consider the lexical entry for *nagovara* ‘persuades’ given in the previous chapter.

$$\Gamma \Delta xFw_\forall y[(\textit{F} \circ \textit{eS} \; x) \land (\textit{PER} \circ \textit{NAGOVARA}_z \circ \textit{MARKA}_z) y) \land w = y \circ \textit{toZ}(L \; x)] :$$

$$z \rightarrow (z \rightarrow z) \rightarrow z \rightarrow z;$$

$$\textit{NP}_{\text{acc, m, sg, 3}} \circ (\textit{NP}_{\text{nom, m, sg, 3}} \circ S_{e, 6}) \circ \textit{NP}_{\text{nom, f, sg, 3}} \circ S_{m, 6};$$

$$\lambda_yFx_\forall.(\textit{persuade} \; y \; (F \; y) \; x) : e \rightarrow (e \rightarrow p) \rightarrow e \rightarrow p$$

We combine *nagovara* ‘persuades’ with its object *Marko*:

$$\Gamma \Delta xFw_\forall y[(\textit{F} \circ \textit{eS} \; x) \land (\textit{PER} \circ \textit{NAGOVARA}_z \circ \textit{MARKA}_z) y) \land$$

$$w = y \circ \textit{toZ}(L \; x)] : (z \rightarrow z) \rightarrow z \rightarrow z;$$

$$\textit{NP}_{\text{nom, m, sg, 3}} \circ S_{e, 6}) \circ \textit{NP}_{\text{nom, f, sg, 3}} \circ S_{m, 6};$$

$$\lambda_yFx_\forall.(\textit{persuade marko} \; (F \; \textit{marko}) \; x) : (e \rightarrow p) \rightarrow e \rightarrow p$$

Now we can combine *nagovara Marka* ‘persuades Marko’ with the controlled verb phrase, resulting in the following sign:

$$\Gamma \Delta yx_\forall[(\textit{PER} \circ \textit{NOMINIRA}_z \circ \textit{SEBE}_z) z) \land x = (\textit{DA}_z \circ z) \land$$

$$\textit{PER}(v \circ \textit{NAGOVARA}_z \circ \textit{MARKA}_z) y) \land w = y \circ \textit{toZ}(L \; x)] : z \rightarrow z;$$

$$\textit{NP}_{\text{nom, f, sg, 3}} \circ S_{m, 6};$$

$$\lambda_x.(\textit{persuade marko} \; (\textit{nominate marko marko}) \; x) : e \rightarrow p$$

Finally, combining this verb phrase with the matrix subject *Ana*, we get the following sign:

180
Now we show an example where the reflexive is the object controller. Consider the sentence Marko sebe smatra pristojnim ‘Marko considers himself polite’. Recall that we analyzed smatra ‘considers’ as taking an accusative noun phrase and an instrumental attributive adjective, ensuring that the accusative object and the adjective have the same gender and number. Below we repeat the lexical entry for the adjective given in the previous chapter, as well as the required tectogrammatical version of the lexical entry for the verb.

(234) $\vdash \lambda w.3xyz[(\text{PER} (\text{NOMINIRA}_z \circ \text{SEBE}_z) z) \land x = (\text{DA}_z \circ z) \land (\text{PER} (\text{ANA}_z \circ \text{NAGOVARA}_z \circ \text{MARKA}_z) y) \land w = y \circ \text{toZ}(L x)] : Z;$ $\text{S}_{m,6}; (\text{persue marko (nominate marko marko ana)} : p$

We cannot combine the reflexive sebe directly with this verb because of the type mismatch. However, we can introduce an accusative trace and combine smatra ‘considers’ with it, then combine the resulting sign with the instrumental adjective, and finally bind the accusative trace which results in the following sign.\(^{30}\)

(235) $\vdash \lambda v.\text{PRISTOJNIM}_z \circ v : z \rightarrow z; \text{N}_{\text{inst},m,sg} \rightarrow \text{N}_{\text{inst},m,sg}; \text{polite} : e \rightarrow p$
(236) $\vdash \lambda v_{FW}.\text{PER}(w \circ \text{SMATRA}_z \circ v \circ (F \circ \text{S})): z \rightarrow (z \rightarrow z) \rightarrow z \rightarrow z;$ $\text{NP}_{\text{acc},m,sg,3} \rightarrow (\text{N}_{\text{inst},m,sg} \rightarrow \text{N}_{\text{inst},m,sg}) \rightarrow \text{NP}_{\text{nom},m,sg,3} \rightarrow \text{S}_{m,6};$
   $\lambda xy_?.(\text{consider}(P x) y) : e \rightarrow (e \rightarrow p) \rightarrow e \rightarrow p$

\(^{30}\)In general, it is a theorem of both intuitionistic (i) and linear (ii) logic, which constitute the type system of our phenogrammatical and semantic component, and tectogrammatical component respectively, that for any formulas $\phi, \psi$ and $\chi$:

i. $\phi \rightarrow \psi \rightarrow \chi \vdash \psi \rightarrow \phi \rightarrow \chi$

ii. $\phi \rightarrow \circ \psi \rightarrow \circ \chi \vdash \circ \psi \rightarrow \circ \phi \rightarrow \circ \chi$
The reflexive can combine with this verb phrase, resulting in the following sign:

\[
(238) \quad \vdash \lambda_w.\mathbf{PER}(w \circ \text{SMATRA}_z \circ \text{SEBE}_z \circ \text{PRISTOJNIM}_z) : z \rightarrow z; \\
\text{NP}_{\text{nom},m,\text{sg},3} \circ \text{S}_{m,6} : \lambda_z.(\text{consider(} \text{polite} \; z) \; z) : e \rightarrow p
\]

Once this verb phrase combines with the subject Marko, we will get a sign associated with the semantic term is \( \vdash (\text{consider (polite marko) marko}) : p \), just as desired.

The lexical entry given for the reflexive above allows it to reflexivize finite verbs. However, unlike finite verbs, participles and infinitives are not phenomenographically keeping a slot for the subject: While the finite verb \( \text{vidi} \) ‘sees’ is phenomenographically of type \( z \rightarrow z \rightarrow z \), its participial counterpart \( \text{vidjela} \) and its infinitival counterpart \( \text{vidjeti} \) are phenomenographically of type \( z \rightarrow z \). So we need to give another lexical entry for a reflexive which can occur in non-finite verb phrases.

We give it the following lexical entry schema:

\[
(239) \quad \vdash \lambda_q.(Q \; \text{SEBE}_z) : (z \rightarrow z) \rightarrow z; (\text{NP}_{\text{acc},f,\text{sg},3} \circ \text{NP}_{\text{nom},f,\text{sg},3} \circ \text{T} \circ \text{NP}_{\text{nom},f,\text{sg},3} \circ \text{T}) \circ \text{NP}_{\text{nom},f,\text{sg},3} \circ \text{T} ; \lambda_{Fz}. F \; z \; z : (e \rightarrow e \rightarrow p) \rightarrow e \rightarrow p
\]

where \( T \) ranges over \( S_{\text{inf}} \) and \( Prd_{\text{pl}} \)

Semantically, this reflexive works just like the one we considered earlier. Textogrammatically, it is looking for infinitival or participial verbs, instead of finite ones. Its phenomen grammatical term reflects the fact that, unlike finite verbs, non-finite verbs build \( S \)-strings and are not expecting to combine with its subject. This reflexive lowers itself into the non-finite verb’s object slot.
The Clitic Reflexive

Moving onto the clitic reflexive se, since the clitics in general cannot be placed in the sentence until all the potential hosts have been accounted for, the clitic reflexive se cannot be placed in the sentence until after the subject is already there. And we can’t get around this easily, because the subject could be a quantificational phrase: even if the clitic reflexive combined with the verb phrase and reserved a slot for the subject, the sentence number parameter would be reduced and so a quantificational subject could not be introduced.

However, once the subject has already combined with the verb phrase, we can no longer ensure that the sentence is interpreted correctly, with the object and the subject being coreferential. Therefore, in order to correctly compose a sentence in which the clitic reflexive clitic se occurs, we need to introduce a complicated hypothesis into the derivation so that the correct interpretation can be arrived at and so that the clitic can find an appropriate host, which may well be the subject.

We will use the following type abbreviations, for legibility purposes:

\[(240)\]

a. \( \text{REFL} = \text{def} \ \prod_{G:Gdr,r:Num,p:Prs}(NP_{\text{acc},g,n,p} \rightarrow NP_{\text{nom},g,n,p} \rightarrow S_{m,6}) \rightarrow NP_{\text{nom},g,n,p} \rightarrow S_{m,6} ) \)

b. \( \text{refl} = \text{def} (e \rightarrow e \rightarrow p) \rightarrow e \rightarrow p \)

c. \( r = \text{def} (z \rightarrow z \rightarrow Z) \rightarrow z \rightarrow Z \)

Below is the reflexive hypothesis we’d need to introduce to get Ana se voli ‘Ana loves herself’.

\[(241)\]

\[ G: r; \text{REFL}_{tsg,3}; G: \text{refl} \vdash G: r; \text{REFL}_{tsg,3}; G: \text{refl} \]
Below is the lexical entry for the clitic reflexive se.

\[ \vdash \lambda_{P_0.}(\text{wac } se)(\lambda_{z}.P(\lambda_{Q_0.Q z})): (x \rightarrow z) \rightarrow z; \]
\[ (\text{REFL}_{f,sg,3} \circ S_{m,n>1} \circ S_{m,1}; \lambda_{P_0.P(\lambda_{Q_0.Q z}):(refl \rightarrow p) \rightarrow p} \]

In the semantic term above, \( \vdash P : refl \rightarrow p, \vdash Q : e \rightarrow e \rightarrow p, \) and so
\[ \vdash \lambda_{Q_0.Q z}(Q z z) : (e \rightarrow e \rightarrow p) \rightarrow e \rightarrow p. \]

In the phenogrammatical term, \( \vdash P : r \rightarrow z, \vdash Q : z \rightarrow z \rightarrow z, \) and so
\[ \vdash \lambda_{Q_0.Q z} : (z \rightarrow z \rightarrow z) \rightarrow z \rightarrow z, \text{ and } \vdash P(\lambda_{Q_0.Q z}) : z. \]

Suppose we are trying to generate the sentence Ana se voli ‘Ana loves herself’.

We introduce the reflexive hypothesis and combine the verb voli with it, resulting in the following sign.

\[ G : x; \text{REFL}_{f,sg,3}; G : refl \vdash G(\lambda_{xy}.\text{PER}(y \circ \text{VOLI}_z \circ x)) : z \rightarrow z; \]
\[ \text{NP}_{\text{nom},f,sg,3} \circ S_{m,6}; G(\lambda_{xy}.\text{love } x y) : e \rightarrow p \]

Now this verb phrase combines with the subject, and then the reflexive hypothesis is withdrawn:

\[ G : x; \text{REFL}_{f,sg,3}; G : refl \vdash (G(\lambda_{xy}.\text{PER}(y \circ \text{VOLI}_z \circ x)) \text{ANA}_z) : z; S_{m,6}; \]
\[ (G(\lambda_{xy}.\text{love } x y) \text{ana}) : p \]

\[ \vdash \lambda_{G_0}(G(\lambda_{xy}.\text{PER}(y \circ \text{VOLI}_z \circ x)) \text{ANA}_z) : x \rightarrow z; \text{REFL}_{f,sg,3} \circ S_{m,6}; \]
\[ \lambda_{G_0}(G(\lambda_{xy}.\text{love } x y) \text{ana}) : refl \rightarrow p \]

The clitic reflexive can combine with the sentence with a withdrawn reflexive hypothesis. The result is the following sign, below which we show the step-by-step derivations of the phenogrammatical and the semantic term.

\[ \vdash (\text{wac } se)(\lambda_{z}.\text{PER}(\text{ANA}_z \circ \text{VOLI}_z \circ z)) : Z; S_{m,1}; (\text{love ana ana}) : p \]
\[ \vdash \lambda_{P_0.}(\text{wac } se)(\lambda_{z}.P(\lambda_{Q_0.Q z}))(\lambda_{G_0}(G(\lambda_{xy}.\text{PER}(y \circ \text{VOLI}_z \circ x)) \text{ANA}_z)); \]
\[ \sim \vdash (\text{wac } se)(\lambda_{z}.(\lambda_{G_0}(G(\lambda_{xy}.\text{PER}(y \circ \text{VOLI}_z \circ x)) \text{ANA}_z))(\lambda_{Q_0.Q z})) \]

184
We also give the lexical entry schema for the clitic reflexive which reflexivizes non-finite verbs.

\[\lambda_{P.z}(\lambda_{G.z}(Q \ z \ z))\]

(249) \[\lambda_{P.z}(\lambda_{G.z}(Q \ z \ z))\]

As in the case of non-clitic reflexives, the main difference between this clitic reflexive and the one defined for finite verbs is in the phenogrammatical term and the tectogrammatical type. In the next section, after we consider clitic auxiliaries which occur in the 1st slot in the cluster, we analyze je ‘is’ and show how to construct representations of sentences in which non-finite verbs are reflexivized.

5.3.6 Auxiliaries

There are three types of verbal clitics in Serbo-Croatian. The clitics of htjeti ‘want, will’ are mainly used to form periphrastic future tense. In the periphrastic future tense construction, they are subject control verbs, combining with either an
infinitival verb phrase or an embedded declarative clause missing a nominative argument.

Another set of verbal clitics are the present tense clitics of *biti* ‘be’ which take a variety of predicative complements, including adjectives, noun phrases, passive participles, prepositional phrases, and, finally, past participles in the periphrastic past tense construction (see Chapter 4). Finally, the aorist clitics of *biti* ‘be’ are used to form the conditional mood and require a past participle complement.

The examples below show various types of sentences containing verbal clitics.

(250) clitics of *htjeti*

a. Ja ću pivo.
   *Ja_1 NOM,sg,1 will_1 sg,1 beer ACC,n,sg*
   ‘I want a beer’

b. Ana će doći.
   *Ana NOM,f,sg,3 will_3 sg,3 come inf*
   ‘Ana will come’

c. Ana će da dođe.
   *Ana NOM,f,sg,3 will_3 sg,3 DA come_3 sg*
   ‘Ana will come’

(251) present tense clitics of *biti*

a. Oni su studenti.
   *they NOM,pl,3 are_3 pl,3 students NOM,m,pl,3*
   ‘They are students’

b. Marko je pametan.
   *Marko NOM,m,sg,3 is_3 sg,3 smart NOM,m,sg*
   ‘Marko is smart’

c. Mi smo iz Sarajeva.
   *we NOM,pl,1 are_1 pl,1 from Sarajevo GEN,n,sg*
   ‘We are from Sarajevo’
d. Knjiga je pročitana.
   book\textsubscript{NOM,f,sg} is\textsubscript{f,sg} read\textsubscript{pas,NOM,f,sg}
   ‘A/the book is/has been read’

e. Ti si došao.
   you\textsubscript{NOM,sg,2} are\textsubscript{sg,2} arrived\textsubscript{ppl,m,sg}
   ‘You arrived’

(252) aorist clitics of \textit{biti}

a. Ana bi htjela kupiti auto.
   Ana\textsubscript{NOM,f,sg,3} would\textsubscript{sg,3} want\textsubscript{ppl,f,sg} buy\textsubscript{inf} car\textsubscript{ACC,m,sg}
   ‘Ana would like to buy a car’

b. Vi biste trebali više učiti.
   you\textsubscript{NOM,pl,2} would\textsubscript{pl,2} need\textsubscript{ppl,m,pl} more study\textsubscript{inf}
   ‘You should study more’

All verbal clitics occur in the first slot in the clitic cluster, except for \textit{je}, the third person present tense clitic of \textit{biti}. We first analyze the slot 1 verbal clitics, and then return to \textit{je}.

\textbf{Slot 1 Auxiliaries}

First we consider slot 1 auxiliaries in main declarative clauses. If we consider only main clauses, precisely because these auxiliaries occur in the first slot in the clitic cluster, we can analyze them as taking an appropriate complement, and then the subject, and encliticizing onto the first word in the resulting string. However, such an analysis would not generalize to other types of clauses.

If we analyze slot 1 clitics as combining with a complement and then the subject, we predict that they encliticize onto some phonological word contained in
the complement or the subject. However, in embedded clauses the clitics obligatorily encliticize onto the complementizer *da*, in polar interrogatives onto the interrogative complementizer *da li*, and in constituent questions onto the fronted *wh* expression. So such an analysis would predict that the clitics do not encliticize onto the complementizer or the fronted *wh* expression, but occur further to the right of that initial element.

A fully general account of verbal clitic placement in different types of clauses can be achieved if we analyze them on a par with nominal clitics, as combining with sentences with an appropriate gap. We introduce the following type abbreviations:

(253) tectogrammatical type abbreviations
a. $IC = \text{def } \prod_{g:Gdr}[(\text{NP}_{nom,g},t,t') \rightarrow S_{inf,6}) \rightarrow \text{NP}_{nom,g},t,t' \rightarrow S_{m,6}]$

b. $FC = \text{def } \prod_{n:N, g:Gdr}[(\text{NP}_{nom,g},t,t') \rightarrow S_{e,n}) \rightarrow \text{NP}_{nom,g},t,t' \rightarrow S_{m,6}]$

c. $PC = \text{def } \prod_{d:D, g:Gdr}[(\text{NP}_{nom,g},t,t') \rightarrow Prd_{d}) \rightarrow \text{NP}_{nom,g},t,t' \rightarrow S_{m,6}]$

(254) semantic type abbreviations
a. $c = \text{def } (e \rightarrow p) \rightarrow e \rightarrow p$

Above, $IC$ is mnemonic for ‘infinitival complement clitic’, $FC$ for ‘finite complement clitic’, and $PC$ for ‘predicative complement clitic’.

Recall that *htjeti* ‘will, want’ clitics can combine with either an infinitival or a finite complement. For the infinitival-taking version of these clitics, the following hypothesis has to be introduced into the derivation:

(255) $G : z \rightarrow z \rightarrow z; IC_{r:Gdr}; G : c \vdash G : z \rightarrow z \rightarrow z; IC_{r:Gdr}; G : c$
The tectogrammatical type of the hypothesis has to be specified for particular person and number values, restricting the choice of a possible enclitic of \textit{htjeti}. We give the following lexical entry for the 3rd person singular \textit{če} which takes an infinitival complement.

(256) \[\lambda_F.(\textsf{wac} \textit{če})(\lambda_z.(P \lambda_{xy}.\textsf{PER}(x \circ y))): (z \to z \to z) \to z;\]
\[
\Pi_{k;K}[(\textsf{IC} \circ S_{k,6}) \circ S_{k,5}]; \lambda_F.(F(\lambda_Fx.\textsf{FUT}(Px)):(c \to p) \to p)
\]

To construct the sentence \textit{Ana če doći} ‘Ana will come’, we have to introduce an appropriate hypothesis, then construct a sentence as if it contained a subject control verb, although it doesn’t. Then, we withdraw the subject control verb hypothesis and create something that \textit{če} can combine with:

(257) \[\lambda_G.(G\textsf{DOČI}_z \textit{ANA}_z): (z \to z \to z) \to z; \textsf{IC}_f \circ S_{m,6};\]
\[\lambda_G.(G\textsf{arrive ana}) : c \to p\]

The result of \textit{če} combining with the sign above is shown below.

(258) \[\textsf{(wac} \textit{če})(\lambda_z.\textsf{PER}(\textsf{DOČI}_z \circ \textit{ANA}_z)) : z; S_5; \textsf{FUT}(\textsf{arrive ana}) : p\]

Tectogrammatically we get a main declarative clause whose number parameter is 5, i.e. the first enclitic slot has been filled. Phenogrammatically, \textit{če} permutes the \textit{S}-strings in its complement, and then encliticizes onto one of them. The term \textsf{PER}(\textsf{DOČI}_z \circ \textit{ANA}_z) is vacuously abstracted on, for typing reason. Unlike pronominal clitics which combine with clauses with an actual phenogrammatical gap, this verbal clitic’s complement doesn’t have a gap that needs to be plugged with the empty \textit{S}-string. Vacuous abstraction takes care of the fact that \textsf{(wac} \textit{če}) is
expecting to insert the empty $S$-string into its complement, because of how we defined $wac$. The phenogrammatical term denotes a set of two strings of languages, where either the clitic host is $Ana$ or $doči$.

Suppose we had to place the same clitic inside an embedded clause. The first thing that needs to be ensured is that the complementizer combines with a sentence before any clitics have been placed inside of it. This restriction is already built into our grammar because we assigned the complementizer the tectogrammatical type $S_{m,6} \rightarrow S_{e,6}$. The complementizer $da$ ensures that its complement sentence has no enclitics already placed inside of by requiring that the number parameter of its argument and result type is 6.

The second issue is guaranteeing that the complementizer itself occurs leftmost in the embedded clause and that it is treated as a length one string of languages, because in our theory enclitics encliticize onto the last phonological word in the clause-initial length one string of languages. This is also already built into our grammar because the complementizer builds phenogrammatical terms consisting of the length one string of languages constructed out of $da$ followed by the length one string of languages constructed out its complement clause.

Here is how we analyze $da$ če $Ana$ $doči$ ‘that Ana will arrive’. The main clause missing a subject control verb combines with the complementizer, and then the hypothesis is withdrawn, resulting in the following sign:

$$\lambda \overline{Gw} \cdot \exists \overline{v} [(G \overline{DO} \overline{C} \overline{I} \overline{Z} \overline{ANA} \overline{v}) \land \overline{w} = (\overline{DA} \overline{z} \circ \overline{v})] : (\overline{z} \rightarrow \overline{z} \rightarrow \overline{Z}) \rightarrow \overline{Z}; \ IC \overline{f} \rightarrow S_{e,6}; \lambda \overline{G} \cdot (G \text{arrive ana}) : \overline{c} \rightarrow \overline{p}$$

190
After će, with the appropriately instantiated $k$ parameter, combines with this sign, we get the following sign:

(260) $\vdash (\text{wac } \acute{c}e)(\lambda_{zw}.\exists_{v}(\text{PER} (\text{DOĆI}_z \circ \text{ANA}_z) \ v) \land w = (\text{DA}_z \circ \nu)) : z; S_{e,5};$ FUT(arrive ana) : p

Since the complementizer $da$ is required to be the first length one S-string in each pronunciation of the embedded clause, će has no choice but to encliticize onto it. This term also denotes a set of two strings of languages, where the first string is required to be constructed out of $da$ onto which the auxiliary encliticized, followed by some permutation of the subject and the infinitive.

Two remarks are in order. Consider the main clause $\text{Maja } \acute{c}e \text{ proćitati knjigu}$ ‘Maja will read a book’. It is represented in the grammar as follows:

(261) $\vdash (\text{wac } \acute{c}e)(\lambda_{zw}.\exists_{v}(\text{PER} (\text{PROĆITATI}_z \circ \text{MAJA}_z \circ \text{KNJIGU}_z) \ v) \land w = (\text{DA}_z \circ \nu)) : z; S_{e,5};$ FUT(read maja book) : p

Phenogrammatically, we predict that all permutations of the infinitive, its complement noun phrase, and the subject are possible, and the clitic encliticizes onto any of these.

First, in reality there are serious phonetic consequences when an enclitic attaches to a certain class of infinitives. Not all infinitives are affected, and whether or not an infinitive is affected by the encliticization of $htjeti$ clitics is itself mor-phophonemically conditioned. The infinitives ending in –ći, for example, are not affected.
These phonetic consequences are partly reflected in orthography. In Croatia, when the affected infinitive hosts a htjeti clitic, it is typically written as Pročitat će Maja knjigu, while elsewhere the standard is to write Pročitaće Maja knjigu. Either way, one syllable is lost, and depending on the coda of the penultimate syllable of the infinitive, even more dramatic changes occur. Here we just note this fact, but cannot adequately represent it in our framework.

Second, our representation of Maja će pročitati knjigu ‘Maja will read a book’ predicts that all of the following are possible:

\[(262) \begin{align*}
a. \text{Maja} & \quad \text{če pročitati knjigu.} \\
Maja_{\text{NOM},f,sg,3} & \quad \text{will}_{\text{sg,3}} \quad \text{read}_{\text{inf}} \quad \text{book}_{\text{ACC},f,sg} \\
& \quad \text{‘Maja will read a/the book’}
\end{align*}\]

b. Maja će knjigu pročitati.

c. Knjigu će pročitati Maja.

d. Knjigu će Maja pročitati.

e. Pročitaće knjigu Maja.

f. Pročitaće Maja knjigu.

However, sometimes the infinitive’s enclitics can attach to the infinitive itself and not climb out, in which case the infinitival verb phrase remains contiguous and must occur on the right edge of the matrix clause. We postpone the analysis of this phenomenon until Chapter 7, where we generalize and refine our analysis of control in general.

