NOTICE

AVIS

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

If pages are missing, contact the university which granted the degree.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

On the Constituent Structure of Slovene

by

Sandi Kodrič

B.Sc., University of Ljubljana, 1991

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in the School

of

Computing Science

National Library
of Canada

Bibliothèque nationale
du Canada

Acquisitions and
Bibliographic Services Branch

Direction des acquisitions et
des services bibliographiques

395 Wellington Street
Ottawa, Ontario
K1A 0N4

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

THE AUTHOR HAS GRANTED AN
IRREVOCABLE NON-EXCLUSIVE
LICENCE ALLOWING THE NATIONAL
LIBRARY OF CANADA TO
REPRODUCE, LOAN, DISTRIBUTE OR
SELL COPIES OF HIS/HER THESIS BY
ANY MEANS AND IN ANY FORM OR
FORMAT, MAKING THIS THESIS
AVAILABLE TO INTERESTED
PERSONS.

L'AUTEUR A ACCORDE UNE LICENCE
IRREVOCABLE ET NON EXCLUSIVE
PERMETTANT A LA BIBLIOTHEQUE
NATIONALE DU CANADA DE
REPRODUIRE, PRETER, DISTRIBUER
OU VENDRE DES COPIES DE SA
THESE DE QUELQUE MANIERE ET
SOUS QUELQUE FORME QUE CE SOIT
POUR METTRE DES EXEMPLAIRES DE
CETTE THESE A LA DISPOSITION DES
PERSONNE INTERESSEES.

THE AUTHOR RETAINS OWNERSHIP
OF THE COPYRIGHT IN HIS/HER
THESIS. NEITHER THE THESIS NOR
SUBSTANTIAL EXTRACTS FROM IT
MAY BE PRINTED OR OTHERWISE
REPRODUCED WITHOUT HIS/HER
PERMISSION.

L'AUTEUR CONSERVE LA PROPRIETE
DU DROIT D'AUTEUR QUI PROTEGE
SA THESE. NI LA THESE NI DES
EXTRAITS SUBSTANTIELS DE CELLE-
CI NE DOIVENT ETRE IMPRIMES OU
AUTREMENT REPRODUITS SANS SON
AUTORISATION.

ISBN   0-612-01124-0

Canada

APPROVAL

Name: Sandi Kodrič

Degree: Master of Science

Title of thesis: On the Constituent Structure of Slovene

Examining Committee: Dr. James J. Weinkam
Chair

Dr. Frederick P. Popowich
Senior Supervisor

Dr. Paul G. McFetridge
Department of Linguistics
Supervisor

Dr Thomas E. Hukari
Associate Professor
Department of Linguistics
University of Victoria
External Examiner

Date Approved: _December 7, 1993_

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

On the Constituent Structure of Slovene.

_____

_____

_____

Author: _____
(signature)

_____Sandi Kodric_____
(name)

_____DEC. 9, 1993_____
(date)

Abstract

Many natural languages exhibit a much higher degree of freedom than English in the ordering of constituents within a clause. In order to use these so-called 'free word order' languages in natural language processing applications, we need grammar models that are adequate from both a linguistic and a computational point of view.

I examine Slovene, one such language, and propose that it is best treated by flattening the traditional hierarchical syntactic structure. I argue that there is little empirical evidence for the finite verb phrase constituent in the clause, and I show that several problems disappear if this assumption is rejected. Instead, I present a model whereby the verb combines with its subject and other complements in a single step. In both finite and nonfinite clauses, there is only one verbal projection in the syntactic structure.

The clause structure is described within the constraint-based formalism of Head-driven Phrase Structure Grammar (HPSG) which is not confined to binary branching and which includes separate specifications of immediate dominance and linear precedence. This analysis of the Slovene clause avoids discontinuous constituency and it accounts for the invariable second position of the clitic cluster by local linear precedence constraints.

I discuss some computational consequences and suggest that the weakened structural constraints do not necessarily result in less efficient parsing. I show how an existing chart parser for HPSG can be adapted to process grammars of this kind.

Acknowledgements

Contents

List of Tables

List of Figures

List of Abbreviations

1	1st person
2	2nd person
2P	clause-second position
2P-CL	second position clitic
3	3rd person
ACC	accusative
ADJ	adjective
AUX	auxiliary
AVM	attribute-value matrix
C	complement
CAT	category
CFG	context-free grammar
CL	clitic
COMP	complementizer
COMPS	complements
CONT	content
DAT	dative
DET	determiner
DTRS	daughters
F	full
FEM	feminine

FS	feature structure
GB	Government-Binding theory
GEND	gender
GPSG	Generalized Phrase Structure Grammar
H	head
HPSG	Head-driven Phrase Structure Grammar
ID	immediate dominance
INF	infinitive
LEX	lexical
LOC	local
LP	linear precedence
MASC	masculine
NOM	nominative
NP	noun phrase
NUM	number
OBJ	object
PER	person
PHON	phonology
PL	plural
PP	prepositional phrase
PRES	present tense
PSP	past participle
QUE	question particle
S	sentence, subject
SG	singular
SUBCAT	subcategorization
SUBJ	subject
SVO	subject-verb-object

SYNSEM	syntax-semantics
UG	Universal Grammar
V	verb
VFORM	verb form
VP	verb phrase

Chapter 1

Introduction

Language is the primary means of communication and recording information. Its huge potential for conveying complex ideas succinctly makes it difficult to understand the underlying cognitive processes. *Computational linguistics* can be seen as the study of language processes as applied to machines. It aims to make it possible for computers to generate and interpret natural language.