To construct a sentence in which there is a finite subject-controlled verb phrase, such as Ana će da vidi Marka ‘Ana will see Marko’ we have to introduce the following hypothesis:
Eventually, the hypothesis is withdrawn, resulting in the following sign:

$$\vdash \lambda H.\lambda x.\exists v.((\text{PER}(x \circ \text{VIDI}_z \circ \text{MARKA}_z) v) \land w = \text{DA}_z \circ v) \ ANA_z : ((z \to z) \to (z \to z)) \to z; \text{FC}_{6,f} \to S_{m,6}; \lambda G.\lambda c.\lambda z.\text{ANA}_z : (z \to z) \to (z \to z); \text{FC}_{6,f}; G : c$$

We give the following lexical entry for $\acute{c}e$ which occurs with a finite verb phrase complement:

$$\lambda P. (\text{wac} \ acute{c}e)(\lambda z.\lambda P(\lambda w.\lambda x.\text{PER} w \circ \text{ANA}_z \circ \text{ANA}_z) y = \text{toZ} k(\lambda i.\lambda e.\lambda s)) : (((z \to z) \to (z \to z)) \to z) \to z; k : k; k; k; \text{FUT}(\text{see marko ana}) : (c \to p) \to p$$

In the phenogrammatical term above, $\vdash P : (((z \to z) \to (z \to z)) \to z)$, and $\vdash Q : z \to z$. Combining this version of $\acute{c}e$ with the finite clause missing its subject results in the following sign:

$$\vdash (\text{wac} \ acute{c}e)(\lambda z.\text{PER} \circ \text{ANA}_z \circ \text{ANA}_z \circ \text{ANA}_z) y = \text{toZ} k(\lambda w.\lambda x.\exists v.((\text{PER}(\text{VIDI}_z \circ \text{MARKA}_z) v) \land w = \text{DA}_z \circ v)) : z; S_5; \text{FUT}(\text{see marko ana}) : p$$

Recall that embedded clauses generally have to remain contiguous and occur on the right edge of the matrix clause. This essentially leaves some phonological word in the noun phrase subject as the possible host. So, the clitic takes the subject, permutes it (in case the subject noun phrase is longer than one), and encliticizes onto the last phonological word in the first $S$-string. In this case, it encliticizes onto $Ana$. Then, it plugs the subject gap in the embedded clause, which results in an $S$-language.
Finally, it fuses the $S$-language which is the permutation of the subject with the clitic properly attached, with the $S$-language that represents the controlled complement. It compacts and the turns into length one $S$-string the controlled finite complement. This is to prevent the finite complement from the intrusion of main clause adverbials. Since language fusion is defined in terms of concatenation, the clitic and its host are required to precede the embedded clause, in which for independent reasons the complementizer must occur left-most. Below we show the step by step derivation of this phenogrammatical term:

(267) \[ \lambda_p (wac \, \acute{\epsilon} e) (\lambda_{zR}(\lambda_{Qw}(\text{PER } w) \bullet z \lambda_y.y = \text{toZ } k(Q(i \in_S))) (\lambda_{H,H}(\lambda_{xw}.\exists_v[[\text{PER} (x \circ \text{VIDI}_z \circ \text{MARKA}_z) v) \land w = \text{DA}_z \circ v]) \text{ANA}_z) \]

\[ \overset{\sim}{\sim} (wac \, \acute{\epsilon} e) (\lambda_{zR}(\lambda_{Qw}(\text{PER } w) \bullet z \lambda_y.y = \text{toZ } k(Q(i \in_S))) (\lambda_{xw}.\exists_v[[\text{PER} (x \circ \text{VIDI}_z \circ \text{MARKA}_z) v) \land w = \text{DA}_z \circ v]) \text{ANA}_z) \]

\[ \overset{\sim}{\sim} (wac \, \acute{\epsilon} e) (\lambda_{zR}(\lambda_{Qw}(\text{PER } w) \bullet z \lambda_y.y = \text{toZ } k(Q(i \in_S))) (\lambda_{xw}.\exists_v[[\text{PER} (x \circ \text{VIDI}_z \circ \text{MARKA}_z) v) \land w = \text{DA}_z \circ v]) \text{ANA}_z) \]

\[ \overset{\sim}{\sim} (wac \, \acute{\epsilon} e) (\lambda_{zR}(\lambda_{Qw}(\text{PER } w) \bullet z \lambda_y.y = \text{toZ } k(Q(i \in_S))) (\lambda_{xw}.\exists_v[[\text{PER} (x \circ \text{VIDI}_z \circ \text{MARKA}_z) v) \land w = \text{DA}_z \circ v]) \text{ANA}_z) \]

If an adverb had been added to the main clause, such as *uskoro* ‘soon’ it too could be a clitic host. The phenogrammatical representation of *Ana \, \acute{\epsilon} e uskoro da vidi*
Marka ‘Ana will soon see Marko’, where uskoro ‘soon’ combined with the main clause verb phrase before the clitic trace was bound is the following:

\[ \vdash (\text{wac} \, \text{c}z) (\lambda_x, w, \exists x [((\text{PER ANA} z) \bullet z \lambda_y. y = \text{toZ k} (\lambda_x. \exists y [((\text{PER VIDI} z \circ \text{MARKA} z) \, v) \land w = \text{DA} z \circ \nu]) v') \land w' = \text{USKORO} z \circ \nu']) : z \]

The adverb is shuffled into the clause, and can occur in exactly three positions: (i) clause initially, hosting the clitic; (ii) immediately after the subject Ana to which the clitic attached; or, (iii) clause finally, to the right of the controlled complement. Since the controlled complement has been turned into a length one S-string, the main clause adverb cannot be shuffled inside of it.

Next, we analyze aorist clitics of biti. The aorist clitics of biti can combine with past participles only. However, in combination with these clitics, they do not express past meaning at all. In our analysis of predicatives in the previous chapter, we assumed that past participles are a kind of predicative complement, and that they carry the past tense meaning. For example, we gave the following lexical entry for spavao ‘slept’.

\[ \vdash \text{SPAVAO} z : z; \text{NP} \, \text{nom,m,sg,3} \rightarrow \text{Prd}_\text{pl}; \lambda_x. \text{PST(sleep } x) : e \rightarrow p \]

To distinguish between conditional mood and past tense, we assume that participles are associated with lexical entries like the following one:

\[ \vdash \text{SPAVAO} z : z; \text{NP} \, \text{nom,m,sg,3} \rightarrow \text{S}_\text{m,6}; \lambda_x. \text{sleep } x : e \rightarrow p \]

Note that this sign has an ordinary declarative sentence as its result type, and that it’s not associated with past meaning. Because of its phenogrammatical term
and type, however, there is no danger of this participle combining with a subject and producing a non-sentence *Marko spavao. This sign is a possible argument of an aorist clitic of biti building a sentence in conditional mood. We then give the following rule schema which converts conditional forming participles into past tense forming ones:

\[(271)\]
\[
\vdash \phi : z; NP_{t,t',t'',t'''} \rightarrow S_{m,6}; \sigma : e \rightarrow p
\]
\[
\vdash \phi : z; NP_{t,t',t'',t'''} \rightarrow Prd_{pl}; \lambda x.PST(\sigma x) : e \rightarrow p \quad \text{[CPL]}
\]

Consider the sentence Maja bi kupila auto ‘Maja would buy a car’. As in the case of other verbal clitics, we introduce a hypothesis corresponding to the auxiliary, and then withdraw it, resulting in the following sign:

\[(272)\]
\[
\vdash \lambda G.(G(KUPILA_{z} \circ AUTO_{z}) MAJA_{z}):(z \rightarrow z \rightarrow z) \rightarrow z;
\]
\[
((NP_{nom,f,sg,3} \rightarrow S_{m,6}) \rightarrow NP_{nom,f,sg,3} \rightarrow S_{m,6}) \rightarrow S_{m,6};
\]
\[
\lambda G.(\text{exists car})(\lambda x,G(\lambda y,\text{buy }x\ y) \text{ maja}) : c \rightarrow p
\]

We give the following lexical entry for bi, where W is some unanalyzed propositional operator:

\[(273)\]
\[
\vdash \lambda F.(wac \ bi)(\lambda z.F(\lambda x.PER x \circ y)) : ((z \rightarrow z \rightarrow z) \rightarrow z) \rightarrow z;
\]
\[
\prod_{k:K}[((NP_{nom,\tau,sg,3} \rightarrow S_{m,6}) \rightarrow NP_{nom,\tau,sg,3} \rightarrow S_{k,6}) \rightarrow S_{k,6}] ;
\]
\[
\lambda F.F(\lambda P_{x}.W(P\ x)) : (c \rightarrow p) \rightarrow p
\]

The sentence Maja bi kupila auto ‘Maja would buy a car’ is then represented by the following sign in the grammar:

\[(274)\]
\[
\vdash (wac \ bi)(\lambda z.PER KUPILA_{z} \circ AUTO_{z} \circ MAJA_{z}) : z; S_{m,5};
\]
\[
(\text{exists car})(\lambda x.W(\text{buy }x\ \text{maja}) : p
\]
The clitic *bi* permutes all the *S*-strings in its complement, then attaches to the last phonological word in the first *S*-string. So, we predict that the sentence is pronounceable six different ways, corresponding to the six different orders of the participle, the subject and the object, with the clitic *bi* attached to the first word in each case.

Finally, we give a lexical entry for a representative of the clitic version of the copula. Recall from the previous chapter that non-aorist finite forms of the verb *biti* have the tectogrammatical type is $\Pi_{g:Gdr, p:D}[\langle \text{NP}_{\text{nom}, g, t, t'} \rightarrow \text{Prd}_d \rangle \rightarrow \text{NP}_{\text{nom}, g, t, t'} \rightarrow \text{S}_{m, 6}]$, where terms of type D specify the kind of the predicative phrase in question.

Excluding *je* which we address in the next section, we analyze present tense *biti* clitics on a par with other verbal clitics, whereby they combine with the predicative complement and then the subject, instead of being placed inside a sentence with an appropriate gap, as with pronominal clitics. We give the following lexical entry for *si*:

\[(275) \quad \vdash \lambda_F. (\text{wac } s_i)(\lambda_z. (\lambda_{x,y}. \text{PER}(x \circ y))): ((z \rightarrow z \rightarrow z) \rightarrow z) \rightarrow z;\]
\[\Pi_{k:K}[\langle \text{PC} \circ \circ \text{S}_{k,6} \circ \circ \text{S}_{k,5} \rangle; \lambda_F. (\lambda_{P_x}. P x)) : (c \rightarrow p) \rightarrow p\]

Phenogrammatically, these enclitics function exactly like *htjeti* clitics which take infinitival verb phrase complements, or aorist *biti* clitics: They permute the *S*-strings in their complement, and then properly attach themselves.
Slot 6 Auxiliary je

The clitic je occurs in the rightmost slot of the clitic cluster. Just like other enclitics, we have to analyze it as combining with a sentence which is missing an expression exactly like je. Below is the schema for the hypothesis that has to be introduced:

\[(276) \ G : z \rightarrow z \rightarrow z ; PC_{T,Gdt,T'} : D ; G : c \vdash G : z \rightarrow z \rightarrow z ; PC_{T,Gdt,T'} : D ; G : c\]

We give je the following lexical entry:

\[(277) \vdash \lambda_F . (\text{wac je})(\lambda_x . P(\lambda_{xy}.\text{PER} x \circ y)) : ((z \rightarrow z \rightarrow z) \rightarrow z) \rightarrow z;\]
\[\prod_{k : K} [(PC -o S_{k,\mu>0}) -o S_{k,0}] ; \lambda_F . (F(\lambda_P.x.P x)) : (c \rightarrow p) \rightarrow p\]

So, this lexical entry is completely analogous to other present tense biti clitics, except for the tectogrammatical number parameter, which je reduces to 0 as it is the rightmost clitic.

To construct a sentence with a pronominal clitic and je, such as Ana ga je vidjela ‘Ana saw him’, we have to introduce an accusative noun phrase trace, and a je trace:

\[(278) \ G : z \rightarrow z \rightarrow z ; PC ; G : c,\]
\[x : z ; NP_{acc,m,sg,3} ; x : e\]
\[\vdash (G(\text{VIDJELA}_z \circ x) \text{ANA}_z) : z ; S_{m,6} ; (G(\text{PST}(\text{see} x)) \text{ana}) : p\]

We first bind the accusative trace, and combine the resulting sign with the accusative clitic, which gives us the following:

\[(279) \ G : z \rightarrow z \rightarrow z ; PC ; G : c \vdash (\text{wac ga})(\lambda_x . G(\text{VIDJELA}_z \circ x) \text{ANA}_z) : z ; S_{m,3} ; (G(\text{see} x) \text{ana}) : p\]

198
Now we bind the je trace and combine je with the resulting sign, which yields the following representation of Ana ga je vidjela ‘Ana saw him’:

\[
\begin{aligned}
&\vdash (\text{wac je})(\lambda z. (\text{wac ga}))(\lambda w. \text{PER}(\text{VIDJELA}_z \circ w \circ \text{ANA}_z)) : Z; \\
&S_{m,0}; \text{PST(see ana x)} : p
\end{aligned}
\]

Essentially, the only two phonological words Ana and vidjela can freely permute. The accusative clitic encliticizes onto the first phonological word, and constructs a length one S-string out of the result. je then encliticizes onto the first phonological word in that string of languages, which is either Ana ga or vidjela ga. So the clitics stack exactly as desired, and the whole sentence is predicted to be pronounceable two different ways, Ana ga je vidjela or Vidjela ga je Ana.

It is well known that in Serbo-Croatian, in the presence of the clitic se, je is typically not pronounced. This is true both for the inherent reflexive and the true reflexive se. For example:

\[(281) \quad \begin{array}{ll}
\text{a.} & \text{Ana se vidjela.} \\
& \text{Ana}_{\text{NOM,f,sg,3}} \text{ se } \text{vidjela.} \\
& \text{Ana saw herself} \\
& \text{cf. Ana se je vidjela.}
\end{array} \]

\[(281) \quad \begin{array}{ll}
\text{b.} & \text{Ana ga se bojala.} \\
& \text{Ana}_{\text{NOM,f,sg,3}} \text{ ga } \text{se } \text{bojala.} \\
& \text{Ana was afraid of him} \\
& \text{cf. Ana ga se je bojala.}
\end{array} \]

Sentences which contain both clitics are not necessarily altogether unacceptable, but we certainly want to at least allow the possibility of je disappearing in the presence of se.
We can easily account for this phenomenon because of the number parameter of sentences. Basically, the clitic je ‘knows’ whether its complement sentence contains a slot 5 clitic (se) or not. In the former case, its argument sentence will have the number parameter 1, but in the latter case its parameter will be greater than 1.

First we show how to use the non-clitic reflexive which reflexivizes non-finite verbs, whose lexical entry schema we presented earlier in the chapter. Suppose we want to generate the sentence *Ana je sebe vidjela* ‘Ana saw herself’. Below we give the needed non-clitic version of the reflexive, as well as the lexical entry for the participle *vidjela*:

\[(282)\]  
\[\begin{align*}
&\vdash \lambda_{Q}. (Q \text{SEBE}_{z}) : (z \to z) \to z; \\
&(\text{NP}_{\text{acc}, f, sg, 3} \to \text{NP}_{\text{nom}, f, sg, 3} \to \text{Prd}_{\text{pl}}) \to \text{NP}_{\text{nom}, f, sg, 3} \to \text{Prd}_{\text{pl}}; \\
&\lambda_{Fz,Fz} : (e \to e \to p) \to e \to p \\
&\vdash \lambda_{x}.\text{VIDJELA}_{x} \circ x : z \to z; \text{NP}_{\text{acc}, f, sg, 3} \to \text{NP}_{\text{nom}, f, sg, 3} \to \text{Prd}_{\text{pl}}; \\
&\lambda_{xy,PST(\text{see } x y)} : e \to e \to p \\
\end{align*}\]

These two signs can combine resulting in the following:

\[(283)\]  
\[\vdash \text{VIDJELA}_{z} \circ \text{SEBE}_{z} : z; \text{NP}_{\text{nom}, f, sg, 3} \to \text{Prd}_{\text{pl}}; \lambda_{z,PST(\text{see } z z)} : e \to p\]

Now we can introduce the hypothesis for the copula, combine the resulting sign with the subject, and then withdraw the hypothesis:

\[(284)\]  
\[\vdash \lambda_{G}.G(\text{VIDJELA}_{z} \circ \text{SEBE}_{z}) \circ \text{ANA}_{z} : ((z \to z \to z) \to z) \to z; \\
(\text{PC}_{\text{pl}} \to \text{S}_{m,6}) \to \text{S}_{m,6}; \lambda_{G}.G(\lambda_{z,PST(\text{see } z z)})\circ \text{ana} : (c \to p) \to p\]

Combining this sign with je, we get the following sign, which has the expected meaning and is predicted to be pronounceable six different ways:

\[(285)\]  
\[\vdash (\text{wac } je)(\lambda_{z,PER(\text{VIDJELA}_{z} \circ \text{SEBE}_{z} \circ \text{ANA}_{z})}) : z; \\
\text{S}_{m,6}; \text{PST(see ana ana)} : p\]
Now we consider the clitic version of this reflexive, whose lexical entry schema we presented earlier in this chapter. Suppose we want to derive Ana se je vidjela ‘Ana saw herself’, and the synonymous Ana se vidjela, where je is dropped. We give the following lexical entry for the clitic se which reflexivizes feminine singular participles.

\[(286) \vdash \lambda P. (\wedge c e) (\lambda z. P (\lambda Q. (i e s))) : (((z \rightarrow z) \rightarrow z) \rightarrow z; (((NP_{acc, fsg, 3} \rightarrow NP_{nom, fsg, 3} \rightarrow Prd_{pl}) \rightarrow NP_{nom, fsg, 3} \rightarrow Prd_{pl}) \rightarrow S_{m,n \geq 1} \rightarrow S_{m,1}; \lambda P (\lambda Q. (Q z z))) : (refl \rightarrow p) \rightarrow p\]

In the derivation of the sentence above, we first introduce the hypothesis corresponding to a reflexive looking for non-finite complements, namely \( \vdash R : (z \rightarrow z) \rightarrow z; NP_{acc, fsg, 3} \rightarrow NP_{nom, fsg, 3} \rightarrow Prd_{pl}; R : refl \). We also introduce a hypothesis corresponding to the copula, \( \vdash C : z \rightarrow z; (NP_{nom, fsg, 3} \rightarrow Prd_{pl}) \rightarrow NP_{nom, fsg, 3} \rightarrow S_{m,6}; C : c \). Then we combine the resulting sign with the subject. Because je reduces the number parameter to 0, we first have to withdraw the reflexive hypothesis, resulting in the following sign:

\[(287) \vdash (NP_{nom, fsg, 3} \rightarrow Prd_{pl}) \rightarrow NP_{nom, fsg, 3} \rightarrow Prd_{pl}; C \]

\( \vdash \lambda R. ((C (R x. \mbox{VIDJELA} z \circ x)) \mbox{ANA} z) : ((z \rightarrow z) \rightarrow z) \rightarrow z; (((NP_{acc, fsg, 3} \rightarrow NP_{nom, fsg, 3} \rightarrow Prd_{pl}) \rightarrow NP_{nom, fsg, 3} \rightarrow Prd_{pl}) \rightarrow S_{m,6}; C : \lambda R. ((C (R \mbox{see}) t) \mbox{ANA} z) : refl \rightarrow p \)

Combining the reflexive looking for non-finite complements with this sign, results in the following:

\[(288) \vdash (NP_{nom, fsg, 3} \rightarrow Prd_{pl}) \rightarrow NP_{nom, fsg, 3} \rightarrow Prd_{pl}; C \]

\( \vdash (\wedge c e) (\lambda z. (C \mbox{VIDJELA} z \mbox{ANA} z)) : z; S_{m,1}; C : (\lambda z. (\mbox{see} z z) : p) \)

Now we can withdraw the other hypothesis, resulting in the following:
This sign can combine with je which we’ve already introduced, or the version of je which remains unpronounced because it ‘knows’ that its complement already contains the clitic reflexive. We list both of these below.

\[ (290) \quad \vdash \lambda \text{c}. (\text{wac}\; \text{se})(\lambda z. (c\; \text{VIDJELA}_z\; \text{ANA}_z)) : (z \to z \to z) \to z; \]
\[ ((\text{NP}_{\text{nom}, f, sg, 3} \to \text{Prd}_{\text{pl}}) \to \text{NP}_{\text{nom}, f, sg, 3} \to \text{Prd}_{\text{pl}}) \to \text{Sm}, 1; \]
\[ \lambda \text{c}. \text{c} (\lambda z. \text{see } z\; z) : c \to p \]

The original one permutes its argument and encliticizes. The other one also introduces permutation, but is not itself pronounced. The latter is constrained to apply only to ‘gappy’ clauses whose parameter is 1, i.e. only those that already contain the reflexive. Below we show the representation of both Ana se je vidjela and Ana se vidjela ‘Ana saw herself’.

\[ (291) \quad \vdash \lambda p. (\text{wac}\; \text{je})(\lambda z. p(\lambda x_y. \text{PER}(x \circ y))) : ((z \to z \to z) \to z) \to z; \]
\[ \prod_{k:K}[(\text{PC} \to \text{S}_{k, l>0}) \to \text{S}_{k, 0}]; \lambda f. (f(\lambda p. x) \to (c \to p) \to p) \]

For each of them, we are predicting two possible pronunciation, depending on whether vidjela or Ana is hosting the clitic(s).
5.3.7 1C and 1W placement

Preliminaries

In this section we explicitly address the 1C and 1W placement of enclitics. Recall that 1C placement refers to encliticization to the last phonological word of the initial constituent, and 1W refers to encliticization to the first phonological word of the initial constituent. However, not everything counts as a ‘constituent’ with respect to 1C placement. Table 5.3 shows the different possibilities of enclitic cluster placement, and marks each possibility in terms of whether the permissive or the conservative version of our grammar predicts that placement or not.

These are pretty much all the types of constituents which allow 1C placement. Clauses and verb phrases, including passive and participial verb phrases, cannot host enclitics. We return to the possibility of infinitives hosting clitics in Chapter 7.

The differences between the permissive and conservative versions of our grammar stem from the differences concerning noun phrase and prepositional phrase composition. Since we analyze enclitics as attaching to the last word in the initial length one \( S \)-string, extending the grammar comes down to specifying in a more fine grained way what can count as that initial length \( S \)-string in a clause.

The problem of generalizing these two grammars to cover clitic placement possibilities comes down to two issues: (i) allow noun phrases and prepositional phrases to fuse into \( S \)-strings of length one in the permissive grammar, and (ii)
<table>
<thead>
<tr>
<th>TYPE</th>
<th>PLACEMENT</th>
<th>permissive grammar</th>
<th>conservative grammar</th>
</tr>
</thead>
<tbody>
<tr>
<td>adverbial phrase</td>
<td>Vrlo brzo je Ana došla.</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Vrlo je brzo Ana došla.</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td></td>
<td>’Ana arrived very fast’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>prepositional adverbials</td>
<td>U velikom gradu je ona živjela.</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>and predicatives</td>
<td>U velikom je gradu ona živjela.</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td></td>
<td>‘She lived in a big city’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>adjective + noun</td>
<td>Pametan momak je došao.</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Pametan je momak došao.</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td></td>
<td>‘A smart young man arrived’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>noun + postnominal modifier</td>
<td>Djevojka iz Beograda je lijepa.</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Djevojka je iz Beograda lijepa</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td></td>
<td>‘The girl from Belgrade is pretty’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>adjective + noun + postnominal modifier</td>
<td>Pametan momak iz Beograda je došao.</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Pametan je momak iz Beograda došao.</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td></td>
<td>‘A smart young man from Belgrade arrived’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>quantificational</td>
<td>Svaka djevojka je lijepa.</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>determiner + noun</td>
<td>Svaka je djevojka lijepa.</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td></td>
<td>‘Every girl is pretty’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.3: Interim summary evaluation of grammar with respect to enclitic cluster placement.
allow ‘reaching into’ the clause-initial length one S-string and split it into a length
two S-strings, the first one built out of the initial phonological word. The latter is
needed in both versions of the grammar.

**Fusing Noun Phrases and Prepositional Phrases**

This is needed for the permissive grammar only, to allow enclitics to attach
to the last phonological word in an initial multi-word noun phrase. All phrasal
noun phrases are quantificational, either because they contain a quantificational
determiner, or because they underwent the [Quant] rule (see Chapter 3). So we
need to write a rule that will target quantificational noun phrases and alter their
phenogrammatical terms appropriately.

\[
\vdash \phi : (z \rightarrow z) \rightarrow z; (\text{NP}_{t,t',t'',3} \rightarrow \text{S}_{m,h}) \rightarrow \text{S}_{m,h}; \sigma : (e \rightarrow p) \rightarrow p
\]

\[
\vdash \lambda z. F(\text{toZ k}(\lambda x. \text{PER } x)) : (z \rightarrow z) \rightarrow z; (\text{NP}_{t,t',t'',3} \rightarrow \text{S}_{m,h}) \rightarrow \text{S}_{m,h}; \\
\sigma : (e \rightarrow p) \rightarrow p
\] [1Z]

We illustrate how the rule works with a concrete example. Consider the noun
phrase *lijepa djevojka* ‘a pretty girl’ after it has undergone the [Quant] rule. Focusing
on the phenoterm only, since the tectogrammatical and the semantic compo-
nent of the sign remain unchanged, here is how the [Z1] rule alters it:

\[
\vdash \lambda z. z \rightarrow z : (z \rightarrow z) \rightarrow z
\]

\[
\vdash \lambda F. F(\text{toZ k}(\lambda (\lambda z. z) \circ \text{DJEVOJKA}_z)((\lambda x. \text{PER } x))) : (z \rightarrow z) \rightarrow z
\] [1Z]

The conclusion of this proof reduces to:

\[
\vdash \lambda F. F(\text{toZ k}(\text{PER } (\lambda z. z) \circ \text{DJEVOJKA}_z)) : (z \rightarrow z) \rightarrow z
\]
So the noun phrase has been turned into a length one string of languages. Should it occur clause initially after it combines with the verb phrase, we will get 1C enclitic placement, because the enclitics would attach to the initial length one string of languages. With the addition of this rule to the permissive grammar, we get 1C placement of clitics for all noun phrases. Noun phrases, of course, do not have to undergo this rule, so 1W placement is still possible.