1.1 Background

This field draws from several disciplines. *Linguistics* seeks to explain grammatical competence—how people come to accept some sentences as grammatical and reject others as ungrammatical. Linguists are looking for language universals, principles that apply to all natural languages. They are trying to find the simplest, most restricted formalized theory of grammar which can account for natural language. The goal is an explanation (description) that covers the most facts with the fewest assumptions and postulations while keeping in mind the range of variation across languages.

Apart from the purely theoretical side, computational linguistics has an 'engineering' side to it. Some researchers are thus concerned with the development of practical

computer systems that deal with natural language: machine translation systems, information retrieval systems, man–machine interfaces, speech recognition/synthesis systems etc. In this 'engineering' subfield, other issues are of primary concern: how to represent knowledge of language and how to use it efficiently in language processes. In other words, how to design efficient algorithms for the various components of a natural language processing system. In this sense, we can view the area of *artificial intelligence* as another 'parent field,' from which several concepts have been inherited. Computational linguistics is thus an interaction of various fields which are concerned with different aspects of language.

Language expresses a relation between sound at one end and meaning at the other— it is structured in several levels. Phonology, for instance, is concerned with the study of speech sounds and relations between them, whereas semantics, on the other side, investigates the meanings of utterances and their parts. The *syntactic* level, which is addressed in this thesis, studies the ways in which words are put together to form phrases, clauses and sentences.

1.2   Motivation

In the recent decades, syntax has been viewed as the core subject of linguistics and it has been extensively investigated. Nonetheless, it remains an area of several competing theories. Even though the theories agree in various respects, one feature that distinguishes them into *transformational* and *nontransformational* is the issue of levels of syntactic representation. Transformational accounts, such as Government-Binding (GB) theory (Chomsky 1981, 1986), assume a sequence of levels and they make use of transformations between the levels. On the other hand, nontransformational theories, such as Generalized Phrase Structure Grammar (GPSG, Gazdar et al. 1985) and Head-driven Phrase Structure Grammar (HPSG, Pollard and Sag 1987, in press), posit only one level of syntactic representation and with it they associate well-formedness conditions of various kinds.

Most current syntactic theories which are to some extent influenced by the original $\bar{\text{X}}$ theory (Jackendoff 1977) are not well suited for the so-called *free word order* languages which exhibit freedom in the ordering of constituents within a clause, as pointed out by Hawkins (1983) and Uszkoreit (1987), among others. The freedom is traditionally explained by the fact that grammatical relations (subject, object etc.) are encoded morphologically (e.g., by inflections in Slavic languages) rather than by linear order as in English. In this sense, most syntactic theories suffer from Anglocentrism, stemming from the influence that the structure of English had on the design of various grammatical frameworks (Uszkoreit 1987:3).

Perhaps the most studied languages with at least partially free word order have been German and Japanese. Uszkoreit (1987) and Gunji (1987) give comprehensive nontransformational accounts of these languages in the GPSG tradition, but both approaches rely considerably on fixed verb positions in their respective languages.

1.3 Thesis overview

Slovene is a South Slavic language with a relatively high degree of freedom in constituent ordering and it also lacks a fixed position of verbs in clauses. The research of Slovene syntactic properties has been rather scarce and many phenomena are still not well understood. From the perspective of linguistic theory, there have been a few investigations in the transformational paradigm (see Golden (1990) for a brief survey), and only some preliminary work in nontransformational phrase structure approaches (Mežnarič 1992).

This thesis is concerned with word order and constituent structure of the Slovene language. More precisely, I will propose a nontransformational model of the simple clause. The thesis is organized as follows. I begin, in chapter 2, by presenting some relevant background. I examine the factors that influence the ordering of constituents in Slovene clauses, and present the difficulties that we encounter if we adopt the standard hierarchical assumptions about the phrase structure. In chapter 3, I go on to

investigate empirical evidence for the clause structure in Slovene. I argue that the assumption of a finite verb phrase constituent in the structure is not sufficiently motivated. I show that some problems disappear, if this assumption is rejected. Instead, I propose a uniform treatment of both finite and nonfinite clauses where verbs combine with all of their arguments in a single step. I then outline an analysis of a simple clause within Head-driven Phrase Structure Grammar (HPSG) in chapter 4. I describe the corresponding rule schema and linear precedence constraints. In chapter 5, I discuss some consequences of the flattened clause structure from the computational perspective. Finally, in chapter 6, I summarize the results and compare this approach with some others.