As for prepositional phrases, we can simply add the predicative and adverbial prepositional phrase lexical entries from the conservative grammar to the permissive grammar.

**Splitting Initial Constituents**

The second extension involves splitting a clause initial length one $S$-string into a length one $S$-string constructed out of the initial phonological word and the rest of that initial $S$-string. This is needed for both the permissive and the conservative grammar.

This would allow the enclitic cluster to be hosted by an initial phonological word which under normal circumstances cannot be separated from the rest of its constituent, such as a noun or an adjective in the presence of a postnominal modifier in the constituent noun phrase.

Recall that Progovac (1996) maintains that only phonological that are normally separable from the remainder of their constituent can host clitics. So, for example,
she doesn’t accept examples where the noun and the postnominal modifier sequence is split by the enclitic cluster. If one wishes to maintain these judgments, then the following rule can simply be omitted from the grammar.

For the grammar which is more permissive and allows enclitics to attach to phonological words which normally can’t be separated from the remainder of their constituent, we first define the function \( \text{br1} : \mathbb{Z} \rightarrow \mathbb{Z} \), for ‘break 1st’. Intuitively, this function looks at the typical phenogrammatical denotation of a sentence (a set of S-strings), and finds the initial length one S-strings. It then ‘zooms in’ and tries to find the p-string that the initial S-string is constructed out of. If it’s constructed out of a length one p-string, meaning that it consists of a single phonological word, nothing happens. But if it consists of a longer s-string, the function breaks it up into two S-strings: one constructed out of the initial phonological word, and the other one constructed out of the other phonological words.

\[
\begin{align*}
\text{br1} & \overset{\text{def}}{=} \lambda_{F_W}.\exists_{v_L}[(F(jv)) \land (\text{fst}_S v(jt)) \land w = \text{toZ}(\lambda_{s.S} = j(\text{fst}_p t)) \circ \text{toZ}(\lambda_{s.S} = \text{rst}_p t) \circ (\text{rst}_S v)] \\
\end{align*}
\]

Then we add the following rule to the grammar:

\[
\begin{align*}
\Gamma \vdash \phi : \mathbb{Z}; S_{m,6}; \sigma : p \\
\therefore \Gamma \vdash (\text{br1 } \phi) : \mathbb{Z}; S_{m,6}; \sigma : p \quad [1W]
\end{align*}
\]

This rule can apply to any declarative main clause which hasn’t had any clitics placed inside of it already. The reason it applies to main clauses only is that in
embedded clauses the complementizer obligatorily hosts the enclitics, and in interro-gatives the obligatory host is either the polar interrogative complementizer or the \textit{wh} expression. The rule doesn’t alter the semantics or the tectogrammatical type of the sign. Phenogrammatically, it constructs a different set of strings of languages than the input one by reaching into the first string of languages, extracting the first phonological word out of it, then constructing a length one string of languages out of that initial phonological word. Now, the enclitic cluster can encliticize into the first phonological word, that is, we get unrestricted 1W placement.

Also note that because of how the function \texttt{br1} is defined, an initial constituent which consists of 3 or more words can never be broken up into two \texttt{S}-strings consisting of one \texttt{S}-string built out of the first two or more phonological words, and the second one built out of the remainder. It always breaks up the initial constituent into an \texttt{S}-string built out of the first phonological word, and the remainder. This is just as desired.

\section{Conclusion}

In this chapter, we gave our theory of encliticization in Serbo-Croatian. The general approach involved treating enclitics as functions looking for sentence with the right kind of gap, and then attaching themselves to the last phonological word out of which the initial length one \texttt{S}-string of their argument is constructed. In contrast to ordinary phenogrammatical combination of expressions which works
at the relatively high level of languages or strings of languages, clitics attach to their hosts at a deeper level, that of phonological words. So, the grammar correctly represents the fact that encliticization builds new phonological words. At the same time, because of how deeply they attach to their hosts, no subsequent ordinary phenogrammatical combination can rip apart the new phonological word created by encliticization.

The number parameter of the tectogrammatical family of sentence types was crucial for enforcing the ordering of enclitics in the clitic cluster, which is itself an idea we borrowed from Morrill and Gavarró (1992) and Kraak (1998). Because the number parameter keeps a very precise record of which slots in the clitic cluster have been filled, we were able to also account for the phenomenon of the clitic je disappearing in presence of se. This phenomenon is not definable semantically or tectogrammatically since it pertains to both the inherent and the true reflexive, and simply depends on whether the penultimate slot in the enclitic cluster is occupied or not. This is precisely what our analysis of this phenomenon depends on as well.

Finally, while we did introduce non-logical rules to get the details of the clitic cluster placement right, it is worth mentioning that in our grammar, there is no difference between 1C and 1W placement of clitics. In both cases, the clitics work the same way, encliticizing onto the last phonological word in the first length one string of languages. The task then is to correctly pick out the class of expressions
which can be treated as initial length one $S$-strings in a clause, and we more or less did that.

There is an outstanding problem concerning the enclitic cluster placement. Namely, in Serbo-Croatian the enclitics are trapped in finite clauses, i.e. there is typically no clitic climbing. However, at this point our grammar cannot distinguish between, say, a sentence missing a local object and the one missing a more deeply embedded object. This makes it possible for a clitic which is an argument of a downstairs verb to occur in the upstairs clitic cluster.

This problem is related to the problem with control, discussed in the previous chapter. In both cases, it is the failure of grammar to distinguish locally missing arguments from more deeply embedded missing arguments that leads to over-generation. We return to both these outstanding issues in Chapter 7.
Chapter 6: Interrogatives

6.1 Introduction

In this chapter, we discuss interrogative clauses in Serbo-Croatian. We first present some data concerning both polar questions and constituent or *wh* questions and then present our analysis.

With respect to constituent questions, we consider a couple of different generalizations concerning word order in multiple *wh* questions. Apart from word order, we also investigate scoping possibilities of *wh* expressions in Serbo-Croatian, which have not been subject to much analysis to the best of our knowledge.

6.2 Data

6.2.1 Polar Questions

There are at least two different strategies for forming polar interrogatives in Serbo-Croatian, which we refer to as verb first and *da* strategy. In both cases the
enclitic interrogative complementizer *li* must occur in the question, and the difference lies in whether *li* encliticizes onto a clause initial finite verb (verb first), or a clause initial *da*.

**da Strategy**

In polar interrogatives with *da*, *da* must occur clause-initially, and hosts the interrogative complementizer *li*. Recall that *li* is itself an enclitic but because of its extremely limited distribution we decided not to treat it on a par with other enclitics in the cluster (see Chapter 5). So, *li* is for us essentially a 0th slot enclitic which must attach to *da* in polar interrogatives, with all the other enclitics attaching to *da* *li*.

Polar questions formed using *da* have roughly the same shape as embedded declarative clauses, except for the addition of *li*. Also like embedded declaratives, so long as the polar question remains contiguous and the complementizer occurs clause-initially, the remainder of the interrogative clause can freely order. Unlike declaratives, where we must distinguish between main and embedded clauses, there is no distinction between main and embedded interrogatives.

The examples below illustrate these generalizations:

(297) no difference between main and embedded questions:

a. Da li Ana spava?
   DA LI Ana$_{\text{NOM},f,sg,3}$ sleep$_{sg,3}$
   ‘Is Ana sleeping?’

b. Pitam se da li Ana spava?
   ask$_{sg,1}$ REFLEDA LI Ana$_{\text{NOM},f,sg,3}$ sleep$_{sg,3}$
'I wonder whether Ana is sleeping.'

(298) *da li* occurs clause-initially:
   a. Da li Ana spava?
      DA LI Ana_{NOM,f,sg,3} sleep_{sg,3}
      ‘Is Ana sleeping?’
   b. *Spava da li Ana?
   c. *Ana da li spava?

(299) other enclitics attach to *da li*:
   a. Da li mu ga je Ana kupila?
      DA LI he_{DAT,m,sg} it_{ACC,m,sg} is_{sg,3} Ana_{NOM,f,sg,3} buy_{ppl,f,sg}
      ‘Did Ana buy him that?’
   b. *Da li Ana mu ga je kupila?
   c. *Da li kupila mu ga je Ana?

(300) so long as *da li* and the enclitics are clause initial, the rest of the material can freely order inside the interrogative clause:
   a. Da li mu je Ana kupila poklon?
      DA LI he_{DAT,m,sg} is_{sg,3} Ana_{NOM,f,sg,3} buy_{ppl,f,sg} present_{ACC,m,sg}
      ‘Did Ana buy him a present?’
   b. Da li mu je poklon kupila Ana?
   c. Da li mu je poklon Ana kupila?
   d. Da li mu je kupila Ana poklon?
   e. Da li mu je kupila poklon Ana?
   f. Da li mu je Ana poklon kupila?

**Verb First Strategy**

Instead of constructing a polar question by making *da* clause-initial and attaching *li* and the rest of enclitics onto it, it is also possible to place a non-clitic finite verb on the left edge of the clause and attach *li* and other enclitics onto it. The generalizations are largely the same as for the *da* question forming strategy—there is no distinction between embedded and main clause questions, and so long
as the finite verb with the clitic cluster attached occurs clause-initially, the rest of the clausal material can freely order.

(301) no difference between main and embedded questions:
   a. Spava _li_ Ana _spava?_
      sleep_{sg,3} LI Ana_{NOM,f,sg,3}
      ‘Is Ana sleeping?’
   b. Pitam se _spava _li_ Ana?
      ask_{sg,1} REFL sleep_{sg,3} LI Ana_{NOM,f,sg,3}
      ‘I wonder whether Ana is sleeping.’

(302) only finite verbs can host _li_ in polar questions:
   a. _Hočeš _li_ doći?
      will_{sg,2} LI come_{inf}
      ‘Will you come?’
   b. * Doći _li_ (ho)češ?
   c. _Jesi _li_ došao?
      are_{sg,2} LI come_{ppl,m,sg}
      ‘Did you come?’
   d. * Došao _li_ (je)si?

(303) other enclitics attach to the finite verb + _li_:
   a. Jeste _li_ mi _ga_ poslali _juče?
      are_{pl,2} LI _DAT,sg it_{ACC,m,sg} send_{ppl,m,pl} yesterday
      ‘Did you send it to me yesterday?’
   b. * Jeste _li_ poslali mi _ga_ _jučer?
   c. * Jeste _li_ _jučer _mi_ ga poslali?

(304) so long as the verb with attached enclitics is clause initial, the rest of the material can freely order inside the interrogative clause:
   a. Jeste _li_ mi _poslali_ _paket_ _juče?
      are_{pl,2} LI _DAT,sg send_{ppl,m,pl} package_{ACC,m,sg} yesterday
      ‘Did you send me the package yesterday?’
   b. Jeste _li_ mi _poslali_ _juče_ _paket?
   c. Jeste _li_ _mi_ _juče_ _poslali_ _paket?

214
6.2.2 Constituent Questions

Serbo-Croatian constituent questions have attracted a lot of attention in mainstream generative grammar (Rudin, 1988a,b; Bošković, 1997, 2002; Stjepanović, 2003; Progovac, 2005), and some attention in other traditions including HPSG (Penn, 1999b) and Type Logical Categorical Grammar (Vermaat, 2005). However, we believe that correct empirical generalizations and consequently an adequate analysis of the who questions in Serbo-Croatian are still lacking. Below we briefly describe main properties of Serbo-Croatian constituent questions, noting when our generalizations (dis)agree with the ones reported in the literature. We first discuss word order and then scope of who expressions.

Word Order

In Serbo-Croatian, on a non-echo question reading, each constituent question has to have exactly one who expression on its left periphery, preceding the enclitic cluster:

(305) a. Koga je Ana vidjela?
   who ACC,sg is sg,3 Ana NOM, f, sg,3 see ppl, f, sg
   ‘Who did Ana see?’

b. * Ana je koga vidjela?

c. * Ana je vidjela koga?

215
The fronted *wh* expression which occurs to the left of the clitic cluster could be a *wh* pronoun as in the preceding examples, or a noun phrase containing a *wh* determiner. In the latter case, the *wh* determiner must occur leftmost in the fronted noun phrase.

It is also possible for just the *wh* determiner to occur to the left of the clitic cluster, with the remainder of the noun phrase occurring somewhere to the right of the enclitic cluster, for example:

A more restrictive set of judgments requires that the rest of the noun phrase must occur immediately to the right of the clitic cluster when the *wh* determiner hosts the enclitics. According to that set of judgments, the example (c) above is ungrammatical.
In a multiple constituent question, one *wh* expression must be fronted as described above, and the other ones must occur somewhere to the right of the enclitic cluster. Since, in general, the verb and its nominal arguments can freely reorder in a clause, the non-fronted *wh* expressions may occur immediately to the right of the clitic cluster, but may also occur further toward the right edge of the clause.

This is somewhat controversial, in that it is usually maintained that any other *wh* expressions must occur *immediately* to the right of the clitic cluster. However, it is also typically simultaneously noted that this is not strictly speaking the case, e.g. Rudin (1988a,b); Bošković (1997); Progovac (2005). We will show how to force all *wh* expressions in a clause to occur on the left periphery, with the clitic cluster encliticizing onto the first one, in accordance with this empirical generalization.

In a multiple constituent question, any *wh* expression can occur to the left of the enclitic cluster, that is, Serbo-Croatian shows no Superiority effects (Rudin (1988a,b); Mihalicek (2008, 2010a)). Any other *wh* expressions in a constituent question are also freely ordered with respect to one another. For example, we accept all of the following:

(309) a. Ko je kome šta dao?  
\[ \text{who}_{\text{NOM},m,sg} \text{ is}_{sg,3} \text{ who}_{\text{DAT},m,sg} \text{ šta}_{\text{ACC},n,sg} \text{ give}_{ppl,m,sg} \]  
‘Who gave what to whom?’

b. ko je šta kome dao?

c. Kome je ko šta dao?

d. Kome je šta ko dao?

31See Bošković (1997, 2001); Stjepanović (2003) for a different empirical generalization concerning the order of *wh* expressions in a clause.
e. Šta je kome ko dao?
f. Šta je ko kome dao?

Scope

Below we discuss certain facts concerning the scope of *wh* expressions in Serbo-Croatian, that haven’t been discussed in the literature previously to the best of our knowledge (modulo the unpublished Mihalicek (2008)). Consider the following Baker ambiguous constituent questions:

(310)  
a. Ko se pita šta smo kome kupili?
     who\textsubscript{NOM,sg} wonders\textsubscript{sg,3} what\textsubscript{ACC,pl} are\textsubscript{1} who\textsubscript{DAT,pl} buy\textsubscript{ppl,pl}  
b. Ko se pita kome smo šta kupili?
     who\textsubscript{NOM} wonders\textsubscript{sg,3} who\textsubscript{DAT} are\textsubscript{1} what\textsubscript{ACC} bought

‘Who wonders what we bought for who?’

Both of these questions, which differ in terms of which *wh* expression occurs immediately to the left of the clitic cluster in the embedded clause, are three way ambiguous in exactly the same way. This is indicated by the fact that all of the following are possible answers to both questions above:

(311)  
a. Dragan.
     [both *šta* and *kome* have embedded scope]
     b. Dragan se pita šta ste njemu kupili, a Gordana se pita šta ste kupili Marku.
        ‘Dragan wonders what you bought for him, and Gordana wonders what you bought for Marko’
        [*šta* has embedded scope, *kome* has root scope]
     c. Dragan se pita kome ste kupili televizor, a Gordana se pita kome ste kupili kompjuter.
Dragan wonders who you bought the TV for and Gordana wonders who you bought the computer for. 

[šta has root scope, kome has embedded scope]

So while exactly one *wh* expression per question must occur on the left periphery, in a multiple *wh* question any *wh* expression, including the fronted one, can outscope the other *wh* expressions. However, *wh* expressions in Serbo-Croatian cannot in general have arbitrary scope. They cannot scope lower than their surface position indicates:

(312) a. Ko Marija zna da će doći?

`whoNOM,m,sg MarijaNOM,f,sg,3 knowssg,3 DA willsg,3 arriveinf`

‘Who does Marija know will arrive?’

# ‘Marija knows who will arrive’

While the root predicate is compatible with both an interrogative or a declarative complement (syntactically speaking), this example shows that *ko* ‘whoNOM’ has to have root scope.

*Wh* expressions also cannot in general scope higher than their surface position indicates:

(313) Marija zna ko će doći.

`MarijaNOM,f,sg,3 knowssg,3 whoNOM,m,sg willsg,3 arriveinf`

# ‘Who does Marija know will arrive?’

‘Marija knows who will arrive’

(314) Marija se pita ko će kome kupiti

`MarijaNOM,f,sg,3 wonderssg,3 whoNOM,m,sg willsg,3 whoDAT,m,sg buyinf poklon. presentACC,m,sg`

‘Marija wonders who will buy a present for who’
So, the only way for a *wh* expression to scope higher than its surface position indicates is if it occurs on the left periphery of a question embedded inside a root constituent question.

### 6.3 Analysis

#### 6.3.1 Semantic Assumptions

In this section, we sketch our hyperintensional theory of interrogative meanings, which is largely based on Pollard (2008b) (see also Mihalicek and Pollard (2012)). Here, we depart slightly from these earlier presentations of the theory with respect to constituent question meanings. We first sketch the analysis of polar question meanings, which is the same as in Pollard (2008b) and Mihalicek and Pollard (2012), and then the analysis of constituent question meanings. For expository purposes, we sketch how the theory of interrogative meanings works using English examples.

**Polar Question Meanings**

We analyze polar questions (meanings of both root and embedded polar interrogative clauses) as having extensions which are singleton sets of true answers, just like Karttunen (1977). On our hyperintensional approach, this means that polar questions have the type $\mathbf{p} \rightarrow \mathbf{p}$; so that the extension at some world $w$ is
then a set of propositions \((p \to t)\) – intuitively, the set of true answers to it. Thus, e.g. whether Ana slept or Did Ana sleep? denotes at some \(w\) a set with exactly one member: either the proposition that Ana slept or that she didn’t, whichever is true at \(w\). We abbreviate the polar question type \(p \to p\) as \(k\).

\[
\text{(315)} \quad k =_{def} p \to p
\]

Now we introduce the constant \(\vdash\text{whether} : p \to k\) together with the following meaning postulate (non-logical axiom):

\[
\text{(316)} \quad \vdash\text{whether} : p \to k
\]

\[
\text{(317)} \quad \text{whether} =_{def} \lambda_{qp}.[p \text{ and } ((p \text{ eq}_p q) \text{ or } (p \text{ eq}_p \neg q))]\]

In the definition of whether we made use of the propositional connectives and, or and not that translate the English sentential connectives \(\text{and}, \text{or} \text{ and } \neg \text{ the case that}\). The following theorems (which follow directly from the facts that (i) the propositions form a preboolean algebra, and (ii) worlds are ultrafilters) relate these propositional connectives to their extensional counterparts:

\[
\begin{align*}
\text{(318)} \quad \text{a. } & \vdash \forall_{pqw}[ (p \text{ and } q)@w = (p@w \land q@w)] \\
\text{b. } & \vdash \forall_{pqw}[ (p \text{ or } q)@w = (p@w \lor q@w)] \\
\text{c. } & \vdash \forall_{pw}[ (\neg p)@w = \neg(p@w)]
\end{align*}
\]

We also made use of the constant \(\text{eq}_p\) (we omit the subscript when the type is clear from the context). This is one of a family of constants \(\text{eq}_S\) of type \(S \to S \to p\) which are used to express, for each hyperintensional meaning type \(S\), propositions that two meanings of type \(S\) are one and the same meaning. The
following meaning postulate states that at any world $w$, the extension of $\text{eq}_S$ at $w$ is the ordinary equality relation on things of type $S$:

\[(319) \vdash \forall_{wxy}[(x \text{eq} y)@w = (x = y)]\]

To give a concrete example, the meaning of a polar question such as *Did Ana sleep?* or *whether Ana slept* would be represented in the grammar as follows:

\[(320) \vdash \lambda_q.q \text{ and } (q \text{eq}_p(\text{sleep ana}) \text{ or } q \text{eq}_p \text{not(}\text{sleep ana)}) : k\]

### Constituent Question Meanings

In Pollard (2008b) and Mihalicek and Pollard (2012) n-ary constituent questions were analyzed as denoting curried functions from $n$ individuals to a singleton set of propositions. Here, we follow a simplification of that theory developed by Carl Pollard\(^{32}\), whereby constituent questions denote sets of true propositions, just like polar questions. That is, we analyze constituent questions as denoting the range of the extension of constituent question meanings in the older version of the theory.

Just as in the older version of the theory, each $wh$ expression can be associated with two different hyperintensional terms. Filler $wh$ expressions convert propositions with a bound individual variable (type $e \rightarrow p$) into a question (type $k$), while *in situ* $wh$ expressions take an argument of type $e \rightarrow k$ (a question with a bound individual variable) and output a question (type $k$).

\(^{32}\)Personal communication to the author, June 9, 2012.
Below we define two hyperintensional constants, which\(_m\) and which\(_i\), corresponding to the filler and \textit{in situ} versions of the determiner \textit{which}.

\begin{align*}
\text{(321)} & \quad \vdash \text{which\(_m\)} : (e \to p) \to (e \to p) \to k \\
& \quad \text{which\(_m\)} =_{\text{def}} \lambda P Q. \text{which\(_i\)} P (\lambda x. \text{whether } (Q x)) \\
\text{(322)} & \quad \vdash \text{which\(_i\)} : (e \to p) \to (e \to k) \to k \\
& \quad \text{which\(_i\)} =_{\text{def}} \lambda P q. \text{exist } P (\lambda x h q)
\end{align*}

For example, the meaning of a unary \textit{wh} question such as \textit{Which student slept?} is represented as follows:

\begin{align*}
\text{(323)} & \quad \vdash \text{which\(_m\)} \text{ student } (\lambda x. \text{sleep } x) : k \\
& \quad \leadsto \vdash \lambda q. \text{exist student } (\lambda x q \text{ and } (q \text{ eq } p (\text{sleep } x) \text{ or } q \text{ eq } p \text{ not } (\text{sleep } x))) : k
\end{align*}

The meaning of a binary constituent question such as \textit{Which student likes which cat?} is represented as follows:

\begin{align*}
\text{(324)} & \quad \vdash \text{which\(_i\)} \text{ cat } (\lambda y. \text{which\(_m\)} \text{ student } (\lambda x. \text{like } y x)) : k \\
& \quad \leadsto \vdash \lambda q. \text{exist cat } (\lambda y. \text{exist student } (\lambda x q \text{ and } (q \text{ eq } p (\text{like } y x) \text{ or } q \text{ eq } p \text{ not } (\text{like } y x))) : k
\end{align*}

The meanings of \textit{wh} pronouns \textit{who} and \textit{what} are represented by the following pairs of hyperintensional constants:

\begin{align*}
\text{(325)} & \quad \vdash \text{who\(_m\)} : (e \to p) \to k \\
& \quad \text{who\(_m\)} =_{\text{def}} \text{ which\(_m\)} \text{ person} \\
& \quad \vdash \text{who\(_i\)} : (e \to k) \to k \\
& \quad \text{who\(_i\)} =_{\text{def}} \text{ which\(_i\)} \text{ person} \\
\text{(326)} & \quad \vdash \text{what\(_m\)} : (e \to p) \to k \\
& \quad \text{what\(_m\)} =_{\text{def}} \text{ which\(_m\)} \text{ thing} \\
& \quad \vdash \text{what\(_i\)} : (e \to k) \to k \\
& \quad \text{what\(_i\)} =_{\text{def}} \text{ which\(_i\)} \text{ thing}
\end{align*}
6.3.2 Polar Questions

Tectogrammatically, we analyze all questions, polar and constituent, main and embedded, as being of type $S_{q,T}$ since all interrogatives have the same distribution in Serbo-Croatian.

*da Strategy*

For the *da* question forming strategy, we simply give the following lexical entry for *da li*:

$$
\lambda_{Z,v}. \exists_v (Z \cdot v) \land w = (t_{oZ} \lambda_{Z,s} = da#li_s) \circ v : Z \rightarrow Z; S_{m,6} \rightarrow S_{q,6}; \rightarrow\mathbf{whether : p} \rightarrow k
$$

Tectogrammatically, it’s looking for a declarative main clause inside of which no clitics have been placed, and outputs a question. Semantically, the meaning of *da li* is represented by the previously defined constant $\vdash \mathbf{whether : p} \rightarrow k$.

Phenogrammatically, *da* comes with *li* encliticized onto it, and *da li* is required to occur clause initially. Whatever order exists within the declarative clause that *da li* takes as its argument, it will be retained in the resulting question. Given that *da li* is tectogrammatically required to combine with clauses inside of which no clitics have been placed, and to occur question-initially, the enclitics have to attach to *da li*. The tectogrammatical types of enclitics just need to be further generalized with respect to the value of the k parameter to allow them to be placed inside questions.