Chapter 2

Preliminaries

Syntactic structure is traditionally represented by *tree diagrams*, and within most phrase structure approaches a few restrictions are usually assumed. Consider the tree in Figure 2.1.

```
                    S
           _____/ \
          NP              VP
      ___/ | \___       |  \___
    DET   ADJ   N     V       NP
     |     |    |     |        |
    The  brown dog  chased   Fido.
```
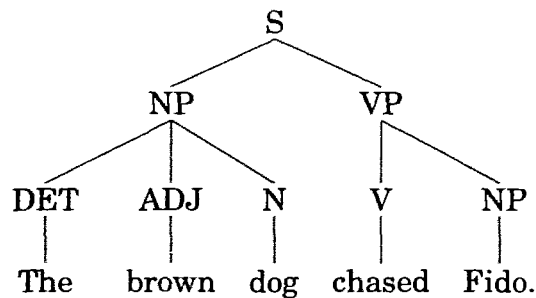
Figure 2.1: Syntactic tree diagram

Terminal nodes (leaves of the tree) represent the sequence of words in the sentence. Nonterminal nodes are labelled with the categorial information, and the tree is rooted in the S (sentence) node. The 'non-tangling' condition means that there should be no crossed branches in the structure[1] (every node in the tree must represent a contiguous substring of the sentence). A *constituent* (a unit of syntactic structure) is any section

---

[1]A notable exception to this condition is Discontinuous Phrase Structure Grammar (DPSG) which allows crossed branches (Bunt 1990).

of the tree that has a single mother (i.e., any subtree).

But there are many phenomena across languages which conflict with these assumptions. For example, (1) illustrates an extraposition of a relative clause in English.

(1) <u>Someone</u> entered the room <u>who I have never met before</u>.

The phrase 'someone who I have never met before' is discontinuous in the sentence. The 'logical' structure of (1)—but one which does not obey the 'non-tangling' restriction—is depicted in Figure 2.2.
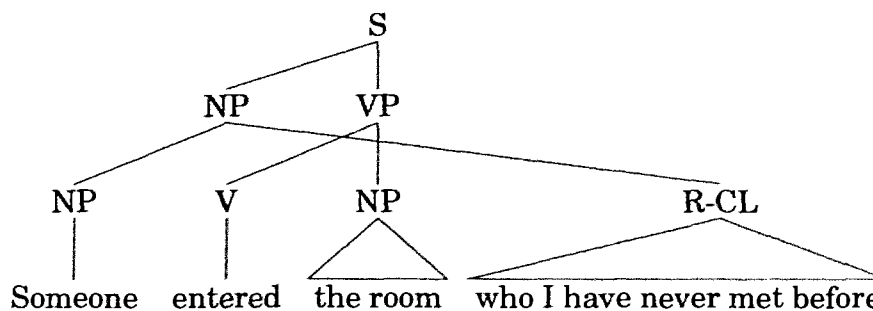
Figure 2.2: Structure with crossed branches

Several ways of accommodating such exceptions to the above assumptions have been proposed. We might:

(a) accept that some syntactically determined (underlying) order is 'scrambled' in the surface string; or

(b) reject the surface word order as the basis for deriving constituent structure; or

(c) 'loosen' the hierarchical constituent structure.

The first option is adopted in most GB-related approaches to flexible word order (see, for instance, Grewendorf and Sternfeld (1990) for German). The second alternative is pursued by Reape (1990) and by Dowty (1990) for German and English, respectively.

In this thesis, however, I will be examining the third option (c). Unlike (a), this option is available to nontransformational accounts with a single level of representation

and it does not depart from the standard view on phrase structure like (b) does (see also chapter 6).

I will first give some relevant facts on Slovene. In section 2.2, I address some general word order issues and then emphasize some word order properties of Slovene that present a problem for the hierarchical approach.

2.1 Basic facts on Slovene

The Slovene language has a rich inflectional morphology. Inflections determine morphological case of noun phrases (NP) which reflects their grammatical relations.[2] With respect to the typological criteria, Slovene is classified as having SVO (subject-verb-object) 'basic order' (Hawkins 1983:338). It shares many properties with other Slavic languages (notably Serbo-Croatian), but the influence of other neighbouring languages can also be seen, as noted by Bennett (1987:281):

> Within the vocabulary, phonology and grammar of Slovene there is ample evidence of its close contact with German and the Central European *Sprachbund* in general (as opposed to the Balkan *Sprachbund*).

Example (2) shows a simple sentence with a nominative subject and an accusative object. Neutral SVO order (2a) is usually preferred in the absence of discourse-related factors (see section 2.4), but other orderings (2b–2d), obtained by 'shuffling,' are grammatical as well, albeit somewhat 'marked.'

(2) a. Janez   piše   pismo.

      John-NOM writes letter-ACC

      'John is writing a letter.'

   b. Pismo   piše   Janez.

      letter-ACC writes John-NOM

---

[2]I will only be concerned with NP complements.

   c. Janez   pismo   piše.
      John-NOM letter-ACC writes

   d. Pismo   Janez   piše.
      letter-ACC John-NOM writes

Past tense is a compound tense in Slovene and it is formed by the present tense form of the auxiliary *biti* 'to be' and the active past participle of the main verb. The participle and the auxiliary agree in number. Additionally, the participle agrees with the subject in gender and number, and the auxiliary agrees with the subject in person and number. In (3), note that the auxiliary must appear in the second position, but the three other constituents may appear in any order. Again, most of the permutations are perceived as marked to various degrees, if used outside an appropriate discourse situation.