The polar question *Da li Ana spava?* ‘Is Ana sleeping?’ is represented in the grammar as follows:
The grammar predicts that this question is pronounceable two different ways, depending on the order of Ana and spava, just as desired. Semantically, this question denotes a set which contains exactly one proposition, that Ana sleeps or that she doesn’t, whichever happens to be true at a given world.

**Verb First Strategy**

Recall that one can also form polar questions in Serbo-Croatian by placing the finite verb, with li encliticized onto it, clause-initially. In order to represent this in the grammar, we will have to list interrogative versions of finite verbs that form questions and not declarative sentences.

We cannot write a non-logical rule mapping the set of finite verbs to their interrogative counterparts because we cannot pick out all the finite verbs as represented in the grammar—they differ in terms of their tectogrammatical, phenogrammatical and semantic type. We could write several different non-logical rules, one for each kind of finite verb depending on its argument structure: one rule for intransitive verbs, one for transitive verbs, etc. We do not pursue that here, however, and just assume that we have interrogative versions of finite verbs listed in the lexicon.

To construct the question Spava li Ana? ‘Is Ana sleeping?’, we need the following lexical entry for the verb:
After it combines with the subject, we get the following sign:

$$\vdash \lambda x. \exists v (\text{PER } x) \land w = (\text{toZ } \lambda z. s = \text{spava#li}_s) \circ v : z \rightarrow z;$$

$\text{NP}_{\text{nom}, f, \text{sg}, 3} \circ S_{q, 6}; \lambda x. \text{whether(sleep } x) : e \rightarrow k$

The verb permutes its subject which makes no difference here where the subject is a single phonological word, but would make a difference in the case of subjects which consist of multiple flexibly ordered words, such as an adjective and a noun. The verb with li encliticized onto it is required to occur question-initially, so in the case of Spava li Ana? ‘Is Ana sleeping?’ there is only one possible pronunciation. Semantically and tectogrammatically this question is the same as its da version, Da li Ana spava?.

6.3.3 Constituent Questions

Unary Constituent Questions

We start by showing how to analyze unary constituent questions in Serbo-Croatian, containing a wh pronoun or a determiner. In a unary wh question, the wh pronoun must host the enclitic cluster, while in the case of wh determiners, either the determiner or the last word of the phrasal noun that the determiner combines with hosts the clitic cluster.

Below we give the lexical entry for ko ‘who$^{\text{NOM}}$’ and koga ‘who$^{\text{ACC}}$’

$$\vdash \lambda x. \exists v (x \in S) \land w = \text{KO}_{z} \circ v : (z \rightarrow Z) \rightarrow Z;$$

$$(\text{NP}_{\text{nom}, m, \text{sg}, 3} \circ S_{m, 6}) \circ S_{q, 6}; \text{who}_{m} : (e \rightarrow p) \rightarrow k$$

226
A *wh* pronoun combines with a declarative clause missing an appropriate argument. Phenogrammatically, its argument clause contains a bound variable which the pronoun gets rid of by feeding it the null $S$-string $e_S$. The pronoun then preposes itself to the remainder of the clause. We borrow this mechanism for fronting *wh* expressions from Muskens (2007b). This way, the grammar ensures that the *wh* pronoun is the first string of languages in the question so that it can eventually host the clitic cluster.

Below we show how to construct a representation of the question *Koga Ana voli? ‘Who does Ana love?’*. We omit phenogrammatical and semantic types and non-essential tectogrammatical parameters. First, an appropriate accusative trace is introduced and the verb combines with it.

(333)  
\[
\frac{\exists x \in \mathcal{V} \left( x \in S \land v \right) \land w = \text{KOGA}_z \circ v : (z \rightarrow z) \rightarrow z;}{\lambda x \in \mathcal{V} \circ \text{PER}(w \circ \text{VOLI}_z \circ x) \circ \text{NP}_{\text{acc}} \circ \text{NP}_{\text{nom}} \circ \text{S}_{\text{m,6}}; \text{love}}
\]

Next, the verb phrase combines with the subject noun phrase, and the accusative trace is bound.

(334)  
\[
\frac{\lambda x \in \mathcal{V} \circ \text{PER}(w \circ \text{VOLI}_z \circ x) \circ \text{NP}_{\text{nom}} \circ \text{S}_{\text{m,6}}; \text{love}}{\lambda x \in \mathcal{V} \circ \text{PER}(\text{ANA}_z \circ \text{VOLI}_z \circ x) \circ \text{S}_{\text{m,6}}; \text{love} \ x \ \text{ana}}
\]
Now, *koga* ‘who_{ACC}’ can combine with the clause with the accusative trace, resulting in the following unary question:

(335) $\vdash \lambda w.\exists v.\left(\text{PER}(\text{ANA}_{z} \circ \text{VOLI}_{z}) \circ v\right) \wedge w = \text{KOGA}_{z} \circ v : \text{Z}; S_{q,6} ;$

$\text{who}_{m} \left(\lambda x.\text{love } x \text{ ana} \right) : \text{k}$

This question is predicted to be pronounceable two different ways, namely *Koga Ana voli?* and *Koga voli Ana?*

Now we analyze unary questions with a *wh* determiner. Recall that the *wh* determiner must occur leftmost in the clause, and the clitic cluster may be hosted by the determiner or by the last word in the determiner’s first argument. We start with the following lexical entry for *koja* ‘which’:

(336) $\vdash \lambda v,w.\exists z.\left(\lambda v.\left(\text{e}_v \in \text{s}_v \circ v\right) \wedge w = \text{toZ} \left(\text{L} \left(\text{KOJA}_{z} \circ y\right)\right) \circ v : \text{Z} \rightarrow \text{z} \rightarrow \text{Z} \rightarrow \text{Z} ; \text{NP}_{\text{nom,f,sg,3}} \circ \circ \left(\text{NP}_{\text{nom,f,sg,3}} \circ \circ \text{S}_{m,6}\right) \circ \circ \text{S}_{q,6} ; \text{which}_{m} : \left(\text{e} \rightarrow \text{p} \right) \rightarrow \left(\text{e} \rightarrow \text{p} \right) \rightarrow \text{k}$

Tectogrammatically, its first argument is a noun which has undergone the [NC] rule (see Chapter 3). This rule changes the tectogrammatical type of a noun into a corresponding noun phrase type, without changing its semantic type. This is to allow for the presence of postnominal modifiers inside the first argument. The determiner agrees with its first argument in case, number and gender. Its second argument is a declarative clause with a bound trace of the same tectogrammatical type as its first argument.

Semantically, it’s represented as the hyperintensional constant \text{which}_m.

Phenogrammatically, it works just like a *wh* pronoun except that it constructs a length one string of languages out of itself concatenated with its first argument,
which is required to occur clause-initially. So this lexical entry will generate questions in which the whole noun phrase containing the \( wh \) determiner hosts the enclitic cluster, e.g. \( Koja \text{ djevojka je došla?} \) ‘Which girl arrived?’

To generate questions in which the determiner hosts the enclitic cluster, we can construct a question with the lexical entry above, and then use the rule [1W] (see Chapter 5) to split the initial length one string of languages into a length two string of languages. This way, we obtain questions such as \( Koja \text{ je djevojka došla?} \) ‘Which girl arrived?’

So the lexical entry above takes care of the less permissive set of judgments which allows the \( wh \) determiner and its first argument to be either contiguous or separated by the enclitic cluster.

A more permissive set of judgments discussed earlier in the chapter allows the first argument of the \( wh \) determiner to occur anywhere further to the right of the enclitic cluster if the determiner hosts the enclitics. For the more permissive grammar, we assume the following lexical entry for \( koja \):

\[
\begin{align*}
\lambda_{y,w}.w & \exists_v (x (t \text{o} Z, L y) v) \land w = KOJA_z \circ v : z \rightarrow (z \rightarrow z) \rightarrow z; \\
\text{NP}_{\text{nom},f,sg,3} & \rightarrow (\text{NP}_{\text{nom},f,sg,3} \rightarrow S_{m,6}) \rightarrow S_{q,6}; \\
\text{which}_m & : (e \rightarrow p) \rightarrow (e \rightarrow p) \rightarrow k
\end{align*}
\]

Semantically and tectogrammatically it works the same as the one given above for the more restrictive grammar. Phenogrammatically, instead of feeding the ‘gappy’ clause the empty \( S \)-string, it feeds it the length one string of languages constructed out of its first argument. This makes it possible for its first argument to reorder
with respect to the other clausal constituents while the *wh* determiner remains clause-initial and hosts the enclitic cluster.

While the less permissive lexical entry and the rule [1W] are sufficient to account for the more restrictive set of judgments, both sets of lexical entries for *wh* determiners are needed for the less restrictive grammar. This is because the rule [1Z], as formulated in Chapter 5, only fuses quantificational noun phrases into length one strings of languages to account for 1C placement of the enclitic cluster.

**Multiple Constituent Questions**

Here we will explore two different generalizations mentioned earlier in the chapter, namely: (i) all *wh* expressions must front with the enclitic cluster occurring after the first one; and, (ii) only one *wh* expression must front, with any others occurring somewhere to the right of the enclitic cluster.

To account for the first generalization, we simply assume that all *wh* expressions have phenogrammatical terms like the ones for the *wh* expressions in unary constituent questions. That is, both filler *wh* expressions (semantic type \((e \rightarrow p) \rightarrow k\)), and the *in situ* ones (semantic type \((e \rightarrow k) \rightarrow k\)) are required to front, phenogrammatically.\(^{33}\)

Consider the question *Ko kome daje poklon?* ‘Who is giving a present to whom?’ According to this set of judgments, the two *wh* expressions can occur in either relative order and have to occur to the left of some permutation of the two non-*wh*

\(^{33}\)We will continue to talk about filler and *in situ* *wh* expressions, using these terms to refer to two different semantic types of *wh* phrases, \((e \rightarrow p) \rightarrow k\) and \((e \rightarrow k) \rightarrow k\) respectively.
words. In our grammar, this question is represented by two signs, one in which ko ‘whonom’ is in situ and left-most, and the other in which kome ‘whodat’ is in situ and leftmost. We need the following two sets of lexical entries:

(338) a. \( \lambda x_y. \exists v (x \in s \, v) \land w = K_0_z \circ v : (z \to z) \to z; \)
\( (NP_{nom,m,sg,3}) \circ (S_{m,6}) \circ (S_{q,6}); who : (e \to p) \to k \)

b. \( \lambda x_y. \exists v (x \in s \, v) \land w = K_0_z \circ v : (z \to z) \to z; \)
\( (NP_{dat,m,sg,3}) \circ (S_{m,6}) \circ (S_{q,6}); who : (e \to k) \to k \)

(339) a. \( \lambda x_y. \exists v (x \in s \, v) \land w = K_0_z \circ v : (z \to z) \to z; \)
\( (NP_{dat,m,sg,3}) \circ (S_{m,6}) \circ (S_{q,6}); who : (e \to p) \to k \)

b. \( \lambda x_y. \exists v (x \in s \, v) \land w = K_0_z \circ v : (z \to z) \to z; \)
\( (NP_{nom,m,sg,3}) \circ (S_{q,6}) \circ (S_{q,6}); who : (e \to k) \to k \)

(338a) gives the lexical entry for the wh filler version of ko ‘whonom’, meaning that it takes a sentence with a bound trace (semantically, of type \( e \to p \)), and outputs a question (semantically of type \( k \)). (338b) is the lexical entry for the in situ version of kome ‘whodat’, meaning that it takes a question with a bound trace (semantically, of type \( e \to k \)), and outputs another question (semantically of type \( k \)).

In (339), we give the in situ version of ko ‘whonom’ and the filler version of kome ‘whodat’. Phenogrammatically, however, all of these wh expressions front.

Using the lexical entries in (338), we can construct the following sign:

(340) \( \lambda w. \exists v v \left( \text{PER} (DAJE_z \circ POKLON z) \, v \right) \land (v' = KO_z \circ v) \land (w = KOME_z \circ v') : k \)

Given the phenogrammatical terms of wh expressions in this version of the grammar, which force them to front, wh words stack at the left edge of the question, hence, the in situ one, which applies after the filler, occurs left-most in the
utterance. In this sign $kome$ ‘who$_{DAT}$’ has to be left-most, followed by the other $wh$ word, followed by some permutation of the verb and the accusative noun phrase. If we had picked a sentence with enclitics, the enclitics would have to attach to the left-most $wh$ expression.

Similarly, using the lexical entries in (339), we can construct the following sign corresponding to the two pronunciations of the question in which $ko$ is left-most, i.e. $Ko$ $kome$ $daje$ poklon? ‘Who is giving a present to whom?’:

(341) $\vdash \lambda_{w, \exists_{\nu\nu}}(\text{PER}(\text{DAJE}_{z} \circ \text{POKLON}_{z}) \nu) \wedge (\nu' = KOME_{z} \circ v) \wedge (w = KO_{z} \circ v') : Z; S_{q, 6}; \text{who}_{1}(\lambda_{y, \text{who}_{m}(\lambda_{x, \text{exists present}(\lambda_{z, \text{give } z y x)}))}) : k$

The other generalization, whereby there is no real multiple $wh$ fronting in Serbo-Croatian and only one $wh$ expression must front, requires the grammar to distinguish between filler and $in$ $situ$ $wh$ words phenogrammatically. Whereas the fillers are required to front and occur on the left periphery of the clause, the $in$ $situ$ $wh$ expressions are lowered into the appropriate gap in the utterance, just like ordinary quantificational noun phrases (see Chapter 3).

According to this set of judgments, $Ko$ $daje$ poklon $kome$? ‘Who is giving a present to whom?’ can be pronounced six different ways, with $ko$ ‘who$_{NOM}$’ occurring question-initially and the other three expressions freely permuting. Similarly, $Kome$ $ko$ $daje$ poklon? ‘Who is giving a present to whom?’ requires only that $kome$ ‘who$_{DAT}$’ occur left-most, so it can also be pronounced six different ways.

We add the following two sets of lexical entries to account for this pattern:

(342) a. $\vdash \lambda_{X_{w}, \exists_{v}}(X \in S \nu) \wedge w = KO_{z} \circ v : (z \rightarrow Z) \rightarrow Z;$

$\left(\text{NP}_{\text{nom}, m, sg, 3} \rightarrow S_{m, 6} \rightarrow S_{q, 6}; \text{who}_{m} : (e \rightarrow p) \rightarrow k\right)$
b. \( \vdash \lambda_x. (X \! \mathbb{E}_{KOME} z) : (\mathbf{z} \rightarrow \mathbf{z}) \rightarrow \mathbf{z}; (\mathbf{NP}_{\text{dat},m,\text{sg},3} \rightarrow S_{q,6}) \rightarrow S_{q,6}; \quad \text{who}_{\mathbf{i}} : (e \rightarrow k) \rightarrow k \)

(343)  
\begin{align*}
a. \quad & \vdash \lambda_{x,w}. \exists_v (X \in S \! v) \wedge w = KOME z \circ v : (\mathbf{z} \rightarrow \mathbf{z}) \rightarrow \mathbf{z}; \\
& (\mathbf{NP}_{\text{dat},m,\text{sg},3} \rightarrow S_{m,6}) \rightarrow S_{q,6}; \text{who}_{m} : (e \rightarrow p) \rightarrow k \\
b. \quad & \vdash \lambda_{x}. (X \! \mathbb{E}_{KO} z) : (\mathbf{z} \rightarrow \mathbf{z}) \rightarrow \mathbf{z}; (\mathbf{NP}_{\text{nom},m,\text{sg},3} \rightarrow S_{q,6}) \rightarrow S_{q,6}; \\
& \quad \text{who}_{\mathbf{i}} : (e \rightarrow k) \rightarrow k \end{align*}

From the lexical entries in (342) we can construct the following sign:

(344)  
\( \vdash \lambda_w. \exists_v (\textbf{PER}(DAJE} z \circ POKLON} z \circ KOME} z \circ v) \wedge (w = \text{KO} z \circ v) : \mathbf{z}; S_{q,6}; \\
\text{who}(\lambda_x. \text{who}_{m}(\lambda_y. \text{exists present}(\lambda_z. \text{give } x y z))) : k \)

This sign represents the question in which \textit{ko} ‘who_{NOM}’ occurs left-most and the other three words can freely reorder with respect to one another. Below is the sign that corresponds to the question in which \textit{kome} ‘who_{DAT}’ occurs left-most, constructed with lexical entries in (343):

(345)  
\( \vdash \lambda_w. \exists_v (\textbf{PER}(DAJE} z \circ POKLON} z \circ \text{KO} z \circ v) \wedge (w = \text{KOME} z \circ v) : \mathbf{z}; S_{q,6}; \\
\text{who}(\lambda_y. \text{who}_{m}(\lambda_x. \text{exists present}(\lambda_z. \text{give } x y z))) : k \)

To summarize, in the less permissive grammar all \textit{wh} expressions front phonogrammatically. The initial \textit{wh} expression hosts the enclitic cluster and any other ones occur immediately to the right of the enclitic cluster.

In the more permissive grammar, the set up is more similar to English. Only one \textit{wh} expression is required to front and occur immediately to the left of the enclitic cluster, while any other ones phonogrammatically behave like ordinary noun phrases, in that they can freely reorder with respect to the verb and any of its other nominal arguments.
Baker Ambiguous Questions

In this section we extend our analysis of constituent questions to account for the patterns of inference in Baker ambiguous questions. We will only be concerned with the grammar developed on the basis of judgments that we consider accurate, i.e. the more permissive grammar that requires only one wh expression per question to front.

We first present our analysis of question embedding verbs, and discuss simpler examples of Baker ambiguity in Serbo-Croatian, for which no extensions of our analysis of constituent questions are needed. Then we turn our attention to the more complex pattern of inference, whereby a wh expression that occurs left-most in the embedded question can have matrix scope.

A typical question-embedding verb in Serbo-Croatian is *pitati se* ‘wonder’; literally ‘ask oneself’. We will treat *pitati se* as an idiom meaning ‘wonder’, with *se* an inherent reflexive (see Chapter 5). Below is the lexical entry for the third person singular present tense version of this verb.

\[
(346) \quad x : z; SE; x : e \vdash \lambda x, y, z \exists y \{ \PER (w \circ PITA z \circ x) y \} \wedge y = v \circ toZ (k x) : z \rightarrow z \rightarrow z; S_{q,6} \rightarrow NP_{nom, m, sg, 3} \rightarrow S_{m,6}; wonder : k \rightarrow e \rightarrow p
\]

This verb can freely reorder with the subject. Its complement interrogative clause is turned into a length one S-string and required to occur on the right edge of the matrix clause (analogous to embedded declarative clauses; see Chapter 4).

A sentence such as *Marko se pita ko spava* ‘Marko is wondering who is sleeping’ is represented in the grammar by the following sign:

234
(347) \( \vdash (w_\text{ac se})(\lambda x.\exists y([\text{PER}(\text{MARKO}_z \circ \text{PITA}_z) \cup y] \wedge w = y \circ \text{toZ k}(\lambda y.\exists y'([\text{PER SPAVA}_z y'] \wedge w' = \text{KO}_z \circ y'])) : \text{Z}; \)
\( S_{m,1}; \text{wonder (who}_m(\lambda x.\text{sleep } x)) \text{ marko : p} \)

Phenogrammatically, this sentence can be pronounced two ways depending on the order of Marko and pita in the main clause. The enclitic se attaches to whatever turns out to be the initial length one S-string. In the embedded clause, the wh word must be left-most.

Now we consider the critical situation, where there is one wh expression in the matrix clause and two in the embedded clause such that either wh expression in the embedded clause can have matrix scope:

(348) Ko se pita kome Marko daje
who\text{NOM,mg,} who\text{DAT,mg,} give\text{mg,}
šta?
what\text{ACC,ng,}
‘Who wonders to whom Marko is giving what?’

The sentence above is three-way ambiguous: while ko must scope at the main clause, kome and šta can both have embedded scope, or one of them can have matrix scope. More precisely, the following terms are all possible representations of the meaning of the question above:

(349) a. \( \vdash \text{who}_m(\lambda x.\text{wonder what}_i(\lambda y.\text{who}_m(\lambda z.\text{give } y z \text{ marko})) x) : \text{k} \)
[both kome and šta have embedded scope]
b. \( \vdash \text{what}_i(\lambda y.\text{who}_m(\lambda x.\text{wonder who}_m(\lambda z.\text{give } y z \text{ marko})) x) : \text{k} \)
[šta has main scope]
c. \( \vdash \text{who}_m(\lambda z.\text{who}_m(\lambda x.\text{wonder what}_i(\lambda y.\text{give } y z \text{ marko})) x) : \text{k} \)
[kome has main scope]
At present, the more permissive version of the grammar can generate two signs whose semantic component is (349a) and (349b) respectively, each of which can be pronounced as in (348). To get the sign which means (349a), we first construct the binary constituent question \( \text{kome Marko daje } \text{štta}, \) which \( \text{pita se} \) \('wonders'\) takes as its argument. Then, \( \text{ko} \) \('\text{who}_{\text{NOM}}'\) combines with the verb phrase which results in the following unary constituent question:

\[
(350) \quad \vdash (\text{wac se}) (\lambda_{yzw} \exists_{v} x (\text{PER} (\text{PITA}_{z} v)) \wedge x = \text{KO}_{z} o v \wedge w = xo) \\
\text{toZ}(k(\lambda_{w'y} \exists_{v'} [\text{PER} (\text{MARKO}_{z} o \text{DAJE}_{z} o \text{ŠTA}_{z}) v'] \wedge w' = \text{KOME}_{z} o v')) : Z \\
S_{m,1}; \text{who}_{m}(\lambda_{x,\text{wonder what}_{x}(\lambda_{y,\text{who}_{m}(\lambda_{z,\text{give y z marko})}) x}) : k
\]

We can also construct a unary constituent question with an unbound accusative trace \( \text{kome Marko daje} \), which \( \text{pita se} \) \('wonders'\) can take as its argument. Then, \( \text{ko} \) \('\text{who}_{\text{NOM}}'\) combines with the verb phrase which results in a unary constituent question. Finally, we bind the accusative trace and combine \( \text{štta} \) \('\text{what}_{\text{ACC}}'\) with that sign, resulting in the following binary question:

\[
(351) \quad \vdash (\text{wac se}) (\lambda_{yzw} \exists_{v} x (\text{PER} (\text{PITA}_{z} v)) \wedge x = \text{KO}_{z} o v \wedge w = xo) \\
\text{toZ}(k(\lambda_{w'y} \exists_{v'} [\text{PER} (\text{MARKO}_{z} o \text{DAJE}_{z} o \text{ŠTA}_{z}) v'] \wedge w' = \text{KOME}_{z} o v')) : Z \\
S_{m,1}; \text{what}_{t}(\lambda_{y,\text{who}_{m}(\lambda_{x,\text{wonder who}_{m}(\lambda_{z,\text{give y z marko})}) x}) : k
\]

The phenogrammatical terms in (350) and (351) are identical; this is because the \( \text{in situ wh} \) word \( \text{štta} \) is lowered into the gap, just like a quantificational noun phrase, so it winds up in the same place regardless of whether it scopes at the embedded or the matrix level. In general, an \( \text{in situ wh} \) word can scope arbitrarily high so long as it scopes over a constituent that is semantically and syntactically already

236
a question with a bound trace, because phenogrammatically it’s lowered into the correct slot.

However, we cannot at present generate the reading in (349c) for the sentence in (348), where the *wh* word which occurs on the left periphery of the embedded clause scopes at the main clause level. This is because in our current setup *wh* expressions come in two flavors: (i) fillers which occur on the left periphery or (ii) *in situ* expressions which are lowered into their complement’s gap. Therefore, it’s not possible for a *wh* expression in our inventory so far to occur on the left periphery of a clause but simultaneously have scope properties of *in situ* *wh* expressions, i.e. scope higher than its surface position indicates.

To get the inference in question and allow a filler *wh* expression to scope higher than its surface position, we introduce a set of special *wh* words looking for arguments which are syntactically and semantically questions missing a filler *wh* expression. Below is a lexical entry for such a version of *kome* ‘*who*_{DAT}’.

\[
\lambda X.\lambda_{Gw}.\exists_v\left(\left(G \in S \land w = KOME_z \circ v\right) \land \left((z \rightarrow z) \rightarrow z\right) \rightarrow z; \\
\left((\left(NP_{dat,m,sg,3} \rightarrow S_{m,6}\right) \rightarrow S_{q,6}\right) \rightarrow S_{q,6}\right) \rightarrow S_{q,6}; \\
\lambda_{Q,who}(\lambda X, Q(\lambda_{P,whether}(P x))): \left((e \rightarrow p) \rightarrow k\right) \rightarrow k
\]

The phenogrammatical term of this sign ensures that *kome* ‘*who*_{DAT}’ occurs on the left edge of the embedded clause. Tectogrammatically, it constructs an interrogative sentence out of an interrogative sentence missing a *wh* expression which turns declarative sentences with noun phrase gaps into interrogatives. Semantically, this *wh* expression is required to have the widest scope in the resulting
question, and the missing pieces of the filler \textit{wh} meaning in the embedded clause are added.

Now we show how to construct a representation of the question \textit{Ko se pita kome Marko daje šta?} ‘Who wonders to whom Marko is giving what?’ with \textit{kome} ‘who_{DAT}’ having the widest scope.