(3) a. Janez   je   bral   knjigo.
       John-NOM AUX-3SG read-SG,PSP,MASC book-ACC

       'John read a book.'

   b. Janez   je   knjigo   bral.
      John-NOM AUX book-ACC read

   c. Knjigo   je   bral Janez.
      book-ACC AUX read John-NOM

   d. Knjigo   je   Janez   bral.
      book-ACC AUX John-NOM read

   e. Bral je   knjigo   Janez.
      read AUX book-ACC John-NOM

   f. Bral je   Janez   knjigo.
      read AUX John-NOM book-ACC

The above examples illustrate the following properties:

1. in compound tenses, the auxiliary invariably appears in the second position;

2. position of the main verb is unrestricted;

3. there are no precedence or adjacency constraints either between the main verb and the auxiliary or between the main verb and any of its arguments.

2.2 Word order

Word order in Slavic languages is affected by the following factors (Ard 1975:95):

(a) grammatical rules of the language: they are relevant for the 'fixed' word order;

(b) discourse situation: it influences the 'free' word order;

(c) rhythmic and stylistic principles.

Their mutual interaction is mainly such that (b) and (c) operate only where (a) leaves some latitude. Indeed, speakers find a sentence 'more unacceptable' if (a) is violated, compared to the violations of (b) or (c). However, the relative 'strength' of particular principles is specific to the language in question, and it usually varies among its speakers.

'Fixed' word order also encompasses certain positions in the clause (such as initial, final or preverbal) which are reserved for elements with a special significance (Dik 1989:359). In Slovene, 'fixed' word order comprises phrase-internal structure (e.g., placement of prepositions, conjunctions and modifiers) and an invariable placement of clitics. Apart from that, the ordering is largely governed by discourse principles (see section 2.4).

2.3 2P cluster

Even though the function of word order in Slovene is to a large extent pragmatic, some syntactic ordering constraints hold, nonetheless. Slovene is one of the languages that obey the Wackernagel Law which informally refers to the following property:[3]

(4) If the clause contains clitics of certain types, these clitics always group together in a *clitic cluster* which occupies the *second position* (2P) in the clause.

The 'second position' originally meant 'after the first word,' as noted by Anderson (1993:72), and the class of 'clitics' was identified with that of accentless elements. In Common Slavic, the clitics were attached to the first accented word of a clause (Ard 1975:96). In this respect, Modern Slovene has undergone a change so that the corresponding items are now positioned on a syntactic rather than phonological basis (Bennett 1987:271).

The clitics immediately follow the *first constituent* of a clause, irrespective of how complex it is. The closely related South Slavic Serbo-Croatian exhibits alternation between the 'syntactic' and 'phonological' position: the cluster can in some cases follow either the first word or the first constituent (Halpern 1992). But this is not the case in Slovene.

Exceptionally, the cluster can be preceded by the clause-initial lexical item (such as an interrogative particle), or it may appear in the initial position itself in the case of ellipsis (Toporišič 1984:539). Phonologically, the 2P cluster is attached either to the preceding or to the following accented word (Toporišič 1984:535).

The class of 2P elements in Slovene includes reduced (weak) object pronominals (genitive, dative, accusative) and various forms of the auxiliary *biti* 'to be.' As is common in 2P languages, the cluster has a rigid internal order which contrasts with a relative freedom elsewhere. For example, in (5) the dative object clitic must precede the accusative one (2P clitics will be italicized in the following examples).

---

[3]Named after J. Wackernagel who first pointed out this placement of clitics (Wackernagel 1892).

(5) a. Janez *mu   ga   je* dal.

John-NOM him-DAT it-ACC AUX given

'John gave it to him.'

b. *Janez *ga   mu   je* dal.

John-NOM it-ACC him-DAT AUX given

c. Dal *mu   ga   je* Janez.

given him-DAT it-ACC AUX John-NOM

Clitic pronouns are in complementary distribution with full object pronouns which have the same distributional properties as NPs. If the full (strong) pronoun *njega* is used instead of the clitic form *ga*, the full pronoun must appear outside the clitic cluster. Compare (5a), (6a) and (6b):

(6) a. *Janez *mu*   njega *je* dal.

John-NOM him-DAT,CL it-ACC,F AUX given

b. Janez *mu   je* dal njega.

John-NOM him-DAT,CL AUX given it-ACC,F

'John gave it to him.'

But unlike Serbo-Croatian (Kaisse 1985:91), there are no full (non-2P) auxiliaries in Slovene. Correspondingly, Slovene auxiliaries can receive contrastive stress; pronominal clitics, in contrast, can not.

The rules governing the internal order of the cluster are quite complex (see Orešnik 1985). For example, the 3rd person, singular present auxiliary *je* follows the pronominals in (5), whereas other present auxiliaries precede them (7).

(7) Ti   *si   mu   ga* dal.

you-NOM AUX-2SG him-DAT it-ACC given

'You gave it to him.'

Mežnarič (1992) describes generalized principles of linear precedence between the various classes of clitics. She proposes a T(C)X structure for the Slovene clause, whereby some fronted constituent or lexical item in the topic position (T) immediately precedes the optional[4] clitic cluster (C) which is in turn followed by everything else (X). This structure resembles the TVX structure of verb-second clauses found in Germanic languages. It has to be borne in mind though, that the clitic cluster here is not a constituent, and it is not sensitive to categorial nature of its phonological host.

2.4 Discourse considerations

Various factors have been identified which influence word order and are related to communication process between the participants in discourse (Yokoyama 1986). For example, a substantial difference is manifested between discourse-initial and non-initial utterances. Following the Prague School tradition, Toporišič (1982) states that the 'free' word order (as opposed to the 'fixed' part of word order) is governed by the principles of Functional Sentence Perspective (Sgall et al. 1973) in that the sentence parts are ordered according to their communicative importance. His division recognizes three segments:

- theme – 'what the sentence is about' (often the 'given' information);

- rheme – 'what is said about the theme' (implying novel information);

- transition – especially in longer clauses, this part is distinguished as part of the rheme; it is the least important part of the message and can often be omitted.

The unmarked sequence in main declarative clauses is:

(8) theme < rheme

Thus, from the communicative point of view, the less important part precedes the more important part. In interrogative sentences, on the other hand, the sequence in

---

[4]Recall that sentences such as (2) contain no 2P clitics.

reversed. Toporišič himself notes that this division often cannot be made unequivocally, particularly in longer and more complex sentences.

In a more recent Prague School approach, Hajičová et al. (1993) employ a scale of 'communicative dynamism' (underlying word order) to divide the utterance into 'topic' part and 'focus' part (topic/focus articulation) where the topic carries the contextual or given information and focus represents the new, salient piece of information intended to update the hearer's knowledge. Focus can be cross-linguistically indicated on different levels of language organization. In 'free word order' languages, it is often signalled by word order. In Slovene, the clause elements are usually arranged so that the 'intonation center,' which coincides with the focus, falls on the last accented word (Toporišič 1984:448); Dik (1989:364) observes the same state of affairs in Polish, Czech and Bulgarian.

As mentioned above, some orderings are only appropriate in certain restricted contexts. Recall the sentence (2b) repeated here as (9).

(9) Pismo   piše   Janez.

    letter-ACC writes John-NOM

    'John is writing the letter.'

Ordered in this way, (9) presupposes[5] a context where the speaker expects the hearer to be already aware of the existence of the letter and the act of writing. For example, it is perfectly acceptable as the answer to the question in (10).

(10) Ali Ana   piše   pismo?

    QUE Ann-NOM writes letter-ACC

    'Is Ann writing the letter?'

Hence, (9) is more accurately translated into English by the cleft sentence in (11).

(11) It is John who is writing the letter. (i.e., not someone else)

---

[5]Here it is also assumed that no word in this sentence bears additional stress (see below).

Thus, what in English is obtained by syntactic devices like clefting, can in Slavic languages be achieved by rearranging the constituents. Additionally, like in many other languages, the effect of (9) can also be accomplished by the focus being signalled prosodically by employing an emphatic stress on the particular word in the utterance (12) and thereby moving the intonation center.

(12) JANEZ piše pismo.
     JOHN-NOM writes letter-ACC
     'JOHN is writing the letter.'

To summarize, constituent order is in interaction with the pragmatic situation established by the previous context and with the prosodic structure of the clause. The cluster of 2P clitics (if any) uniformly occupies the second position in the clause. The rest is essentially unrestricted—major constituents of the clause are ordered according to their communicative prominence: the new (most important) piece of information generally comes last.

Chapter 3

Clause structure

In this chapter, I first discuss the issue of configurationality and then examine the VP constituency in Slovene. In section 3.4, I conclude that, if we take into account the properties of the subject, there is no motivation for a finite VP node (separate from S). I then summarize the proposed model of the Slovene clause.

3.1 Configurationality

It has been suggested in recent years that languages with relatively unconstrained word order have nonconfigurational syntactic representations (Chomsky 1981, Hale 1983); as opposed to the 'ordinary' hierarchical or configurational structure. In many cases across languages, the difference amounts to the existence of a VP node in the constituent structure, as illustrated by Figures 3.1 and 3.2.