In constructing the embedded question, we first introduce an appropriate filler \textit{wh} trace:

\begin{equation}
K : (z \rightarrow z) \rightarrow z ; (\text{NP}_{\text{dat,m,sg,3}} \rightarrow S_{m,6}) \rightarrow S_{q,6} ; K : (e \rightarrow p) \rightarrow k \\
\vdash K : (z \rightarrow z) \rightarrow z ; (\text{NP}_{\text{dat,m,sg,3}} \rightarrow S_{m,6}) \rightarrow S_{q,6} ; K : (e \rightarrow p) \rightarrow k
\end{equation}

We construct the embedded question with this \textit{wh} trace and the actual \textit{wh} expression, the \textit{in situ} version of šta ‘what_{ACC}’, resulting in the following sign:

\begin{equation}
K : (z \rightarrow z) \rightarrow z ; (\text{NP}_{\text{dat,m,sg,3}} \rightarrow S_{m,6}) \rightarrow S_{q,6} ; K : (e \rightarrow p) \rightarrow k \\
\vdash K (\lambda_{x}.\text{PER}(\text{MARKO}_{z} \circ \text{DAJE}_{z} \circ \text{ŠTA}_{z} \circ w)) : z ; S_{q,6} ; \text{what}_{i}(\lambda_{y}.K(\lambda_{x}.\text{give } y \times \text{marko})) : k
\end{equation}

Now we can combine the embedding verb with this sign. For expository purposes, we will not use \textit{pitati se} ‘wonder’ but \textit{pitati} ‘ask’ as the embedding verb, to avoid the additional complications with the inherent reflexive. The following sign is the result of \textit{pitati} ‘ask’ combining with the sign above:

\begin{equation}
K : (z \rightarrow z) \rightarrow z ; (\text{NP}_{\text{dat,m,sg,3}} \rightarrow S_{m,6}) \rightarrow S_{q,6} ; K : (e \rightarrow p) \rightarrow k \\
\vdash \lambda_{xy}.\exists_{v}(\text{PER} (x \circ \text{PITA}_{z} v) \land y = v \circ \text{toZ} k (\lambda_{w}.\text{PER}(\text{MARKO}_{z} \circ \text{DAJE}_{z} \circ \text{ŠTA}_{z} \circ w))) : z \rightarrow z ; \text{NP}_{\text{nom,m,sg,3}} \rightarrow S_{m,6} ; \\
\text{ask} (\text{what}_{i}(\lambda_{y}.K(\lambda_{x}.\text{give } y \times \text{marko}))) : e \rightarrow p
\end{equation}

Now, the verb phrase above can combine with \textit{ko} ‘who_{NOM}’ resulting in the following sign:

238
Finally, we bind the filler \(wh\) trace, and combine the resulting sign with the special version of \(kome\)'s \(wh_{DAT}\). The resulting sign is of tectogrammatical type \(S_{q,6}\).

Below we show the step by step reduction of the phenogrammatical and the semantic term of the resulting sign.

\[(357)\]

\(\lambda x.((\lambda w'.\exists y v'(\text{PER} \text{PITA}_z v) \land v' = v \circ \text{toZ} k \ (K (\lambda w'.\text{PER}(\text{MARKO}_z \circ \text{DAJE}_z \circ \text{ŠTA}_z \circ w)))) \land w' = \text{KO}_z \circ v') : Z\)

\(\overset{\sim}{\lambda w'.\exists y v'(\text{PER} \text{PITA}_z v) \land v' = v \circ \text{toZ} k \ ((\lambda z.\exists y \text{GeS}y \land x = \text{KOME}_z \circ y) (\lambda w'.\text{PER}(\text{MARKO}_z \circ \text{DAJE}_z \circ \text{ŠTA}_z \circ w)))) \land w' = \text{KO}_z \circ v' : Z\)

\(\overset{\sim}{\lambda w'.\exists y v'(\text{PER} \text{PITA}_z v) \land v' = v \circ \text{toZ} k \ (\lambda x.\exists y (\lambda w'.\text{PER}(\text{MARKO}_z \circ \text{DAJE}_z \circ \text{ŠTA}_z \circ w))) \text{GeS} y \land x = \text{KOME}_z \circ y) \land w' = \text{KO}_z \circ v' : Z\)

\(\overset{\sim}{\lambda w'.\exists y v'(\text{PER} \text{PITA}_z v) \land v' = v \circ \text{toZ} k \ (\lambda x.\exists y (\lambda w'.\text{PER}(\text{MARKO}_z \circ \text{DAJE}_z \circ \text{ŠTA}_z)) y \land x = \text{KOME}_z \circ y) \land w' = \text{KO}_z \circ v' : Z\)

The wide scoping \(wh\) word \(kome\) is forced to occur on the left edge of the embedded clause; \(ko\) occurs on the left edge of the main clause; \(šta\) occurs somewhere in the embedded clause. The entire embedded clause is turned into a length one \(S\)-string and occurs on the right edge of the main clause.

\[(358)\]

\(\lambda w.\text{who}(\lambda w.\text{Q}(\lambda P.\text{whether}(P w))) \)

\(\lambda x.\text{who}_{m}(\lambda z.\text{ask} (\lambda y. K(\lambda x.\text{give} y x \text{marko}))) z) : k\)
Semantically, the *in situ* šta ‘what’ is forced to have the semantics of a filler *wh* expression, because it turns out to be the unique *wh* which scopes in the embedded clause. Note that:

(359) \( \text{what}_i(\lambda_y.\text{whether}(\text{give } w \text{ marko})) =_{	ext{def}} \text{what}_m(\lambda_y.\text{give } y w \text{ marko}) \)

### 6.4 Conclusion

In this chapter, we sketched a theory of interrogatives in Serbo-Croatian. We analyzed two common strategies for forming polar questions. The interrogative complementizer *li* which occurs in polar questions was not treated like other enclitics. Since its distribution is very limited, we analyzed it as already lexically encliticized onto its host, either *da* or a finite verb. Its host is required to occur clause-initially and is always a length one *S*-string, so the other enclitics have to attach to *li* and its host, which gives us the right predictions concerning the enclitic cluster placement.

With respect to constituent questions, we showed how to force multiple *wh* expressions to front, but also constructed a more permissive grammar in which
only one *wh* expression per question is required to occur on the left edge. We pursued the latter generalization, that is in accordance with our judgments, in more detail, in particular with respect to Baker ambiguous multiple constituent questions.

The theory of interrogative presented here overgenerates. Specifically, the grammar currently allows *wh* extraction out of interrogatives. In the next chapter, we address this problem, as well as other related locality issues, including control and clitic climbing.
Chapter 7: Locality Constraints

7.1 Introduction

In previous chapters, we have defined a grammar of Serbo-Croatian that overgenerates with respect to finite control, enclitic placement and \textit{wh} extraction. Specifically, the grammar incorrectly predicts that clitics can climb out of arbitrary clauses, that long distance subject control is possible, and that \textit{wh} extraction out of any clause, including polar and constituent questions, is possible.

The root of the problem is in the inference rules: \([\text{Ax}]\) allows us to introduce any hypothesis (or trace) into a proof, and \([\neg \circ \text{I}]\) allows us to discharge that hypothesis (or bind the trace) at any point in the proof. When we bind a trace, we have no information about whether that trace is local, originating in the root clause, or whether it is more deeply embedded. Further, we are not able to keep track of the type of constituent in which the trace originates. When we discharge a hypothesis, we cannot tell whether that hypothesis was used to construct an embedded declarative clause or an embedded question, for example.
Clitic placement, subject control and *wh* extraction, as well as the placement of quantificational noun phrases, all involve introducing a certain kind of trace and then binding it at an appropriate point. Constraints on what kind of trace may be introduced to construct a proof of some sign, and at what point that trace can be bound we refer to as ‘locality constraints’.

In this chapter, we first describe the relevant locality constraints in Serbo-Croatian and point out, where appropriate, how our grammar overgenerates. Then, we propose a simple tectogrammatical solution to these overgeneration problems. Finally, we summarize and evaluate the resulting, more fine-grained, grammar.

### 7.2 Constraints on Trace Binding

#### 7.2.1 Quantification

Our current grammar predicts unrestricted scoping of all quantificational noun phrases. There are no locality constraints with respect to the scope of existentially quantifying expressions, within the fragment of Serbo-Croatian considered in this thesis. They can bind into its complement from an arbitrary distance. We think that the following string is semantically ambiguous between the wide and the narrow reading of the embedded quantificational noun phrase.

(360) Ana_

misli da Marko voli nekoga.  
Ana_{NOM,f,3sg} think_{3sg} DA Marko_{NOM,m,3sg} love_{3sg} someone_{ACC,m,3sg,3}  
‘Ana thinks that there’s a person such that Marko loves that person’  
‘There’s a person such that Ana thinks Marko loves that person’
Further, *nekoga* ‘someone’ can also take wide or narrow scope with respect to a question in which it originates. The following strings are also semantically ambiguous:

(361) a. Ana se pita da li Marko voli Ana\textsubscript{NOM,f,sg,3} REFL ask\textsubscript{sg,3} DA LI Marko\textsubscript{NOM,m,sg,3} love\textsubscript{3,sg} nekoga. someone\textsubscript{ACC,m,sg,3}
‘Ana wonders whether there’s a person x such that Marko loves x’
‘There’s a person x such that Ana wonders whether Marko loves x’
b. Ana se pita ko voli nekoga. Ana\textsubscript{NOM,f,sg,3} REFL ask\textsubscript{sg,3} who\textsubscript{NOM,m,sg,3} love\textsubscript{3,sg} someone\textsubscript{ACC,m,sg,3}
‘Ana wonders which person x is such that there’s a person y who x loves’
‘There’s a person y and Ana wonders who loves y’

In declarative environments, universally quantifying noun phrases can also take wide or narrow scope:

(362) Ana misli da Marko voli svakoga. Ana\textsubscript{NOM,f,sg,3} think\textsubscript{3sg} DA Marko\textsubscript{NOM,m,sg,3} love\textsubscript{3,sg} everyone\textsubscript{ACC,m,sg,3}
‘Ana thinks that every person is such that Marko loves that person’
‘Every person is such that Ana thinks that Marko loves that person’

However, universally quantifying expressions cannot escape from questions:

(363) a. Ana se pita da li Marko voli Ana\textsubscript{NOM,f,sg,3} REFL ask\textsubscript{sg,3} DA LI Marko\textsubscript{NOM,m,sg,3} love\textsubscript{3,sg} svakoga. everyone\textsubscript{ACC,m,sg,3}
‘Ana wonders for every person x whether Marko loves x’
# ‘For every person x Ana wonders whether Marko loves x’
b. Ana se pita ko voli svakoga. Ana\textsubscript{NOM,f,sg,3} REFL ask\textsubscript{sg,3} who\textsubscript{NOM,m,sg,3} love\textsubscript{3,sg} everyone\textsubscript{ACC,m,sg,3}
Ana wonders which person x is such that for every person y x loves y.’
# ‘For every person y Ana wonders who loves y’

To get wide universal readings not available for strings in (363a) and (363b), a resumptive strategy is used, illustrated in examples below:

(364) a. Za svaku ženu, Ana se pita da li joj se Marko udvara.
   For every woman, Ana wonders whether Marko is flirting with her.

b. Za svaku ženu, Ana se pita kojoj se udvara.
   For every woman, Ana wonders who is flirting with her.

In the examples above joj ‘to her’ is the resumptive pronoun, whose case is determined by the verb (in this case udvarati se ‘to court, to flirt’). However, the preposition za ‘for’ determines the case of its complement noun phrase, so the wide scoping svaku ženu ‘every woman’ occurs in accusative case, and not in the case required by the verb. Thus, the case of the resumptive pronoun and the case of the preposed existentially quantifying noun phrase do not necessarily match, although their gender and number do. We simply note the existence of this construction which allows the quantificational noun phrase to have wide scope, but we do not pursue its analysis here.
Finally, we consider the relative scope of existentially and universally quantifying noun phrases. As discussed in Chapter 3, in declarative sentences, both wide universal and wide existential readings are possible.

(365) a. Neko voli svakoga.
   someone_{NOM,m,sg} love_{3,sg} everyone_{ACC,m,sg,3}
   ‘There’s a person x such that x loves everyone’
   ‘For every person y there’s a person x such that x loves y’

b. Svako voli nekoga.
   everyone_{NOM,m,sg,3} love_{3,sg} someone_{ACC,m,sg,3}
   ‘There’s a person x such that everyone loves x’
   ‘For every person y there’s a person x such that y loves x’

Strings with embedded declaratives, and with multiple quantificational expressions, are multiply semantically ambiguous. The string below can be interpreted many different ways, for example, the existentially quantifying noun phrase can have wide matrix scope (a), narrow matrix scope (b), wide embedded scope (c) or narrow embedded scope (d). The embedded universally quantifying noun phrase can also have matrix scope (e).

(366) Svaki profesor misli da je neki
   every_{NOM,m,sg} professor_{NOM,m,sg} think_{3,sg} DA be_{3,sg} some_{NOM,m,sg}
   student_{NOM,m,sg} read_{ppl,m,sg} every_{ACC,f,sg} book_{ACC,f,sg}
   student_{NOM,m,sg} svaku knjigu.
   svaku knjigu.
   a. ‘There’s a student such that every professor thinks that student read every book’ (i.e. every professor thinks that Sandy read every book)
   b. ‘For every professor there’s a student such that the professor thinks that student read every book’ (i.e. Mark thinks that Sandy read every book, Robin thinks that Tom read every book, etc.)
   c. ‘Every professor thinks that there’s a student who read every book’
   d. ‘Every professor thinks that every book was read by some student or other’
e. ‘For every book, every professor thinks that there’s some student who read it’

The interpretation of sentences with embedded interrogatives, and with multiple quantificational expressions, is more restricted since universally quantifying noun phrases cannot escape from questions. It is impossible to interpret the string below with the embedded universally quantifying noun phrase taking matrix scope:

(367) Marko se pita da li je neki student pročitao svaku knjigu.
Marko NOM,m,sg 3 REFLEX ask 3,sg DA LI be 3,sg every NOM,m,sg student NOM,m,sg readppl,m,sg some ACC,f,sg book ACC,f,sg
a. ‘Marko wonders whether there’s a student who read every book’
b. ‘Marko wonders whether every book was read by some student or other’
c. ‘There’s a student such that Marko wonders whether that student read every book’
d. ‘For every book, Marko wonders whether there’s a student who read it’

In sum, existentially and universally quantifying noun phrases are subject to different locality constraints: while the existentially quantifying ones can always scope arbitrarily high, the universally quantifying ones cannot escape from interrogatives. Similarly, it seems that universally quantifying noun phrases have to scope locally if they occur in relative clauses, but this doesn’t seem to be the case for existentially quantifying ones. For example:

(368) Neki student koji je pročitao svaku knjigu je bio umoran.
Neki NOM,m,sg student NOM,m,sg which NOM,m,sg,3 be 3,sg readppl,m,sg some ACC,f,sg book ACC,f,sg be 3,sg book ACC,f,sg bepppl,m,sg tired NOM,m,sg

247
a. ‘There’s a student who read every book and was tired’
b. ‘For every book there’s a student who read it and was tired’

(369) Svaki student koji je pročitao
    everyNOM,m,sg studentNOM,m,sg whichNOM,m,sg,3 be3,sg readppl,m,sg
    neku knjigu je bio umoran.
    nekuACC,f,sg bookACC,f,sg be3sg,3 beppl,m,sg tiredNOM,m,sg

a. ‘Every student who read some book or other was tired’
b. ‘There’s a book such that every student who read it was tired’

Since we do not analyze relative clauses at all, we simply note these facts but do not present an account locality constraints in relative clauses.

This section by no means constitutes an exhaustive characterization of quantification in Serbo-Croatian. A better understanding of this empirical domain would require an investigation of (i) the universally quantifying pronouns svi ‘everyone’ and sve ‘everything’ and the accompanying determiners meaning ‘all’, in addition to svako ‘everyone, each person’ and the accompanying determiners meaning ‘every, each’; (ii) existentially quantifying bare nouns, as opposed to the pronoun neko ‘someone’ or noun phrases containing the determiner neki ‘some’; and, (iii) the interaction of quantifier scope with negation, which we avoid here since we do not have an analysis of negative concord or negative polarity items in Serbo-Croatian.

Here, based on the data considered above, we entertain the following preliminary generalization: existentially quantifying expressions have unrestricted scope in Serbo-Croatian, while universally quantifying ones have to scope locally if they
originate in questions. Our grammar, however, at present predicts unrestricted scope of both kinds of quantificational expressions.

### 7.2.2 Finite Control

Recall that we analyzed finite controlled complements as embedded verb phrases, i.e. embedded clauses with bound subject traces. This is because in finite controlled complements the complementizer *da* obligatorily occurs. However, control is local, in that the subject or object controller may only control the immediately embedded verb phrase, and not a more deeply embedded one.

With respect to finite control, our grammar currently generates non-sentences in (a), where long distance control is attempted, in addition to sentences in (b). Below, $x_i$ indicates the location of the gap, and $i$ indexes the noun phrase the gap is supposed to be coreferential with:

(370) subject control:

a. * Ana$_i$ mora da oni misle da $x_i$ dode.
   Ana$_{NOM,f,sg,3}$ mora$_{sg,3}$ DA they$_{NOM,m,pl,3}$ think$_{pl,3}$ DA arrive$_{sg,3}$
   [where Ana is supposed to control the embedded verb phrase *dode*]

b. Ana$_i$ mora da $x_i$ dode.
   Ana$_{NOM,f,sg,3}$ mora$_{sg,3}$ DA arrive$_{sg,3}$
   ‘Ana must come’

(371) object control:

a. * Ana nagovara Marka$_i$ da oni
   Ana$_{NOM,f,sg,3}$ persuade$_{3,sg}$ Marka$_{ACC,m,sg,3}$ DA they$_{NOM,m,pl,3}$
   misle da $x_i$ dode.
   think$_{pl,3}$ DA arrive$_{sg,3}$
   [where Marka is supposed to control the embedded verb phrase *dode*]
b. Ana nagovara Marko da x₁ dode.
   \[ \text{Ana}_{\text{NOM},f,sg,3} \text{ persuade}_{\text{3sg}} \text{ Marko}_{\text{ACC},m,sg,3} \text{ DA} \text{ arrive}_{\text{sg,3}} \]
   ‘Ana persuades Marko to come’

### 7.2.3 Enclitic Placement in Finite Clauses

Our grammar also allows clitic climbing out of finite clauses which is not possible in Serbo-Croatian. Below, in non-sentences in (a) the embedded clitics have escaped into the matrix clause. In reality, only examples as in (b) are acceptable. Our grammar, however, generates both. Clitics are in boldface.

(372) pronominal clitics:
   a. * Ana \text{ ga} misli da Sanja vidi.
      \[ \text{Ana}_{\text{NOM},f,sg,3} \text{ he}_{\text{ACC},m,sg,3} \text{ think}_{\text{3sg}} \text{ DA} \text{ Sanja}_{\text{NOM},f,sg,3} \text{ see}_{\text{sg,3}} \]
      ‘Ana thinks Sanja sees him’
   b. Ana misli da \text{ ga} Sanja vidi.

(373) reflexive clitics:
   a. * Ana \text{ se} misli da Marko vidi.
      \[ \text{Ana}_{\text{NOM},f,sg,3} \text{ REFL think}_{\text{3sg}} \text{ DA} \text{ see}_{\text{sg,3}} \]
      ‘Ana thinks Marko sees himself’
   b. Ana misli da \text{ se} Marko vidi.

(374) verbal clitics:
   a. * Ana \text{ je} misli da Marko kupio
      \[ \text{Ana}_{\text{NOM},f,sg,3} \text{ REFL think}_{\text{3sg}} \text{ DA} \text{ Marko}_{\text{NOM},m,sg,3} \text{ buy}_{\text{pl,m,sg}} \]
      \text{book}_{\text{ACC},f,sg,3}
      ‘Ana thinks Marko bought the book’
   b. Ana misli da \text{ je} Marko kupio knjigu.
7.2.4 Enclitic Placement in Controlled Complements

Clitics almost always occur in the finite controlled complement and don’t climb out to the matrix clause, constructed by the control verb. However, sometimes clitics have the option of climbing out of finite controlled verb phrases but only to the next finite clause. Progovac (2005) distinguishes between I-verbs (indicative-like), and S-verbs (subjunctive-like), and points out that if the subject control verb is an S-verb, clitics can climb out of the controlled verb phrase and occur in matrix clause built by that S-verb; the (a) example below is shown with her judgment, while the (b) example is uncontroversially acceptable.

(375) clitics of controlled verb phrases embedded by S-verbs:
   a. ? Ana ga želi da vidi.
      Ana_{NOM,f,sg,3} he_{ACC,m,3,sg} want_{3sg} DA see_{3sg}
      ‘Ana wants to see him’
   b. Ana želi da ga vidi.

However, if an S-verb embeds an entire finite clause, and not a finite controlled verb phrase, clitic climbing is not possible, as shown below. In sum, clitics can only climb out of finite controlled verb phrases which are arguments of S-verbs.

(376) clitics of clauses embedded by S-verbs:
   a. * Ana ga želi da Marko vidi.
      Ana_{NOM,f,sg,3} he_{ACC,m,3,sg} want_{3sg} DA Marko_{NOM,m,sg,3} see_{3sg}
      ‘Ana wants Marko to see him’
   b. Ana želi da ga Marko vidi.

Further, clitics which originate in an infinitival verb phrase have the same options, regardless of whether the control verb is an I- or an S-verb. Clitics can and
usually do climb out of controlled infinitival verb phrases, but only climb out to
the next finite clause, not arbitrarily high. At the same time, clitics can also encliti-
cize onto the infinitive itself. For example:

(377) clitics of controlled infinitival verb phrases
   a. * Ana ga misli da Marko želi
tAnaNOM,f,sg,3 itACC,m,3,sg think3sg DA MarkoNOM,m,sg,3 want3sg
dati Maji.
giveinf MajaDAT,f,sg,3
   ‘Ana thinks that Marko wants to give it to Maja’
   b. Ana misli da ga Marko želi dati Maji.
   c. ? Ana misli da Marko želi dati ga Maji.

Our judgments are indicated in the examples above. Whatever one’s judgments
concerning the placement of the enclitics onto the infinitive, the (b) example is
uncontroversially good and the (a) example uncontroversially bad. We will show
how the grammar can be optionally modified to allow sentences like (c), while
excluding examples like (a) and still generating (b).

Finally, predicative complements behave just like controlled complements of
I-verbs: the clitics must climb out to the matrix finite clause (a, c); they cannot
occur in the predicative complement in which they originate (b); and, they cannot
climb out higher than the matrix finite clause (d).

(378) clitics of predicative complements
   a. Ana mu ga je poslala.
      AnaNOM,f,sg,3 heDAT,m,sg,3 itACC,m,sg,3 be3sg sendpl,f,sg
      ‘Ana sent it to him’
   b. * Ana je poslala mu ga.
      AnaNOM,f,sg,3 be3sg sendpl,f,sg heDAT,m,sg,3 itACC,m,sg,3
In case of nested complements which all allow clitic climbing, the clitics climb out until they get to a finite clause, for example:

(379)  Ana  mu  ga  je  htjela  poslati.  
       Ana_{NOM,f,sg,3}  he_{DAT,m,sg,3}  it_{ACC,m,sg,3}  think_{sg,3}  DA  be_{3,sg}  Ana_{NOM,f,sg,3}  send_{pl,f,sg}  
       ‘Ana wanted to send it to him’

Here, *mu ‘to him’ and ga ‘it’ climb out of the infinitival complement, and over the participle, to get to the main finite clause level.

7.2.5 Wh Extraction

Just like quantificational noun phrases, *wh traces can also bind from an arbitrary distance, so long as they originate in an embedded declarative clause:

(380)  a.  Koga  Ana  misli  da  Marko  voli?  
       who_{ACC,m,sg,3} Ana_{NOM,f,sg,3}  think_{3,sg}  DA  Marko_{NOM,m,sg,3}  love_{3,sg}  
       ‘Who does Ana think Marko loves?’

  b.  Koga  Ana  misli  da  ko  voli?  
       who_{ACC,m,sg,3} Ana_{NOM,f,sg,3}  think_{3,sg}  DA  who_{NOM,m,sg,3}  love_{3,sg}  
       ‘For which x and y does Ana think that x loves y?’
However, no *wh* extraction is allowed out of questions, polar or constituent:

(381) * Koga who se Ana who pita da li Marko who voli?
     who_{ACC,m,sg,3} who_{ACC,m,sg,3} REFLEX who_{NOM,f,sg,3} who_{NOM,m,sg,3} who_{NOM,m,sg,3} who_{LOVE,3sg}
     ‘Who does Ana wonder whether Marko loves?’

(382) * Koga who se Ana who pita ko who voli?
     who_{ACC,m,sg,3} who_{ACC,m,sg,3} REFLEX who_{NOM,f,sg,3} who_{NOM,m,sg,3} who_{NOM,m,sg,3} who_{LOVE,3sg}
     ‘Who does Ana wonder who loves?’

As for *wh* in situ traces, if they originate in a *wh* question which is embedded inside another *wh* question, they can be bound at the embedded or the root level; this gives rise to Baker’s ambiguity.