The view of two radically different classes of languages was later refuted when it became evident that several languages possess some of the features thought to be characteristics of nonconfigurational languages. Instead, researchers in the GB paradigm attempted to formulate the 'configurationality' parameter of variation among languages (see, for instance, Marácz and Muysken 1989).

```
              S
          ╱───────╲
       SUBJ        VP
        ╱╲       ╱────╲
       ╱──╲     V      OBJ
      John     │       ╱╲
               │      ╱──╲
             loves    Mary
```
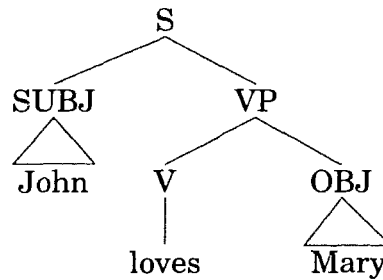
Figure 3.1: 'Configurational' (hierarchical) structure

```
             S
         ╱───┬───╲
      SUBJ   V    OBJ
       ╱╲    │    ╱╲
      ╱──╲   │   ╱──╲
      John  loves Mary
```
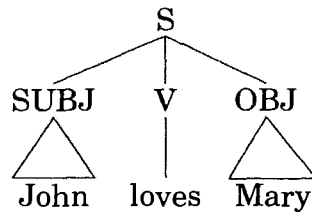
Figure 3.2: 'Nonconfigurational' (flat) structure

The criterion considered to be the most important was subject-object asymmetry—the two behave differently in some respects. This was claimed to suggest that the verb and the object form a VP node. However, as Speas (1990:162) points out, 'the claim that [the subject and the object] must be *structurally* distinguishable is a theory-internal prediction made by GB,' and it is not shared by some other theories. Furthermore, the status of the VP constituent in general is somewhat questionable. Several syntactic theories do not recognize it, for example Functional Grammar (Dik 1989), Relational Grammar (Perlmutter 1983), and Word Grammar (Hudson 1984).

In what follows, I investigate empirical (theory-independent) motivation for the existence of a VP node in Slovene. This view of not posing further constituent structure than can be motivated was expressed by Dowty (1990:34).

Since hierarchical syntactic structure is so often assumed, syntacticians don't usually ask questions—at least beyond the elementary syntax course—as to what the nature of evidence for a constituent structure in a particular

language is: we just take whatever our favorite syntactic theory would predict as the expected one for the string of words in questions—by current $\overline{X}$ theory, etc.—unless and until that assumption is contradicted by some particular fact.

I will argue for a flat structure of the Slovene clause (like in Figure 3.2, without the VP node). In chapter 4, the clause will be described within HPSG, a theoretical framework with two properties, important for the topic of this thesis:

(a) it is not confined to binary branching (as recent versions of the GB theory are); and

(b) it does not assume a homomorphism of mappings between syntactic and semantic representations, i.e., GB's Projection Principle (Radford 1988:552).

3.2 VP constituency

Phrase structure approaches regard constituency as the central manifestation of syntactic structure.[1] The evidence for the VP constituent in particular has been used as one of the prime diagnostics against the flat clause structure. The term 'VP' standardly denotes a constituent which is a projection of the verb and contains its non-subject arguments and possibly adverbials, but is distinct from the S category.

The argumentation whether some construction is a constituent mainly involves distributional and semantic criteria[2] which I now investigate in some detail.

3.2.1 Distributional evidence

If the distribution of some parts of the construction is the same as the distribution of the whole construction, then it is taken to be a constituent. 'Surface' VP constituency

---

[1]Within dependency approaches, on the other hand, constituency is not fundamentally important (Mel'čuk 1988).

[2]Below the phrase level, phonological and prosodic criteria also become important. See Miller (1992) for discussion.

tests rely on the assumption that the verb and its object behave as a unit in syntactic processes. For example, English VPs can be deleted (13a), preposed (13b) or replaced by a pro-form (13c).

(13) John said he would [read the book]

    a. and he did [].

    b. and [read the book] he did.

    c. and <u>so</u> he did.

Analogous tests do not provide us with conclusive results in Slovene. First, (14) exemplifies a deletion which is roughly comparable to (13a). It can affect the verb only (14a), but not the VP in (14b). The object clitic *ga* can not be deleted (a similar observation is made by Bolta 1989:152).

(14) a. Janez pravi, da   bo   pojedel   zajtrk   in res ga bo.

       John says COMP will eaten-PSP,MASC breakfast-ACC and indeed it-ACC will.

       'John says he will eat the breakfast and, indeed, he will.'

    b. *Janez pravi, da   bo   pojedel   zajtrk   in res bo.

       John says COMP will eaten-PSP,MASC breakfast-ACC and indeed will.

Second, VP-ellipsis is another textbook VP test—Radford (1988:83) claims that 'only VPs can undergo ellipsis (under appropriate discourse conditions).' But in Slovene, ellipsis is not limited to VPs: we can also omit just the object (15a) or just the transitive verb (15b). This too suggests that the verb and the object are not so tightly bound together as they are in English.

(15) a. Janez   knjigo   piše, Ana   pa   bere.

       John-NOM book-ACC writes Ann-NOM whereas reads

       'John is writing the book, whereas Ann is reading it.'

   b. Janez   piše   knjigo, Ana   pa   pismo.

     John-nom writes book-acc Ann-nom whereas letter-acc

     'John is writing a book whereas Ann is writing a letter.'

Next, yet another test (somewhat analogous to English VP-preposing) is provided by the fronting to the topic (clause-initial) position, as in (16). The similar pre-AUX position has also been taken as providing a test for constituency in Warlpiri (cf. Laughren 1989:329, Siewierska 1988:159).

(16) *Bral   knjigo je Janez.

     read-psp book-acc AUX John-nom

As (16) shows, finite verb-object constructions cannot be fronted.[3] If we disregard coordinate structures and parentheticals, any material that may precede the 2P cluster forms a constituent. But whenever the verb is fronted, nothing else may be fronted with it. The pre-AUX test therefore also gives a strong indication against constituency.

We can conclude that the distributional evidence does not provide enough support for positing the finite VP constituent.

3.2.2 Semantic criteria

As Miller (1992:14) notes, semantic criteria are often implicit. They represent 'the notion that syntactic scope should be isomorphic to semantic scope to the furthest extent possible' and that 'semantic function–argument structures correspond to syntactic constituents.' Miller also observes however, that this latter criterion is not reliable in the case of the VP. In the particular case of Slovene, Bolta (1989:151) shows that idiomatic expressions cannot be used to argue for a VP constituent convincingly.

Semantic issues are beyond the scope of this thesis but it is important to note that the subject/predicate dichotomy remains expressible within HPSG because HPSG does not share GB's assumption of the projection principle (for example, this principle would

---

[3]This is also the case in Serbo-Croatian (Rivero 1991:333).

imply that whenever the semantic representation is 'logically dyadic,' the syntactic representation must also have exactly two parts).

3.3 Status of subjects

There are several features in Slovene that require the subject to be distinguished from other verbal arguments: null subjects, control, and reflexivization. The properties of the subject in Slovene are somewhat different from those in English. They lead to a different clause structure which will be summarized in section 3.4.

First, Slovene is a *null-subject* (subject-drop) language (like, for instance, Russian, Italian or Spanish). When the subject of the clause can be recovered from the context, it is usually phonetically unrealized. Consider the following sentences without overt subjects.

(17) a. Bere knjigo.

reads book-ACC

'He/she is reading a book.'

b. Knjigo sem bral.

book-ACC AUX-1SG read-PSP,MASC

'I was reading a book.'

Second, Slovene belongs to the class of 'subject-antecedent' languages (Bolta 1990). Only the subject (either overt (18a) or null (18b)) can be coreferenced with a reflexive pronoun (reflexive and its antecedents are marked by subscripted indices).

(18) a. Janez je Petra povabil k sebi.

John-NOM AUX Peter-ACC invited to self-DAT

'John$_i$ invited Peter$_j$ to his$_{i/*j}$ place.'

b. Petra   je   povabil   k sebi.

Peter-ACC AUX invited-PSP,MASC to self-DAT

'He$_i$ invited Peter$_j$ to his$_{i/*j}$ place.'

Next, 'control' verbs take complements which are headed by untensed verbs (such as the infinitive in (19)) and always have controlled (unexpressed) subjects. The term *control* standardly refers to the coindexing relation between the unexpressed (understood) subject of the untensed clause and some controller NP in the matrix clause.

(19) Ana   poskuša [brati   knjigo].

Ann-NOM tries   read-INF book-ACC

'Ann is trying to read the book.'

Controlled complements form constituents (phrases) according to the pre-AUX test (20).

(20) [Brati   knjigo]   je   poskušala   Ana.

read-INF book-ACC AUX tried-PSP,FEM Ann-NOM

'Ann was trying to read the book.'

Controlled complements are nonfinite verb phrases and they too allow 'shuffling' of verbal arguments (as observed in finite clauses). Moreover, Slovene nonfinite verbs can never take overt subjects (Bolta 1986:426). There are no comparable constructions to the English one in (21) in which *John* is in some analyses considered to be the subject of the infinitival clause (Radford 1988:304).

(21) It is impossible for John to catch that bus.

Note that in HPSG the 'subject' is defined lexically rather than as a tree-configurational notion. Correspondingly, we can refer to the subject without relying on the *structural* distinction[4] (further ramifications for the HPSG analysis will be discussed in section 4.4).

---

[4]Cf. HPSG treatment of control and binding in Pollard and Sag (in press).

3.4 Flat clause structure

I now turn to formulating the structure of the clause. Borsley (1984) argued for an analysis of English where the VP's differ from S's merely in the feature ±SUBJ. He proposed that S be specified as +SUBJ and VP as –SUBJ as depicted in Figure 3.3 (number 2 indicates the bar level in his proposal).
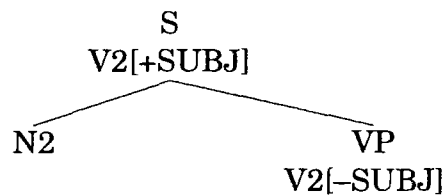
Figure 3.3: S/VP distinction in SUBJ feature

It is this observation that will be applied here—S and VP are of the same bar level and projection status. But while English VP and S need to be represented by two distinct nodes because there is independent evidence for processes that involve the VP, I argue that this is not so in Slovene.

The key note here is the lack of syntactic evidence for a finite VP node distinct from S: no syntactic process seems to refer to the finite verb-object construction (i.e., without the subject) exclusively. This suggests that the verb combines with all its arguments in one step (rather than with non-subject arguments first and with the subject in the next step).