(383) Ko who y who se pita ko who z who koga who x who voli?
     who_{NOM,m,sg,3} who_{NOM,m,sg,3} who_{NOM,3sg} who_{NOM,m,sg,3} who_{NOM,m,sg,3} who_{NOM,m,sg,3} who_{LOVE,3sg}
     ‘For which x does y wonder which z loves x?’
     ‘Which y wonders which z loves which x?’

However, in situ *wh* traces cannot originate in a polar question at all; they cannot be bound at the level of the polar question (a), or at the level of the matrix *wh* question (b). Examples below are only acceptable on an echo reading.34

(384) a. * Da li ko who spava?
     DA who_{NOM,m,sg,3} who_{NOM,m,sg,3} who_{SLEEP,3sg}
     ‘Does who sleep?’

b. * Ko who x who se pita da li ko who y who spava?
     who_{NOM,m,sg,3} who_{NOM,m,sg,3} who_{NOM,m,sg,3} who_{NOM,3sg} who_{NOM,m,sg,3} who_{NOM,m,sg,3} who_{SLEEP,3sg}
     ‘For which y does x wonder whether y sleeps?’

34In addition to the echo question interpretation, it is also possible to interpret the *ko* ‘who’ which occurs in the polar question as an indefinite, so that (384a) means something like ‘Is someone/anyone sleeping?’ and (384b) ‘Who wonders whether someone/anyone is sleeping? We ignore such interpretations here.
7.2.6 Summary

In sum, there are several different types of traces which behave differently with respect to where they can be bound. Some of these traces, such as existential quantificational traces, can be bound arbitrarily high; others, such as clitic traces, have to be bound locally. Below is a list of the different kinds of traces:

- existential quantificational traces
- universal quantificational traces
- clitic traces
- subject traces in controlled complements
- wh filler traces
- wh in situ traces

We have to distinguish between subject traces in controlled complements and other subject traces, that might be bound by a quantificational or a *wh* expression. This is because the former must be bound locally before the control verb combines with the controlled complement, since long distance control is illicit. Nominative quantificational or *wh* traces can, however, be bound arbitrarily high.

Where different traces can and cannot be bound also depends on the constituent in which the trace originates and the kind of expression which embeds that constituent.

For example, consider an embedded declarative clause. If such a clause has clitic traces, we need to prevent them from being passed onto the matrix level so
that the clitics can’t climb out of embedded declaratives. We also don’t want an embedded declarative to pass on an unbound subject trace in a way that would allow long distance control, which is illicit. However, unbound quantificational and \textit{wh} traces in the embedded clause should be passed on to the matrix clause, because \textit{wh} and quantificational traces can be bound from afar.

As another example, consider a constituent question. A constituent question cannot pass on any clitic traces, since this would allow clitic climbing; it cannot pass on subject traces in a way that would allow long distance control; it cannot pass on any filler \textit{wh} traces because \textit{wh} extraction from questions is not allowed; and, it cannot pass on any universal quantificational traces. However, a constituent question has to be able to pass on existential quantificational traces, as well as \textit{in situ} \textit{wh} traces; the latter is necessary to obtain Baker ambiguous readings when a multiple \textit{wh} question is embedded inside another \textit{wh} question. Polar questions behave much like \textit{wh} questions except that they must also prevent \textit{in situ} \textit{wh} traces from escaping.

Next, consider finite controlled complements. They need to be able to pass on all \textit{wh} and quantificational traces to the next level. Further, they have to contain a local subject trace which can then be bound so that the matrix subject can be interpreted as coreferential with the subject trace in the controlled verb phrase. If that control verb is an I-verb, the clitic traces cannot be passed on because clitic climbing out of controlled verb phrases is not allowed in this case. However, if
the control verb is an S-verb, the clitic traces should have the option of climbing out of the controlled verb phrase and into the matrix clause (but no higher than that).

So, different clause-like constituents (declaratives, interrogatives and controlled complements) have to act as barriers for certain kinds of traces, preventing them from being passed on, and act as holes for other kinds of traces, allowing them to escape to the matrix level. Table 7.1 shows for each clause-like constituent, and each type of trace, whether such a trace can escape from the constituent in question or not. It clearly shows the gradation in terms of (i) how good different kinds of traces are at escaping, and (ii) how good different kinds of constituents are at preventing traces from escaping.

Clitic traces are more or less restricted to their local finite clauses; traces of subjects of controlled verb phrases can escape out of the controlled verb phrase in which they originate but no higher; universal quantificational traces and wh filler expressions cannot escape out of questions; wh in situ traces cannot escape from polar questions only; and, finally, existential quantificational traces can escape from arbitrary constituents.

Looking at different constituent types, whereas predicative and finite controlled complements of I-verbs pass on all traces to the matrix level, polar questions are barriers to almost all traces, only allowing existential quantificational traces to escape.
<table>
<thead>
<tr>
<th>CONSTITUENT TYPE</th>
<th>clitic</th>
<th>subject of controlled VP</th>
<th>wh filler; ∃</th>
<th>wh in situ</th>
<th>∃</th>
</tr>
</thead>
<tbody>
<tr>
<td>polar question</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td><em>wh</em> question</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>declarative clause</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>finite controlled complement of I-verbs, infinitival complement</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>finite controlled complement of S-verbs, predicative complement, infinitival complement</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 7.1: Which traces can (√) or cannot (✗) escape from which constituents.
In the next section, we sketch the tectogrammatical mechanism which we use to implement these locality constraints in our grammar, which is what the remainder of the chapter is dedicated to.

### 7.3 Implementing Locality Constraints

We propose a purely tectogrammatical mechanism for implementing locality constraints. For that reason, in this chapter we suppress the presentation of phenogrammatical and semantic typed terms, and only show how to modify the tectogrammatical representations of expressions.

Our implementation involves adding an additional natural number parameter to sentence types and predicatives, as well as noun phrases. We superscript this natural number parameter to type symbols to improve readability. For example, the following are now tectogrammatical types in our grammar:

\[(385)\]

\[\begin{align*}
\text{a. } S_{m,6}^1 & \quad \text{where } \vdash m : K, \vdash 6 : \text{Nat}, \text{ and } \vdash 1 : \text{Nat} \\
\text{b. } \text{Prd}_{pl}^3 & \quad \text{where } \vdash pl : D, \text{ and } \vdash 3 : \text{Nat} \\
\text{c. } \text{NP}_{\text{nom},f,sg,3}^5 & \quad \text{where } \vdash \text{nom} : \text{Cse}, \vdash f : \text{Gdr}, \vdash sg : \text{Num}, \vdash 3 : \text{Prs} \text{ and } \vdash 5 : \text{Nat}
\end{align*}\]

In the remainder of this chapter, we omit the presentation of the \textit{Gdr}, \textit{Prs} and \textit{Num} parameters on noun phrase types; so we will write \(\text{NP}_{\text{nom}}^5\) instead of the full type given in (c) above.
Recall that our task is essentially to distinguish different kinds of traces, and then for each kind of trace, restrict its options with respect to whether it is passed to the matrix level or not by a given constituent, and when it can be bound.

We use the newly introduced number parameter to decorate each kind of noun phrase trace. The correspondence between different kinds of noun phrase traces and the tectogrammatical types of such traces, for a given value of $c$, is shown in Table 7.2.

Of course, the [Ax] rule cannot itself be constrained to only introduce noun phrase hypotheses whose number parameter is 1, 2, 3 or 5. [Ax] allows us to introduce noun phrase traces with an arbitrary number parameter. However, the expressions whose complements have bound traces (clitics, quantificational noun phrases, $wh$ expressions and control verbs) will require that the bound trace in question have a specific number parameter, as in Table 7.2. Lexical noun phrases all come decorated with number 1, just like quantificational noun phrase traces.

A transitive finite verb constructs a sentence whose clitic counting number parameter is 6, and whose trace number parameter is the product of the trace

<table>
<thead>
<tr>
<th>TRACE TECTOGRAMMATICAL TYPE</th>
<th>clitic</th>
<th>subject of controlled VP $wh$ filler; $\forall$</th>
<th>$wh$ in situ $\exists$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TECTOGRAMMATICAL NP^2_c</td>
<td>NP^2_c</td>
<td>NP^3_c</td>
<td>NP^5_c</td>
</tr>
<tr>
<td>NP^1_c</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.2: Tectogrammatical types of traces.
number parameters of its noun phrase arguments. For example, if a transitive verb combines with a lexical noun phrase as its subject, and a clitic trace as its object, the trace number parameter of the resulting clause will be 2. If it combines with a filler _wh_ trace and a clitic trace, the clause trace parameter will be 6.

Since 1 is the identity for multiplication, lexical noun phrases add nothing to the clausal trace parameter. Similarly, since existentially quantifying noun phrase traces can escape from any constituent and be bound arbitrarily high, clauses do not have to keep track of how many such traces they have, and so we parametrize quantificational noun phrase traces with 1.

For other kinds of traces, the trace number parameter on clausal types keeps track of _how many of each kind of trace_ there are in the clause. This is a direct consequence of the following: (i) the fact that the clitic, subject, universal quantificational and _wh_ traces are parametrized by primes, (ii) the fact that the clausal trace number parameter is a product of those primes, and, finally, (iii) the unique factorization theorem, which states that every integer greater than 1 is uniquely decomposable into a product over a multiset of primes.\(^{35}\)

\(^{35}\)Many thanks to Bart Snapp for his suggestion to use products instead of sums, and the discussion of the properties of multiplication and exponentiation, as opposed to addition. The key difference for us here is that sums don’t retain much information about how they are constructed, i.e. what the summands are and how many of each kind there are in the sum. But because of the prime factorization theorem, from a product we can recover its internal structure in terms of its prime factors, and the number of occurrences of each prime factor in a product.
For us, this means that if a sentence contains a clitic trace, for $n>0$, its trace number parameter will be of the form $2n$; if it contains a $wh$ filler trace it will be of the form $3n$; and, if it contains a $wh$ in situ trace it will be of the form $5n$.

Further, through exponentiation, itself defined in terms of multiplication, we can keep track of how many traces of a given type a clause contains. It’s important to recall that any number raised to 0 is 1. So, if a sentence’s trace number parameter can be represented as $2^i3^n5^m$ then it contains exactly $n$ $wh$ filler traces, $m$ $wh$ in situ traces, and $i$ traces of clitics or of subjects in controlled verb phrases. This is all a consequence of the unique factorization theorem.

As a typographical convention, we will denote multiplication by juxtaposition, and exponentiation by superscripting the exponent to the number which is being raised to that power.

So, clauses keep track of how many of each kind of trace they contain, and embedding expressions can then require that their complement contain or not contain such-and-such a trace.

For example, if a an embedding expression requires that its complement clause has the trace parameter 1, it means that all traces of its complement, except the existential quantificational ones, have to be bound and taken care of before the complement can be embedded. As another example, if an embedding expression requires that its complement’s trace number parameter be representable as $5^n$, it means that it will only pass on $wh$ in situ and existential quantificational traces.
If an embedding expression requires that its complement trace product be repre-
sentable as $5^m3^m$ then it will pass on both filler and in situ wh traces, as well as
quantificational traces, but not clitic or controlled subject traces.

Enclitics, universally quantifying noun phrases, wh expressions and control
verbs require that its complement clause have a bound trace with such-and-such
number parameter, and after combining with it, they remove a factor correspond-
ing to the trace from the trace product of the resulting clause. For example, if a
pronominal clitic combines with an expression whose type is $\text{NP}^2_{\text{acc}} \rightarrow \text{S}^4_{m,6}$, it
removes the clitic trace (factor of 2) from the sentence’s trace product and outputs
an expression of type $\text{S}^2_{m,6}$.

In the next section, we show how to use this tectogrammatical mechanism to
enforce locality constraints in our Serbo-Croatian grammar.

7.4 Enforcing Locality Constraints

7.4.1 Declarative Clauses

Recall that traces of clitics and subjects in controlled verb phrases cannot be
passed up to the matrix clause from an embedded declarative clause. In other
words, clitics cannot climb out of finite embedded clauses, and long distance con-
trol is not possible. However, both kinds of wh traces as well as quantificational
traces have to be able to escape from the embedded declarative.

We cannot build these locality constraints into the complementizer, because in
embedded declaratives the complementizer obligatorily hosts the clitics, and so
the clitic traces have to be bound and the clitics appropriately placed only after
the complementizer has picked up its complement clause. For this reason, we
required the complementizer to combine with finite clauses inside of which no
clitics have been placed (i.e. with clitic parameter 6). So, we analyze the comple-
mentizer itself as a hole for all traces, and give it the following tectogrammatical
representation:

\[(386) \vdash da : \prod_{i \in \mathbb{N}} [S_{m,6}^i \circ S_{e,6}^i] \]

As before, the complementizer only changes the K parameter to e, and simply
passes along to the matrix level any traces that its complement clause might have.
It’s the embedding verbs which we analyze as forcing their complement to have
no clitic traces, or traces of subjects which might be controlled. Below is the tec-
togrammatical representation of the verb misli ‘thinks’:

\[(387) \vdash misli : \prod_{n, i, j, k \in \mathbb{N}} [S_{e,n}^{3^i5^k} \circ NP_{nom}^j \circ S_{m,6}^{3^i5^k}] \]

If a natural number can be represented as \(3^i5^k\), then 2 is not one of its fac-
tors. So, if a sentence’s trace parameter can be represented as \(3^i5^k\), then it doesn’t
contain any noun phrase traces whose number parameter is 2. In this way, the em-
bedding verb forces its complement clause to not contain any clitic or controlled
subject traces. It does, however, allow \(wh\) traces to be passed on to the matrix
level, as well as any quantificational traces. The number parameter of the matrix
subject \((j)\) is then just added to the trace product as another factor, resulting in
\(3^i5^kj\).
Quantificational Traces

Suppose we want to construct the following sentence, with an existentially quantifying pronoun:

(388) a. Ana misli da Marko voli nekoga.
    Ana_{NOM,f,sg,3} think_{sg,3} DA Marko_{NOM,m,sg,3} love_{sg,3} someone_{ACC}
    ‘Ana thinks that there’s someone who Marko loves’
    ‘There’s someone who Ana thinks Marko loves’

Below we give the tectogrammatical representations needed:

(389) a. ⊢ ana : NP_{nom}^{1}
    b. ⊢ marko : NP_{nom}^{1}
    c. ⊢ voli : \prod_{j,k:Nat}[\prod_{NP_{acc}^{j} \rightarrow NP_{nom}^{k}} S_{m,6}^{i}]
    d. ⊢ nekoga : \prod_{i:Nat}[\prod_{NP_{acc}^{i}} S_{m,6}^{i}]

This verb voli ‘loves’ keeps track of the number parameters of its noun phrase arguments (j and k), and constructs a main clause whose trace parameter is the product of j and k. The lexical noun phrases come with parameter 1. Given the tectogrammatical type of the quantificational noun phrase nekoga ‘someone’, it can only combine with clauses with a bound noun phrase trace whose parameter is 1. This ensures that quantificational noun phrase traces always have to be decorated with 1. This existentially quantifying noun phrase does nothing to alter its complement’s trace parameter since existential quantificational traces can always be passed up.

Below we show a proof of the sentences above. We only show the tectogrammatical types in the proof. We first introduce an accusative trace whose number
parameter is 1, since the quantificational expression *nekoga* ‘someone’ is looking for a sentence missing an accusative noun phrase whose number parameter is 1. We proceed to combine the verb *voli* ‘loves’, with appropriately instantiated number parameters, with the accusative trace and then the subject noun phrase.

(390)

\[
\begin{align*}
\text{Marko} & \vdash \text{NP}^{1}_{\text{acc}} \neg \text{NP}^{1}_{\text{nom}} \rightarrow \text{S}^{1}_{m,6} \\
\vdash \text{NP}^{1}_{\text{nom}} & \vdash \text{NP}^{1}_{\text{acc}} \rightarrow \text{NP}^{1}_{\text{nom}} \rightarrow \text{S}^{1}_{m,6} \rightarrow \text{E} \\
\vdash \text{NP}^{1}_{\text{acc}} & \vdash \text{S}^{1}_{m,6}
\end{align*}
\]

At this point, the proof can continue in two different ways. The first option is to bind the trace, introduce the quantificational noun phrase and then embed the resulting sentence. This gives the narrow reading.

(391)

\[
\begin{align*}
\text{complementizer } da & \vdash \text{NP}^{1}_{\text{acc}} \rightarrow \text{NP}^{1}_{\text{nom}} \rightarrow \text{NP}^{1}_{\text{acc}} \rightarrow \text{S}^{1}_{m,6} \rightarrow \text{E} \\
\vdash \text{NP}^{1}_{\text{nom}} & \vdash \text{NP}^{1}_{\text{acc}} \rightarrow \text{NP}^{1}_{\text{nom}} \rightarrow \text{S}^{1}_{m,6} \rightarrow \text{E}
\end{align*}
\]

Note that the *i* and *k* parameters of the embedding verb *misli* ‘thinks’ were both instantiated as 0, so that \(3^{i}5^{k}\) is 1 since the embedded clause doesn’t have any *wh* traces.
The second option is to not bind the quantificational trace at the embedded level and instead embed the sentence with the trace, and bind it at the main clause level. This gives the wide scope reading.

(392)

Recall that universal quantificational traces seem to behave just like filler wh traces with respect to locality constraints; therefore, they are parametrized by 3. We give the following tectogrammatical representation to the universally quantifying pronoun svako ‘everyone’:

(393)  ⊨ svako : \prod_{i:Nat}[(NP^3_{nom} \rightarrow S^3_{i,6}) \rightarrow S^i_{m,6}]

This pronoun combines with a sentence with a bound nominative trace whose number parameter is 3 (NP^3_{nom}), and removes a factor of 3 from the resulting sentence’s trace product. Since declarative embedding verbs such as misli ‘thinks’ allow their complement clause to have 3 as a factor, universal quantificational traces originating in an embedded clause can nonetheless have matrix scope.

267
Wh Traces

Next, we show how *wh* traces can escape from embedded declaratives and be bound at the matrix level. Below we show tectogrammatical representations of the filler *wh* expression ko ‘who_{NOM}’ and the *in situ* *wh* expression koga ‘who_{ACC}’.

(394) a. \( \vdash \text{ko} : \prod_{\pi : \text{Nat}} [\big( \text{NP}_3^{\text{nom}} \rightarrow S_{m,6}^{3n} \big) \rightarrow S_{q,6}^n] \)
   
b. \( \vdash \text{koga} : \prod_{\pi : \text{Nat}} [\big( \text{NP}_5^{\text{acc}} \rightarrow S_{q,6}^{5n} \big) \rightarrow S_{q,6}^n] \)

Consider the following sentence, in which both *wh* expressions escape from the embedded clause:

(395) Ko who\text{NOM}_{m,sg,3} Ana who\text{NOM}_{f,sg,3} misli think_{3,sg} DA who\text{ACC}_{m,sg,3} koga who_{ACC}_{m,sg,3} voli? love_{sg,3}

‘Who does Ana think loves who?’

Below we show how to construct a proof of this sentence. The verb *voli* ‘loves’ has to have its number parameters instantiated as 5 and 3, so that it can combine with a *wh in situ* and a filler *wh* trace.

(396)

\[
\frac{\text{wh filler trace}}{\text{NP}_3^{\text{nom}} \vdash \text{NP}_3^{\text{nom}}} \quad \text{[Ax]} \quad \frac{\text{voli ‘loves’}}{\text{NP}_5^{\text{acc}} \vdash \text{NP}_3^{\text{nom}} \rightarrow S_{m,6}^{15}} \quad \frac{\text{wh in situ trace}}{\text{NP}_5^{\text{acc}} \vdash \text{NP}_5^{\text{acc}}} \quad \text{[Ax]} \quad \frac{\text{NP}_3^{\text{nom}} \vdash \text{NP}_3^{\text{nom}}}{\text{NP}_5^{\text{acc}} \vdash S_{m,6}^{15}} \quad \text{[\text{-E}]} \quad \text{NP}_5^{\text{acc}, \text{NP}_3^{\text{nom}} \vdash S_{m,6}^{15}} \quad \text{[\text{-E}]} \]

At this point, we could bind the filler and then the *in situ* *wh* trace and get a constituent question, but a declarative embedding verb such as *misli* ‘thinks’ could not take it as a complement, for both tectogrammatical (it requires a declarative
clause argument) and semantic reasons (it requires an argument of type $p$, not $k$).

Instead, we keep the traces unbound, in the context, and embed the clause.

Now, we bind the $wh$ traces and construct a main constituent question. Recall that for semantic and tectogrammatical reasons we first have to bind the filler trace and place the filler $wh$ expression, and then the in situ, as in the proof below.

Clitic Traces

Finally, we refine the tectogrammatical types of clitics and show how their placement in the finite clause in which they originate is mandated. We start with pronominal clitics, and give $ga$ ‘him’ and $mu$ ‘to him’ the following tectogrammatical types:
The clitics require that their complement have a bound clitic trace with parameter 2, and hence, that the clause with a clitic trace have the trace parameter expressible as $2n$. Once they combine with their complement, they remove 2 as a factor from the clause’s trace product parameter.

Consider the following example which shows again that the clitics must occur in the embedded clause in which they originate:

(400) a. Marko misli da mu ga Ana šalje.
    'Marko thinks that Ana sends it to him'


Below is the tectogrammatical representation of the verb šalje ‘sends’.

(401)   \(\frac{\text{šalje : }\Pi_{i,j,k:\text{Nat}}[\text{NP}_{\text{dat}}^i \rightarrow \text{NP}_{\text{acc}}^j \rightarrow \text{NP}_{\text{nom}}^k \rightarrow S_{m,o}^{ijk}]}{\text{šalje}}\)

To construct the sentence above, we instantiate the verb’s number parameters as 2, 2 and 1, for the clitic dative and accusative object and the lexical noun phrase subject respectively. Below is the proof of the embedded clause, with two clitic traces:
There is no way for a clause whose type is $S_{e,6}^4$ to be embedded by *misli* ‘thinks’, or any other such verb, because it requires that its complement’s trace product be expressible as $3^i5^k$. But 4 does not have 3 or 5 as one of its factors; conversely, 4 is not a factor of $3^i5^k$ for any $i$ and $k$. So, we must bind the clitic traces and place the clitics into the embedded clause before the embedding verb combines with it, as below:

(403)

After all the clitics have been properly placed, and the embedded clause’s trace product is expressible as $3^i5^k$ (in this case, with $i$ and $k$ both instantiated as 0), the embedding verb can pick it up as its argument.
Recall that the reflexive and verbal clitics introduce traces which are not noun phrases. The extension of our mechanism to those clitics may not be obvious. Below we show how reflexive clitics are trapped in finite clauses, and, since all verbal clitics are either control or predicative verbs, we return to their analysis later in the chapter, when we discuss locality of control.

For the inherent reflexive *se*, recall that the verbs requiring it are given lexical entries which already contain a trace. We now add the requirement that such verbs are lexically provided with a clitic gap, i.e. that 2 already occurs in the trace product of the main clause they are constructing. Below we show the tectogrammatical representation of *boji ‘is afraid of’*, which is required to occur with the inherent reflexive, as well as the inherent reflexive.

\[(404) \begin{align*}
    a. \quad & x : \text{SE} \vdash \text{boji} : \prod_{i,j} \text{Nat} [\text{NP}^i_{\text{gen}} \circ \text{NP}^j_{\text{nom}} \circ S^{2ij}_{\text{m},6}] \\
    b. \quad & \vdash \text{se} : \prod_{i,n} \text{Nat, k : K} [(\text{SE} \circ S^{2i}_{k,n>1}) \circ S^i_{k,1}] 
\end{align*} \]

The verb *boji ‘is afraid of’* constructs a trace product of the sentence by multiplying its genitive \((i)\) and nominative \((j)\) arguments’ number parameters with 2, where 2 in the product intuitively means that there’s also a clitic which has to be dealt with locally, namely the inherent reflexive. Just as with pronominal clitics, the inherent reflexive must be placed into a finite clause before that clause can be embedded, because a clause with an inherent reflexive trace has 2 as one its trace product factors. So, the inherent reflexive trace has to be removed by the clitic *se* before the clause can be picked up by an embedding verb.
Recall that for the true reflexive we had to introduce a complicated reflexive trace, to ensure the correct semantic interpretation of the clause containing it. Here we simply refine the tectogrammatical type of the reflexive hypothesis and the reflexive itself to ensure that it’s placed locally. Below we redefine the type $\text{REFL}$, the type of the reflexive trace, and give the tectogrammatical representation of the reflexive clitic.

\[(405) \quad a. \quad \text{REFL} =: \prod_{i: \text{Nat}} [(\text{NP}^2_{\text{acc}} \rightarrow \text{NP}^i_{\text{nom}} \rightarrow \text{S}^2_{m,6}) \rightarrow \text{NP}^i_{\text{nom}} \rightarrow \text{S}^2_{m,6}] \]
\[b. \quad \vdash \text{se} : \prod_{k: \text{K}, n, i: \text{Nat}} [(\text{REFL} \rightarrow \text{S}^2_{k,n>1}) \rightarrow \text{S}^i_{k,1}] \]

The reflexive clitic trace forces the sentence in whose construction it is used to contain 2 as one of the factors of its trace product, where 2 corresponds to the reflexive clitic.

So, even though the traces of reflexive clitics are not noun phrases, the trick is to still ensure that that clauses containing them contain a factor of 2 in their trace product. This is accomplished by the verb’s lexical entry in the case of the inherent reflexive, and by the reflexive hypothesis in the case of the true reflexive. The reflexive hypothesis, in turn, has to be of the type presented above because the clitic reflexive requires that its complement have a bound trace of precisely that form. Therefore, the reflexive clitics cannot climb out of finite clauses for exactly the same reason as other clitics considered above.
Traces of Controlled VP Subjects

Long distance control is prevented in exactly the same way in which clitic climbing is prevented. Since traces of controlled verb phrase subjects are parametrized by 2, a sentence containing such a trace cannot be embedded by ordinary declarative embedding verbs such as misli ‘thinks’ because they do not allow their complement’s trace product to have 2 as a factor. We discuss this further in the next section, which is dedicated to control and predication.

7.4.2 Controlled and Predicative Complements

In this section, we show how to implement locality constraints in control and predication, ensuring that the local subject trace of the complement is bound, and not a more deeply embedded one. Since clitic verbs are all either control or predicative verbs, we analyze them in this section as well. We also show how to force clitics to stay in the finite controlled complement, and how to allow them to climb out to the next finite clause if the control verb is an S-verb the predicative or. Finally, we show how to mandate clitic climbing out of predicatives, and how to allow clitics of infinitival controlled complements to either climb out or occur on the infinitive itself.

Finite Control

We start with finite controlled complements. Consider the following sentence:

(406) Ana hoče da ga vidi.
Ana_{NOM,f,sg,3} want_{3,sg} DA he_{ACC,m,sg,3} see_{3,sg}
‘Ana wants to see him’

We tectogrammatically represent the subject control hoče ‘wants’ as follows:

\[(407) \quad \vdash \text{hoče} : \prod_{n,i,j,k}[((\text{NP}^2_{\text{nom}} \rightarrow S_{e,n}^{3/5/2}) \rightarrow \text{NP}^k_{\text{nom}} \rightarrow S_{m,6}^{3/5/2})]
\]

The subject trace in a controlled complement has to be of type NP\text{nom}^2, because hoče ‘wants’ requires that its complement have a bound trace of exactly that type. Any clause containing such a trace would have 2 as a factor in its trace product. Therefore, it couldn’t be embedded by an ordinary embedding verb such as misli ‘thinks’ because such verbs do not allow its complement clauses to have 2 as factor in their trace product. This guarantees that control is local.

Further, the control verb hoče ‘wants’ requires that its complement have exactly one occurrence of 2 in its trace product. If its complement’s trace product had more than one occurrence of 2, it would not be expressible as \(3^{5/2}\). So, all clitics originating in the finite controlled verb phrase are required to be placed in the finite controlled complement, before the control verb picks it up. Since there are no nominative clitics, NP\text{nom}^2 can only be the controlled complement’s subject trace, and not a clitic trace.

Quantificational and \(wh\) traces, if any, are passed on by the control verb. The matrix clause subject’s number parameter \((k)\) is added as a factor to the main clause trace product.

Below we show how to construct the proof of the sentence Ana hoče da ga vidi

‘Ana wants to see him’.
This clause cannot be embedded by a verb like *misli* ‘thinks’ because its trace products is a power of 2 and *misli* ‘thinks’ doesn’t allow its complement’s trace product to have 2 as a factor. So, long distance control is prevented.

Further, clitic climbing is not allowed either. Suppose that at this point we bind the subject trace, and leave the clitic trace unbound, attempting to pass it on to the matrix level. That gives us a sign whose tectogrammatical type is \( \text{NP}_{\text{nom}}^2 \rightarrow S_{4,6}^4 \). However, *hoče* ‘wants’ requires that its complement be of type \( \text{NP}_{\text{nom}}^2 \rightarrow S_{e,6}^{3i/2} \) for some values of \( i \) and \( j \). But there are no values of \( i \) and \( j \) that would make \( 3^i5^j/2 \) equal to 4; i.e. no matter how we instantiate \( i \) and \( j \), the product \( 3^i5^j/2 \) has only one occurrence of factor 2. So we are forced to place the clitic before the control verb picks up its complement. The only way to continue the proof above is as follows:

(409)

\[
\frac{\text{subject trace}}{\text{NP}_{\text{nom}}^2 \vdash \text{NP}_{\text{nom}}^2} \quad \frac{\text{clitic trace}}{\text{NP}_{\text{acc}}^2 \vdash \text{NP}_{\text{acc}}^2 \quad \text{Ax}} \quad \frac{\text{NP}_{\text{nom}}^2 \rightarrow \text{NP}_{\text{nom}}^2 \rightarrow S_{m,6}^4}{\text{NP}_{\text{acc}}^2 \rightarrow \text{NP}_{\text{acc}}^2 \quad \text{Ax}} \quad \frac{\text{NP}_{\text{nom}}^2 \rightarrow \text{NP}_{\text{nom}}^2 \rightarrow S_{m,6}^4 \rightarrow S_{m,6}^4 \rightarrow S_{m,6}^4}{\text{NP}_{\text{acc}}^2 \rightarrow \text{NP}_{\text{acc}}^2 \quad \text{Ax}} \\
\vdash S_{m,6}^4 \rightarrow S_{e,6}^4 \\
\vdash S_{m,6}^4 \rightarrow S_{e,6}^4 \\
\vdash \text{NP}_{\text{acc}}^2, \text{NP}_{\text{nom}}^2 \vdash S_{e,6}^4
\]

276
We enforce the locality of object control similarly. Below we give the tectogrammatical representation of nagovara ‘persuades’, as in Ana nagovara Maju da ga vidi ‘Ana persuades Maja to see him’, where the matrix object Maju is controlling da ga vidi ‘to see him’.

\[
\frac{\text{Ana} \quad \text{hoče ‘wants’}}{\vdash \text{NP}^1_{\text{nom}}} \quad \vdash \text{NP}^2_{\text{nom}} \rightarrow S^2_{e,4} \rightarrow \text{NP}^1_{\text{nom}} \rightarrow S^1_{e,4} \quad \vdash \text{S}^1_{e,4} [\rightarrow E]
\]

\[
\vdash \text{NP}^1_{\text{nom}} \rightarrow S^1_{e,4} [\rightarrow E]
\]

\[
\vdash \text{S}^1_{e,4}
\]

(410) \( \vdash \text{nagovara : } \prod_{n,m,i,j,k} \text{Nat} [\text{NP}^6_{\text{acc}} \rightarrow (\text{NP}^2_{\text{nom}} \rightarrow S^3_{e,n} \rightarrow \text{NP}^m_{\text{nom}} \rightarrow S^3_{m,0} \rightarrow S^3_{/km}]
\]

Next, we show how to analyze clitic subject control verbs, namely clitics of htjetì ‘will, want’. Recall that we have to introduce a trace corresponding to the control verb, and then bind that hypothesis at the finite matrix clause level. Below we redefine the type FC as the type of control verb traces, and then give the tectogrammatical representation of the clitic ´ce ‘will’.

\[
\text{a. FC} = \text{def } \prod_{n,i,j,m} \text{Nat} [\text{NP}^2_{\text{nom}} \rightarrow S^3_{e,n} \rightarrow \text{NP}^m_{\text{nom}} \rightarrow S^3_{m,0} \rightarrow S^3_{/2m}]
\]

\[
\text{b. } \vdash \text{´ce : } \prod_{k,K,n} \text{Nat} [\text{FC} \rightarrow S^2_{k,0} \rightarrow S^2_{n}]\]

The only difference between a non-clitic control verb and a control verb clitic trace is that the former removes the only occurrence of 2, corresponding to the subject trace, from its complement’s trace product, whereas the latter does not. This is because although it removes the subject trace (2), it adds a clitic trace (also 2), so that we know that a clitic is missing.
The control verb clitic trace prevents the clitics from climbing out of the finite complement and allows *wh* and quantificational traces to be passed on in exactly the same way as the control verb *hoće* ‘wants’ analyzed above.

Below we show how to construct the proof of the interrogative *Ana će da ga vidi* ‘Ana will see him’. This proof proceeds exactly like the proof of *Ana hoće da ga vidi* ‘Ana wants see him’ up until the point when the subject trace is bound. Then, we introduce the verbal clitic trace, and combine the verb phrase with the subject. Finally, we bind the clitic trace and place the verbal clitic.

(412)

I-Verbs vs. S-Verbs

Control verbs discussed above require that the clitics be placed inside the finite controlled complement, and doesn’t let them climb out to the matrix clause. Recall that according to Progovac (2005) S-verbs may allow clitics of their controlled complement to climb out, so that both sentences below are acceptable:

(413) a. Ana želi da ga vidi.
   Ana_{NOM,f,sg,3} want_{3,sg} DA he_{ACC,m,sg,3} see_{3,sg}
   ‘Ana wants/wishes to see him’
b. Ana ga želi da vidi.

We provide verbs like želi ‘wants’ with two lexical entries, one which is just like hoče ‘wants’ and doesn’t allow clitic climbing, and the other which does. These are given below:

\[(414) \quad \begin{align*}
\text{a. } & \vdash želi : \prod_{n,i,j,k}[(\text{NP}_{2\text{nom}} \rightarrow S_{e,n}^{3i/2}) \rightarrow \text{NP}_{\text{nom}}^k \rightarrow S_{m,6}^{3/5k}] \\
\text{b. } & \vdash želi_{cc} : \prod_{n,i}[(\text{NP}_{2\text{nom}} \rightarrow S_{e,n}^{2i}) \rightarrow \text{NP}_{\text{nom}}^k \rightarrow S_{m,6}^{ki}] 
\end{align*}\]

Whereas the version of želi ‘wants’ which does not allow clitic climbing requires that its complement’s result type have exactly one factor of 2 in its trace product (\(S_{e,n}^{3i/2}\)) corresponding to the subject trace, želi ‘wants’ which does allow clitic climbing allows possibly multiple factors of 2 (\(S_{e,n}^{2i}\)). This ensures that its will pass on clitic traces to be bound at the matrix level.

Below is the proof of Ana ga želi da vidi ‘Ana wants to see him’ in which the clitic occurs in the matrix clause. It proceeds exactly like the proof for Ana hoče da ga vidi ‘Ana wants to see him’ given above up until the the embedded clause with the clitic trace and the subject trace is constructed. The only difference is that with želi we are not forced to place all the clitics inside the controlled complement, because it can embed a complement whose trace product contains multiple factors of 2.
In the proof above, we bind the accusative clitic trace at the main clause level, and
then place the clitic in the main clause, getting the clitic climbing word order.

In reality, only pronominal clitics may climb out of controlled complements
embedded by S-verbs. Our analysis allows verbal clitics to climb out too. How-
ever, the reason verbal clitics don’t climb out of S-verb controlled complements is
because verbal clitics are not allowed to occur in them at all.

Recall that S stands for ‘subjunctive like’. Complements of S-verbs are severely
restricted in terms of the tense and aspect. No future or past tense, or conditional
mood is allowed in S-verb controlled complements. Serbo-Croatian verbal clitics
are used precisely for the formation of past tense (imperfective present clitics of
*biti* ‘be’), future tense (clitics of *htjeti* ‘want’), and conditional mood (aorist clitics
of *biti* ‘be’). The following examples, where the S-verb controlled complement is
in future tense (a), past tense (b) and conditional mood (c), are all impossible, for
semantic reasons:
Perfective present is the norm for controlled complements (a), though imperfective present is allowed to some extent as well (b):

(416)  
Ana NOM,f,sg,3 want3,sg DA will3,sg DA come3,sg  
[attempted: ‘Ana wants that she will come’]  
b. * Ana želi da je došla.  
Ana NOM,f,sg,3 want3,sg DA is3,sg come,pl,f,sg  
[attempted: ‘Ana wants that she has come’]  
c. * Ana želi da bi došla.  
Ana NOM,f,sg,3 want3,sg DA would3,sg come,pl,f,sg  
[attempted: ‘Ana wants that she would come’]  

As for copular constructions other than past tense, instead of the imperfective forms of biti ‘be’ whether full or clitic, it is much better to use the perfective forms of that verb in a controlled complement, and the perfective forms do not have clitic counterparts:

(417)  
a. Ana želi da pročita knjigu.  
Ana NOM,f,sg,3 want3,sg DA read3,sg,pf book ACC,f,sg,3  
‘Ana wants to read a book’  
b. Ana želi da čita knjigu.  
Ana NOM,f,sg,3 want3,sg DA read3,sg,impf book ACC,f,sg,3  
‘Ana wants to be reading a book’  

(418)  
a. ? Ana želi da je pametna.  
Ana NOM,f,sg,3 want3,sg DA is3,sg,impf smart NOM,f,sg  
‘Ana wants to be smart’  
b. Ana želi da bude pametna.  
Ana NOM,f,sg,3 want3,sg DA is3,sg,pf smart NOM,f,sg  
‘Ana wants to be smart’
In sum, for independent reasons, verbal clitics typically cannot occur in S-controlled complements, let alone climb out of them. Our grammar does not take into account these tense and aspect restrictions at all, and so it overgenerates. We believe that these restrictions are in the purview of a more sophisticated theory of Serbo-Croatian meanings and the syntax/semantics interface. The articulation of such a theory, is unfortunately, beyond the scope of this thesis so we leave this issue unresolved.

**Infinitival Control**

If a controlled complement is infinitival, we want to make it possible for its clitics to climb out to the finite clause (a), but also for them to occur on the infinitive itself (b):

(419)  

\[
\begin{align*}
\text{a. } & \text{Ana } \epsilon e \text{ mu ga poslati.} \\
& \text{Ana will send it to him' } \\
\text{b. } & \text{Ana } \epsilon e \text{ poslati mu ga.} 
\end{align*}
\]

If the clitics occur in the infinitival verb phrase, they have to encliticize onto the infinitive (a) and cannot encliticize onto some other non-clitic argument of the infinitival verb (b). The entire infinitival verb phrase must remain contiguous and its constituents cannot permute with the main clause constituents (c, d). Finally, the infinitive itself has to be leftmost in such an infinitival verb phrase (e), and the whole infinitival verb phrase must occur on the right edge of the matrix clause (f) and cannot host main clause clitics (g):
In contrast, if the infinitive’s clitics climb out to the matrix clause, the infinitive itself and any of its non-clitic arguments are just treated as ordinary main clause arguments and can freely reorder with respect to them, so that all of the following are possible:

First we analyze the latter case, when clitics obligatorily climb out of the infinitival verb phrase, and all the non-clitic main clause and infinitival verb’s complements can freely reorder. Infinitival verbs have lexical entries like the following for poslati:

\[ \lambda_{x,y}.\text{POSLATI}_z \circ x \circ y : z \to z \to z; \]
\[
\prod_{i,k: \text{Nat}} [\text{NP}_{\text{dat}} \to \text{NP}_{\text{acc}} \to \text{NP}_{\text{nom}} \to S^2_{\text{inf},i}]; \text{send} : e \to e \to e \to p
\]
So, this infinitival verb constructs an infinitival clause whose trace product is \(2ik\), where 2 corresponds to the missing subject, \(i\) to the parameter of its dative argument, and \(k\) to the parameter of its accusative argument. In reality, such verbs can never combine with the subject for phenogrammatical reasons: they combine with two \(S\)-strings corresponding to the two objects and construct another \(S\)-string, but there is not room for the subject noun phrase.

Tectogrammatically, a non-clitic control verb such as \(hoće\) ‘wants’ which can combine with an infinitival verb phrase, has the following representation:

\[
\begin{aligned}
\exists_{\text{hoće}_{\text{inf}}}: & \prod_{i,k: \text{Nat}} (\text{NP}^2_{\text{nom}} \rightarrow S^2_{\text{inf},6}) \rightarrow \text{NP}^k_{\text{nom}} \rightarrow S^k_{m,6} \\
\end{aligned}
\]

It doesn’t allow any clitics to occur on its infinitive because it requires that the infinitival verb phrase have clitic parameter 6. So, it forces any clitics of the infinitive to climb out to the main clause level. After it combines with the infinitival verb phrase, it removes a factor of 2 from the trace product, corresponding to the missing subject.

Recall that in constructing a clause which contains \(će\) ‘will’, the clitic version of \(hoće\), which takes an infinitival complement, we introduce a hypothesis corresponding to the control verb which requires an infinitival complement. We redefine \(IC\), the type of that trace as follows.

\[
\begin{aligned}
\text{IC} & = \prod_{i,k: \text{Nat}} (\text{NP}^2_{\text{nom}} \rightarrow S^2_{\text{inf},6}) \rightarrow \text{NP}^k_{\text{nom}} \rightarrow S^k_{m,6} \\
\end{aligned}
\]

A trace of this type combines with an infinitival verb phrase missing a subject, inside of which no clitics have been placed (clitic parameter 6). Whatever the trace...
product of the result type of its complement 2 is one of its factors (2i), corresponding to the missing subject. This trace then removes the subject trace 2 from the product, but adds one for itself, the clitic control verb, and adds the main subject’s parameter as a factor as well resulting in \( S_{m,6}^{2k} \). Phenogrammatically and semantically, the trace is the same as explained in Chapter 5.

We then refine the lexical entry for ‘će ‘will’ which controls an infinitival complement as follows:

\[
\begin{align*}
\lambda_P. (\text{wac } & \text{će}) (\lambda_z. (P \lambda_x. \text{PER}(x \circ y))) : ((z \rightarrow z \rightarrow z) \rightarrow z) \rightarrow z; \\
\prod_{k: \text{K}, i: \text{Nat}} [(\text{IC} \rightarrow S_{k,6}^{2i}) \rightarrow S_{k,5}^{i}]; \lambda_{F}. (F(\lambda_x. \text{FUT}(P x)) : (c \rightarrow p) \rightarrow p
\end{align*}
\]

Tectogrammatically, it removes a factor of 2 from its complement’s result type’s trace product, corresponding to the clitic ‘će ‘will’. Phenogramamatically, as in Chapter 5, it forces all non-clitic arguments of the infinitive and the infinitive itself to permute with main clause constituents. Below we show how to construct the proof of \( \text{Ana } \text{će } \text{ga } \text{poslati } \text{Marku} \), ‘Ana will send it to Marko’ focusing on tectogrammatical types only.

\[
\begin{align*}
\text{poslati } \text{‘to send’} & \\
\vdash & \text{NP}_{\text{dat}}^{1} \rightarrow \text{NP}_{\text{acc}}^{2} \rightarrow \text{NP}_{\text{nom}}^{2} \rightarrow S_{\text{inf},6}^{4} & \vdash & \text{NP}_{\text{dat}}^{1} \\
\text{NP}_{\text{acc}}^{2} \rightarrow & \text{NP}_{\text{nom}}^{2} \rightarrow S_{\text{inf},6}^{4} & \text{accusative clitic trace} & \text{NP}_{\text{acc}}^{2} \vdash & \text{NP}_{\text{acc}}^{2} \\
\text{NP}_{\text{acc}}^{2} \vdash & \text{NP}_{\text{nom}}^{2} \rightarrow S_{\text{inf},6}^{4} & \text{[\rightarrow E]} & \text{[\rightarrow E]}
\end{align*}
\]

At this point we cannot bind the clitic trace and place the accusative clitic. For one, we do not have lexical entries for clitics which can occur in the infinitival

\[285\]
verb phrases (not yet). More importantly, both a control verb which takes an infinitival complement and a corresponding trace require that no clitics have been placed inside the infinitival verb phrase because their argument type, for some $i$, is $(\text{NP}_{\text{nom}}^2 \rightarrow S_{\text{inf},i}^{2i})$, that is, the clitic parameter has to be 6. So, if we tried to place the accusative clitic, it would reduce the clitic parameter to 3, and then the control verb or its trace could not combine with it.

Instead, we introduce the control verb clitic hypothesis, and proceed to combine the result with the matrix clause subject. Then we place the verbal clitic, followed by the accusative clitic.

(427)

In this way, the clitics all cluster in the main clause and the sentence *Ana će ga poslati Marku*, ‘Ana will send it to Marko’ is pronounceable six different ways, because of the phenogrammatical representation of *će* which allows all non-clitics to freely permute.

286
When the clitics occur on the infinitive, however, the infinitival verb phrase behaves like a little embedded clause: it must remain contiguous and occur on the right periphery of the main clause, and the infinitive, onto which the clitics are encliticized, must be initial in the infinitival verb phrase, much like the complementizer in finite embedded clauses. As a result, the sentence *Ana će poslati ga Marku*, ‘Ana will send it to Marko’ can be pronounced just one way. *Ana hoće poslati ga Marku*, ‘Ana wants to send it to Marko’ in which the non-clitic *hoće* ‘wants’ occurs, can be pronounced two different ways because *Ana* and *hoće* can permute in the main clause.

In the spirit of infinitival verb phrases containing clitics behaving like clauses in terms of word order, we provide alternative lexical entries for infinitives which, phenogrammatically, construct sets of $S$-strings, and not just permutable $S$-strings. Below is the lexical entry for the version of *poslati* ‘to send’ which builds an infinitival clause inside of which clitics can be placed.

\[
\lambda_{x,y,w} w = \text{POSLATI}_z \circ x \circ y : z \rightarrow z \rightarrow z;
\]
\[
\Pi_{i,k: \text{Nat}} [\text{NP}^i_{\text{dat}} \rightarrow \text{NP}^k_{\text{acc}} \rightarrow \text{NP}^2_{\text{nom}} \rightarrow S^2_{\text{inf,6}}, \text{send} : e \rightarrow e \rightarrow e \rightarrow p]
\]

Infinitival verb phrase constructed by means of this lexical entry still cannot combine with the subject for tectogrammatical reasons: it requires that it subject be $\text{NP}^2_{\text{nom}}$, but no lexical noun phrases have parameter 2, only controlled subject traces.
Suppose we want to construct the infinitival verb phrase \textit{poslati ga Marku} ‘to send it to Marko’. We combine the infinitive with its dative argument, and an accusative clitic, resulting in the following sign:

\begin{equation}
\lambda_{wy,w} \vdash \text{POSLATI}_z \circ \text{MARKU}_z \circ y : z; \\
\text{NP}_{\text{acc}}^2 \rightarrow \text{NP}_{\text{nom}}^2 \rightarrow \text{NP}_{\text{inf},6}^2; \lambda_{yx}.\text{send marko } y x : e \rightarrow p
\end{equation}

After withdrawing the clitic hypothesis, we get the following sign:

\begin{equation}
\lambda_{wy,w} \vdash \text{POSLATI}_z \circ \text{MARKU}_z \circ y : z; \\
\text{NP}_{\text{acc}}^2 \rightarrow \text{NP}_{\text{nom}}^2 \rightarrow \text{NP}_{\text{inf},6}^2; \lambda_{yx}.\text{send marko } y z : e \rightarrow e \rightarrow p
\end{equation}

The pronominal clitics have to now be modified to tectogrammatically and semantically combine with such an infinitival verb phrase. Phneogrammatically, such a verb phrase is of the right type and the right form for the existing phenogrammatically representations of clitics to be placed correctly, onto the infinitive which constitutes the initial length one string in the infinitival verb phrase. So, below is the version of \textit{ga} ‘it’ which can tectogrammatically combine with this infinitival verb phrase:

\begin{equation}
\lambda_{wy,w} \vdash \text{POSLATI}_z \circ \text{MARKU}_z \circ y : z; \\
\text{NP}_{\text{acc}}^2 \rightarrow \text{NP}_{\text{nom}}^2 \rightarrow \text{NP}_{\text{inf},6}^2; \lambda_{yx}.\text{send marko } y z : e \rightarrow e \rightarrow p
\end{equation}

Combining this clitic with the infinitival verb phrase missing an accusative clitic, we get the following sign:

\begin{equation}
\lambda_{wy,w} \vdash \text{POSLATI}_z \circ \text{MARKU}_z \circ y : z; \text{NP}_{\text{nom}}^2 \rightarrow \text{NP}_{\text{inf},6}^2; \lambda_{yx}.\text{send marko } y : e \rightarrow p
\end{equation}

Next, we have to modify control verbs which combine with infinitival verb phrases inside of which the clitics are placed and so no clitic climbing is allowed. Below
we show the lexical entry of ho´ce ‘wants’ which will require that its infinitival complement not contain any clitic traces, preventing clitic climbing, and that it remain contiguous and at the right edge of the sentence.

\[(433) \vdash \lambda_{F_{y},[PER(y \circ \hbox{HO\'CE}_z) \bullet z \lambda_z.z = \hbox{toZ}(k \ F)]} : Z \rightarrow z \rightarrow Z;\]

\[
\prod_{i,j,k,n: \hbox{Nat}}[(\hbox{NP}^2_{\hbox{nom}} \rightarrow \hbox{S}^{3/5_{l,2}}_{\hbox{inf},n}) \rightarrow \hbox{NP}^i_{\hbox{nom}} \rightarrow \hbox{S}^{3/5_{l,2}}_{m,6}; \lambda_{P_{x}.\hbox{want}} x(P \ x) : c]
\]

Phenogrammatically, it permutes itself with the subject and then fuses the resulting S-language PER(y \circ \hbox{HO\'CE}_z) with the S-language obtained by compacting and then constructing a singleton S-language out of its complement \(F\). Tectogrammatically, it passes on all \(wh\) traces and quantificational traces, but it blocks clitic traces. Recall that \(c\) is defined to be \((e \rightarrow p) \rightarrow e \rightarrow p\).

The sentence Ana ho´ce poslati ga Marku, ‘Ana wants to send it to Marko’ is represented by the following sign:

\[(434) \vdash \hbox{PER}(\hbox{ANA}_z \circ \hbox{HO\'CE}_z) \bullet z \lambda_z.z = \hbox{toZ}(\hbox{wac} \ ga \ (\hbox{POSLATI}_z \circ \hbox{MARKU}_z \circ y)) : Z; \hbox{S}^1_{m,6}; \\hbox{want} \ hana \ (\hbox{send} \ hanka \ y \ hana) : p\]

To use the clitic version of ho´ce ‘wants’ which combines with a clause-like infinitival verb phrase, we first define the following tectogrammatical type. IC_{no} stands for ‘infinitival complement, no clitic climbing’.

\[(435) \quad IC_{no} =_{\hbox{def}} \prod_{i,j,k,n: \hbox{Nat}}[(\hbox{NP}^2_{\hbox{nom}} \rightarrow \hbox{S}^{3/5_{l,2}}_{\hbox{inf},n}) \rightarrow \hbox{NP}^i_{\hbox{nom}} \rightarrow \hbox{S}^{3/5_{l,2}}_{m,6}]
\]

This tectogrammatical type ensures that the infinitival verb phrase doesn’t have any clitic traces, just the subject trace. Its ultimate result type, \(\hbox{S}^{3/5_{l,2}}_{m,6}\), does contain a factor of 2 in the trace product, corresponding to the verbal clitic.
To use će ‘will’ which combines with a clause-like infinitival verb phrase, we introduce the following trace:

\[(436) \quad G : Z \to z \to z; IC_{no} ; G : c \vdash G : Z \to z \to z; IC_{no} ; G : c\]

Once poslati ga Marku ‘to send it to Marko’ combines with this trace, then with the main clause subject Ana, and finally the verbal clitic trace is bound, we get the following sign:

\[(437) \quad \vdash \lambda_{\gamma}.G ((\text{wac ga})(\lambda_{yw}.w = \text{POSLATI}_{z} \circ \text{MARKU}_{z} \circ y)) \text{ANA}_{z} : (Z \to z \to z) \to z; IC_{no} \to S_{m,6}^{2}; \lambda_{G}.(\lambda_{x}.(\text{send marko } y \ x)) \text{ ana} : c \to p\]

The lexical entry for će ‘will’ is given below:

\[(438) \quad \vdash \lambda_{\gamma}.(\text{wac će})(\lambda_{z}.h(\lambda_{iy}.(\text{PER } y) \bullet_{z} \lambda_{x}.x = \text{toZ}(k \ V))) : ((Z \to z \to z) \to z; \prod_{k \in K,i: \text{Nat}}[(IC_{no} \to S_{k,6}^{2}) \to S_{k,5}^{1}]; \lambda_{H}.(\lambda_{px}.\text{FUT}(P x)) : (c \to p) \to p\]

Phenogrammatically, this clitic permutes any permutable main clause constituents, but requires that the embedded infinitival verb phrase remain contiguous and at the right edge of the sentence. The sentence Ana će poslati ga Marku, ‘Ana will send it to Marko’ is represented in the grammar by the following sign:

\[(439) \quad \vdash (\text{wac će})(\lambda_{z}.(\text{PER } \text{ANA}_{z}) \bullet_{z} \lambda_{x}.x = \text{toZ}(k ((\text{wac ga})(\lambda_{yw}.w = \text{POSLATI}_{z} \circ \text{MARKU}_{z} \circ y))) : Z; S_{m,5}^{1}; \text{FUT}(\text{send marko } y \ \text{ana}) : p\]

In sum, requiring that the infinitives pass up all clitic traces to the matrix clause was straightforward. Allowing them to host clitics in which case the infinitival verb phrases truly exhibit clause-like behavior, required systematic additions of lexical entries for infinitives which build clause-like verb phrases, control verbs
which embed them (including their clitic versions), and clitics which can be placed inside such infinitival quasi clauses.

**Predicatives**

Recall that predicative complements cannot host any clitics, not even if they contain verbal forms such as participles. Consider the following example:

(440) a. Ana je vidjela Marka.
   Ana\textsubscript{NOM, f, sg, 3} be\textsubscript{3, sg} see\textsubscript{pl, f, sg} Marko\textsubscript{ACC, m, sg, 3}
   ‘Ana saw Marko’

b. * Ana vidjela je Marka.

c. Ana je Marka vidjela.

d. Marka je Ana vidjela.

e. Marka je vidjela Ana.

f. Vidjela je Ana Marka.

g. Vidjela je Marka Ana.

The participle can’t host its own clitics and it can freely permute with other constituents. Tectogrammatically, we represent the participle as follows:

(441) \( \vdash \) vidjela : \( \Pi_{i: \text{Nat}}[\text{NP}^i_{\text{acc}} \circ \text{NP}^2_{\text{nom}} \circ \text{Prd}^{2i}_{\text{pl}}] \)

The non-clitic version of the copula *jeste* ‘is’ is then tectogrammatically represented as follows:

(442) \( \vdash \) jeste : \( \Pi_{d:D, i,k: \text{Nat}}[(\text{NP}^2_{\text{nom}} \circ \text{Prd}^{2i}_{\text{pl}}) \circ \text{NP}^k_{\text{nom}} \circ S^{ik}_{m, 6}] \)

It passes all the traces that its predicative complement contains to the matrix level. To use the clitic version *je* ‘is’, recall that we need to introduce a trace corresponding to the copula. Below we redefine \textbf{PC} as the type of the copula trace:
Below is the tectogrammatical representation of the clitic je ‘is’:

\[
\vdash \text{je} : \prod_{k : \text{K}, i : \text{m} : \text{Nat}} [ (\text{PC} \rightsquigarrow S_{k,n>0}^{2i}) \rightarrow S_{k,0}^{i} ]
\]

Just like other clitics, it removes \(2\) as a factor of its complement’s trace product.

Below we show how to construct the proof of the sentence Ana ga je vidjela ‘Ana saw him’.

\[
\begin{array}{ccc}
\text{copula trace} & \text{vidjela ‘saw’} & \text{accusative clitic trace} \\
\hline \\
\vdash \text{PC} \vdash \text{PC} & \vdash \text{NP}_{\text{acc}}^{2} \rightarrow \text{NP}_{\text{nom}}^{2} \rightarrow \text{Prd}_{\text{pl}}^{4} & \vdash \text{NP}_{\text{acc}}^{2} \vdash \text{NP}_{\text{nom}}^{2} \rightarrow \text{Prd}_{\text{pl}}^{4} \rightarrow [\rightarrow \text{E}] \\
\hline \\
\vdash \text{NP}_{\text{acc}}^{2} \vdash \text{NP}_{\text{nom}}^{2} \rightarrow \text{S}_{m,6}^{4} & \vdash \text{PC} \vdash \text{NP}_{\text{acc}}^{2} \vdash \text{S}_{m,6}^{4} \rightarrow [\rightarrow \text{I}] & \\
\hline \\
\vdash \text{PC} \vdash \text{NP}_{\text{acc}}^{2} \rightarrow \text{S}_{m,6}^{2} & \vdash \text{PC} \vdash \text{NP}_{\text{acc}}^{2} \rightarrow \text{S}_{m,6}^{2} \rightarrow [\rightarrow \text{E}] \\
\hline \\
\vdash \text{PC} \vdash \text{NP}_{\text{acc}}^{2} \rightarrow \text{S}_{m,3}^{2} & \vdash \text{PC} \vdash \text{NP}_{\text{acc}}^{2} \rightarrow \text{S}_{m,3}^{2} \rightarrow [\rightarrow \text{E}] \\
\end{array}
\]

However, participles of subject control verbs have to be associated with multiple lexical entries, just as their finite counterparts. This is because they can (a) combine with a finite verb phrase, (b) combine with an infinitival verb phrase and force clitic climbing, or (c) combine with an infinitival verb phrase and treat their infinitival complement like a clause inside of which the clitics occur. Below are
lexical entries for the three versions of hoće ‘wants’. We omit the semantic representation because it is the same in all three cases.

(446) a. finite complement:
\[\lambda_{F_{vw}}.\exists_{xy}[(F \circ S \circ x) \land (\text{PER} \circ \text{HOĆE}_{z} \circ y) \land w = y \circ \text{toZ}(L \circ x)] : (z \rightarrow Z) \rightarrow z \rightarrow Z;\]
\[\Pi_{g:Gdr,i,j}:\text{Nat}[(N_{\text{nom},g,sg,3} \rightarrow S_{e,6}^{3/5}) \rightarrow N_{\text{nom},g,sg,3} \rightarrow S_{m,6}^{3/5/k}]; \text{want : c}\]

b. infinitival complement + clitic climbing:
\[\vdash \lambda_{xy}.\text{PER}(y \circ \text{HOĆE}_{z} \circ x) : z \rightarrow z \rightarrow Z;\]
\[\Pi_{g:Gdr,i,j}:\text{Nat}[(N_{\text{nom},g,sg,3} \rightarrow S_{i}^{2i}) \rightarrow N_{\text{nom},g,sg,3} \rightarrow S_{m,6}^{i/k}]; \text{want : c}\]

c. clause-like infinitival complement:
\[\vdash \lambda_{F_{y}}.\text{PER}(y \circ \text{HOĆE}_{z}) \bullet_{z} \lambda_{x,x = \text{toZ}(k \circ F)} : Z \rightarrow z \rightarrow Z;\]
\[\Pi_{i,j,k}:\text{Nat}[(N_{\text{nom},g,sg,3} \rightarrow S_{\text{inf},6}^{3/5}) \rightarrow N_{\text{nom},g,sg,3} \rightarrow S_{m,6}^{3/5/k}]; \text{want : c}\]

Below are the lexical entries for the three versions of the participle htjela ‘wanted’:

(447) a. finite complement:
\[\lambda_{G}.(\lambda_{v,v = \text{HTJELA}_{z}}) \bullet_{z} (G \circ S) : (z \rightarrow Z) \rightarrow Z;\]
\[\Pi_{i,j,k}:\text{Nat}[(N_{\text{nom},f,sg,3} \rightarrow S_{\text{inf},6}^{2i}) \rightarrow N_{\text{nom},f,sg,3} \rightarrow \text{Prd}_{pl}^{i/k}]; \text{want : c}\]

b. infinitival complement + clitic climbing:
\[\vdash \lambda_{x,\text{HTJELA}_{z}} \circ x : z \rightarrow z;\]
\[\Pi_{i,j,k}:\text{Nat}[(N_{\text{nom},f,sg,3} \rightarrow S_{\text{inf},6}^{2i}) \rightarrow N_{\text{nom},f,sg,3} \rightarrow S_{m,6}^{i/k}]; \text{want : c}\]

c. clause-like infinitival complement:
\[\vdash \lambda_{F_{y}}.\lambda_{v,v = \text{HTJELA}_{z}} \bullet_{z} \lambda_{x,x = \text{toZ}(k \circ F)} : Z \rightarrow Z;\]
\[\Pi_{i,j,k}:\text{Nat}[(N_{\text{nom},f,sg,3} \rightarrow S_{\text{inf},6}^{3/5}) \rightarrow N_{\text{nom},f,sg,3} \rightarrow S_{m,6}^{3/5/k}]; \text{want : c}\]

The second version of htjela ‘wanted’ in (b) above is the most straightforward one: it takes an infinitival verb phrase which is not clause-like (phenogrammatical type z) and concatenates itself with it. This means that the participle, the infinitive,
and any non-clitic complements of the infinitive are subject to the matrix verb’s permutation. The lexical entries given for the copula and its non-clitic version can combine with such participial constituents to give sentences like (440) above.

However, both versions of htjela ‘wanted’ which combine with clause-like constituents have to ensure that their complements occur on the right edge of the matrix clause. The participle itself can to some degree permute with main clause constituents. For example:

(448) a. Ana je htjela vidjeti ga.
   AnaNOM,f,s3 be3,s3 want_pl seeInf heACC,m,s3
   ‘Ana wanted to see him’

b. Htjela je vidjeti ga Ana.

(449) a. Ana je htjela da ga vidi.
   AnaNOM,f,s3 be3,s3 want_pl DA heACC,m,s3 see3,s3
   ‘Ana wanted to see him’

b. Htjela je Ana da ga vidi.

This means that we have to add additional lexical entries for the copula which can combine with predicative complements constructed by means of participles which combine with finite complements or clause-like infinitival complements.

Such predicative complements are phenogrammatically of type Z and have a more intricate phenogrammatical structure. We give the non-clitic jeste ‘is’ the following lexical entry:

(450) \( \lambda_{Vxw} \exists_{vy} [(Vv) \land (\text{PER}(x \circ \text{JESTE}_z \circ \text{toZ}(\text{fst}_s v)) y) \land w = y \circ (\text{rst}_s v)] : Z \rightarrow Z \rightarrow Z; \prod_{i,k,Nat} [\text{NP}^2_{\text{nom} \circ \text{Prd}^2_{ipl}} \rightarrow \text{NP}^k_{\text{nom} \circ S^k_{m6}}] ; \lambda_{PxP} \cdot Px : c \)
Phenogrammatically, *jeste ‘is’* takes off the first length one string of its complement, which has to be the participle itself. It permutes the participle and the subject with itself, but requires that its participle’s complement occur on the right edge of the clause.

Its clitic version is associated with the following lexical entry:

\[
\lambda_{hi} \cdot \text{wac} \cdot \text{je} \cdot (\lambda_{z} \cdot \text{H} \cdot (\lambda_{Vyx} \cdot \exists_{v} \cdot (v \cdot v) \wedge x = \text{PER}(y \circ (\text{fst}_{S} \cdot v) \circ (\text{rst}_{S} \cdot v))) : ((Z \to z \to Z) \to Z) \to Z; \prod_{i,k,i,n} [(\text{FC} \cdot \circ S_{k,n > 0}^{2i} \cdot \circ S_{k,0}^{i})]; \lambda_{P_{x}} \cdot P \cdot x : c
\]

Phenogrammatically, it takes off the participle (*fst*\(_{S} v*) and lets it permute with the subject, and requires that the rest of the participial complement (*rst*\(_{S} v*) occurs on the right edge. It encliticizes onto either the participle itself or some matrix clause constituent.

### 7.4.3 Interrogatives

The last thing we need to describe is how to prevent *wh* extraction out of interrogatives. We need to allow *in situ* *wh* traces to escape from embedded *wh* questions to allow for Baker’s ambiguity, but we have to block filler *wh* traces from originating in constituent questions. For polar questions, we must block both kinds of *wh* traces from originating inside of them. For both polar and constituent embedded questions, we cannot allow them to contain unresolved clitic traces. That is, enclitics cannot escape from questions.
Polar Questions

Recall that we analyzed two ways of constructing polar interrogatives in Serbo-Croatian: one which involves the interrogative complementizer *da li*, and the other one which requires the finite verb to occur question-initially with *li* encliticized onto it.

Below is the tectogrammatical representation of *da li* sensitive to the traces that its complement contains:

\[(452) \vdash (\text{da li}) : \Pi_{i: \text{Nat}} [S_{m,6}^{2i} \rightarrow S_{q,6}^{2i}]\]

The interrogative complementizer allows its complement clause’s trace product to only consist of factors of 2, corresponding to the clitics. Therefore, no *wh* traces are allowed.

A question forming verb, such as *voli li* in *Voli li Ana Marka?* ‘Does Ana love Marko?’ has the following tectogrammatical representation:

\[(453) \vdash (\text{voli li}) : \Pi_{i: \text{Nat}} [\text{NP}_{\text{acc}}^{2i} \rightarrow \text{NP}_{\text{nom}}^{2i} \rightarrow S_{q,6}^{2i}]\]

It explicitly requires that its subject be either a lexical noun phrase or a quantificational trace, and that its object have the parameter expressible as $2^i$, that is, in practice, either 1 (if it’s a lexical noun phrase or a quantificational trace) or 2 (if it’s a clitic trace). So, we are explicitly ruling out any *wh* traces. In sum, expressions which form polar questions do not allow there to be any *wh* traces at all.

Below we show how exactly the proof of *Da li Ana voli koga?* ‘Does Ana love who?’ fails.
(454)

\[ \Pi_{i: \text{Nat}} [S_{m,6}^{2i} \rightarrow S_{q,6}^{2i}] \]

\[ \text{Ana} \quad \vdash \text{NP}_{\text{nom}}^i \rightarrow \text{NP}_{\text{nom}}^i \rightarrow S_{m,6}^5 \]

\[ \text{voli} \text{‘loves’} \quad \text{in situ wh trace} \]

\[ \text{NP}_{\text{acc}}^5 \vdash \text{NP}_{\text{acc}}^5 \rightarrow S_{m,6}^5 \]

\[ \rightarrow \text{E} \]

The problem is that there is no way to instantiate \( i \) in the type for the complementizer so that it can combine with a sentence whose trace product is 5; 2 is not a factor of 5. So the proof fails.

**Constituent Questions**

As discussed earlier in this chapter, we analyze \( \text{wh} \) filler expressions as combining with declarative clauses with a \( \text{wh} \) filler trace, and \( \text{wh in situ} \) expressions as combining with questions missing an \( \text{in situ} \ \text{wh} \) trace. We repeat the tectogrammatical representations of the filler \( \text{wh} \) expression \( \text{ko} \text{‘who_{NOM}’} \) and the \( \text{in situ} \ \text{wh} \) expression \( \text{koga} \text{‘who_{ACC}’} \) considered earlier.

(455)

a. \( \vdash \text{ko} : \Pi_{n: \text{Nat}} [ (\text{NP}_{\text{nom}}^3 \rightarrow S_{m,6}^{3n} ) \rightarrow S_{q,6}^n ] \)

b. \( \vdash \text{koga} : \Pi_{n: \text{Nat}} [ (\text{NP}_{\text{acc}}^5 \rightarrow S_{q,6}^{5n} ) \rightarrow S_{q,6}^n ] \)

Question embedding expression allow \( \text{wh in situ} \) traces to be passed on, but block any \( \text{wh} \) filler traces as well as clitic traces. Below is the tectogrammatical representation of the verb \( \text{pita} \text{‘wonders’} \) which we analyzed in Chapter 6 as requiring an inherent reflexive.

(456)

\( x : \text{SE} \vdash \text{pita} : \Pi_{i,k,n: \text{Nat}} [S_{q,n}^{5i} \rightarrow \text{NP}_{\text{nom}}^k \rightarrow S_{m,6}^{5i/2k}] \)

297
Its complement is only allowed to have trace product expressible as $i^j$ which means that it can contain only quantificational and *wh in situ* traces. This verb builds a clause which has 2 as a factor, corresponding to the inherent reflexive.

Below we show how *wh* extraction out of a *wh* question fails, so that "$\text{Ko se Ana pita koga voli 'Who}\text{NOM} \text{ does Ana wonder who}\text{ACC} \text{ loves?'}$" is correctly excluded from the grammar.

\[(457)\]

\[
\begin{align*}
\text{wh filler trace} & \quad \exec{\text{voli 'loves'}} \quad \text{wh filler trace} \\
\text{NP}_\text{nom}^3 \vdash \text{NP}_\text{nom}^3 & \quad \text{NP}_\text{acc}^3 \vdash \text{NP}_\text{nom}^3 \rightarrow \text{S}_\text{m,6}^6 & \quad \text{NP}_\text{acc}^3 \vdash \text{NP}_\text{acc}^3 \rightarrow \text{S}_\text{m,6}^6 \\
\text{NP}_\text{acc}^3, \text{NP}_\text{nom}^3 \vdash \text{S}_\text{m,6}^6 & \quad \text{NP}_\text{nom}^3 \vdash \text{NP}_\text{acc}^3 \rightarrow \text{S}_\text{m,6}^6
\end{align*}
\]

\[
\begin{align*}
\text{wh filler} & \\
\vdash (\text{NP}_\text{acc}^3 \rightarrow \text{S}_\text{m,6}^6) & \rightarrow \text{S}_\text{m,6}^3 & \vdash (\text{NP}_\text{nom}^3 \rightarrow \text{S}_\text{m,6}^6) \rightarrow \text{S}_\text{m,6}^3
\end{align*}
\]

The problem is that there is no way to instantiate $i$ in the type of *pita se 'wonders'* so that it can combine with this question; no matter what $i$ we pick 3 cannot be a factor of $i^j$. So the proof fails, and *wh* extraction out of constituent questions is prevented.

### 7.5 Conclusion

In this chapter, we discussed various locality constraints in Serbo-Croatian. We noted that in a logical grammar with hypothetical reasoning such as ours, locality-based overgeneration stems from the inability of the grammar to control directly
when the hypothetical reasoning rule can apply. We systematically considered
different kinds of constituents and different kinds of traces, to map out precisely
what kind of trace can be passed up from what kind of clause to the matrix clause.

We then proposed a simple mechanism for enforcing locality constraints. It in-
volved decorating each kind of noun phrase and noun phrase trace with a number,
and having clause-forming expressions construct products of their arguments’
number parameters. Each trace type that the grammar has to be sensitive to,
such as \textit{wh} traces, clitic traces and controlled subject traces, was decorated with a
prime number, so that from a clause’s trace product we could recover the informa-
tion about what types of traces it contained, and how many of each type of trace.
In this manner, we have substantially refined the grammar presented in earlier
chapters to prevent clitic climbing, extraction from questions and long distance
control.

We have also refined our theory of control in general, by carefully distinguishing
between (i) finite control when the control verb is an I-verb, (ii) finite con-
trol when the control verb is an S-verb, (iii) infinitival control when the infinitive
doesn’t host the clitics, and (iv) infinitival clitics when the infinitival verb phrase
behaves like a clause in that it hosts its own clitics and remains contiguous, at the
right edge of the matrix clause. This caused some systematic additions to the rest
of the system, namely finer distinctions among lexical entries of subject control
verbs, including their participial and clitic counterparts.
The mechanism we proposed for enforcing locality constraints is straightforward, mathematically very simple, and can be added to any logical grammar whose tectogrammatical type system admits dependent types. Further, we believe that such a mechanism could easily translate to grammars of other languages where similar locality constraints have to be enforced.
Chapter 8: Conclusion

We wanted to construct a theory of Serbo-Croatian grammar, expressed in a logical formalism. We put together a suitable formalism, with three distinct components, namely, phenogrammar, tectogrammar and semantics. The phenomenological component of the grammar was elaborate, but it allowed us to represent and combine expressions at different levels of the type hierarchy, and manipulate them accordingly via reordering or order-inducing functions. In this way, we were able to exercise a great deal of control over word order.

Our grammar generates a substantial fragment of Serbo-Croatian. We analyzed declarative and interrogative clauses, subject and object control and predicative structures, and proclitics and enclitics. Still, many important empirical phenomena were not considered, including topicalization and relative clauses, among many others.

While we strove to provide a fairly detailed analysis, there are still many loose ends. For example, although we discussed various postnominal modifiers, we
had nothing to say about the order of those modifiers in noun phrases, in case more than one occur in the same phrase.

Our theory is surely a wrong theory of Serbo-Croatian grammar, but we believe the project to be a success nonetheless, both from the empirical and the theoretical perspective. While there are outstanding problems with our theory and just as many empirical phenomena we had nothing to say about, this is still the most comprehensive and detailed descriptive grammar of Serbo-Croatian that we know of.

Further, it is, to the best of our knowledge, the most comprehensive grammar of any Slavic language in the categorial grammar tradition. We think this is important because it shows that it is possible to devise a usable categorial grammar which can deal with complex word order and agreement patterns, not really present in English or other related languages that a lot of categorial work has been on.

We also devised a theory of 2P clitics in a categorial grammar, which hasn’t been done before, to the best of our knowledge. Since Serbo-Croatian enclitics cannot be anchored to any specific constituent type and can be hosted by a variety of constituents so long as they occur second in the clause, we believe that having an architecture which has a phenogrammatical level of representation helped us give an empirically adequate and conceptually satisfying analysis. In virtue of
having a distinct phenogrammatical component, we were able to arrive at a single theory-internal phenogrammatical generalization concerning the placement of the enclitic cluster in various types of clauses. Phenogrammar and tectogrammat jointly helped us enforce the correct placement of the clitic cluster, which accords with our belief that clitic placement is a syntax/word order interface phenomenon.

Finally, we sketched a tectogrammatical mechanism for enforcing locality constraints. With it, we were able to exert a great deal of control over under what circumstances different kinds of traces can or must be bound, or passed onto the matrix level. For example, we were able to allow clitic climbing when needed, but prevent it in most cases; to ensure that control is a local phenomenon; and to prevent extraction from interrogatives.

It is our hope that eventually our theory of grammar could be extended to include other empirical phenomena in Serbo-Croatian, particularly negative concord. We also hope that parts of our analysis could be generalized to formal grammars of other languages, such as our analysis of 2P clitics and locality constraints.

We hope that we succeeded in (re)raising important theoretical questions concerning grammatical architecture as well as word order, particularly with respect to the assumption that there is a ‘canonical’ word order from which ‘non-canonical’ orders are derived, and that different word order is a sufficient criterion for distinguishing linguistic expressions.
Bibliography


Mihalicek, Vedrana. 2010c. (Not so) free word order in Lambda Grammar: The case of BCMS. Presented at Formal Approaches to South Slavic and Balkan Languages, Dubrovnik, Croatia, October 4-6, 2010.