Thus, I will maintain that all verbal arguments are immediate daughters of the verb's maximal projection in both finite and nonfinite cases. But there are some differences between the two types (see Table 3.1):

- a finite verb takes either an overt subject (thus forming what we may provisionally call S[+SUBJ]) or a null subject (forming S[–SUBJ]);

- a nonfinite verb never takes an overt subject (its projection is labelled 'VP').

label	verb	overt subject
S	finite	$+/-$
VP	nonfinite	$-$

Table 3.1: S and VP in Slovene

Clearly, once we accept that the verb and the object do not form a syntactic unit, we don't have to account for the discontinuity in cases where they fail to appear adjacently in the clause as, for instance, in the sentence (2d) repeated as (22).

(22) Pismo   Janez   piše.

letter-ACC John-NOM writes

'John is writing a letter.'

This sentence would be represented by the tree in Figure 3.4 rather than by the one in Figure 3.5 which contains a discontinuous VP and crossed branches.

S[+SUBJ]

```
         S[+SUBJ]
        /    |    \
      NP    NP     V
      |     |      |
    Pismo  Janez  piše
```
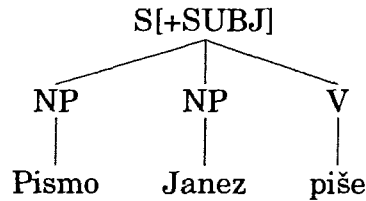
Figure 3.4: Representation without the VP

```
        S
        |  \
        NP  VP
       /|    \
     NP |     V
     |  |     |
   Pismo Janez piše
```
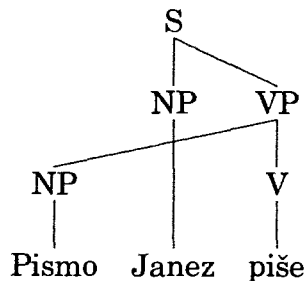
Figure 3.5: Representation with the VP

As for the clitic cluster, this thesis cannot give a comprehensive treatment. For the

particular case of Slovene, I present a purely syntactic account of the composition and the placement of the cluster in section 4.5. I contend, however, that a more general treatment (including, for example, clitic climbing which is associated with Slovene nonfinite verbs) would have to involve morphological and phonological processes.[5]

---

[5]Much of the recent work on clitics emphasizes their morpho-phonological nature by viewing the clitics as 'phrasal affixes.' See Anderson (1992) and Miller (1992).

Chapter 4

HPSG account of the data

Having surveyed the data and proposed the structure of the clause, I will next introduce
the relevant details of HPSG and then outline the model of the Slovene clause within
this framework.

4.1 Head-driven Phrase Structure Grammar

Head-driven Phrase Structure Grammar is an integrated theory of natural language
syntax and semantics (Pollard and Sag 1987, in press). The theory is *nonderivational*
in that it does not employ transformational operations or movements between distinct
levels of syntactic structures. Rather, the attributes of linguistic structures are related
by structure sharing, i.e., 'token identity between substructures of a given structure in
accordance with lexical specifications or grammatical principles' (Pollard and Sag in
press, p. 2). The theory is *declarative* in the sense that an interaction of universal and
language-specific constraints on linguistic expressions (words and phrases) determines
their well-formedness (admissibility) in a particular language, independently of the
order that they are applied in.

Linguistic entities are modelled by *sorted feature structures* (FSs). The sort of the FS
indicates what kind of object it is modelling. The finite set of sort symbols is assumed

to be partially ordered. Informally, FSs are record-like collections of attribute-value pairs. What attributes (features) are appropriate for some FS is determined by its sort. The value of an attribute may be an embedded FS (possibly atomic, i.e., with no appropriate attributes), a list, or set of FSs. Substructures in a given FS may be shared, that is, two attributes may have one and the same object as their value. HPSG employs *unification* as the major information-combining operation between FSs. A formalized underlying 'feature logic' which lays foundations for the computational implementation of the theory is set forth by Carpenter (1992).

4.2 Signs

Any *sign* (roughly, the information associated with a word or a phrase) that conforms to all of the relevant principles constitutes a prediction of the theory. In a sign, information is organized in distinct levels. These are also called top-level attributes and include the sign's phonology (PHON) and syntax-semantics (SYNSEM). Top-level attributes are internally structured, bringing together (under the same attribute) information that forms a natural class. For example, a fundamental assumption of HPSG is that some signs act as *heads* and share certain features with their phrasal projections. These features are subsumed under the SYNSEM | LOC | CAT | HEAD attribute (such a sequence of attributes is termed a *path*[1] and is used to refer to embedded attributes).

In the lexicon, heads are specified with their *subcategorization* information, that is, what constituents they have to combine with (subcategorize for). For example, the verb *sees* bears a requirement that it has to combine with a nominative subject and accusative object to form a sentence. A list-valued[2] feature SUBCAT is used as a repository of such requirements, as shown in (23).

(23) SYNSEM | LOC | CAT | SUBCAT ⟨NP[*nom*], NP[*acc*]⟩

---

[1]Throughout this chapter I often use the abbreviated feature names instead of full paths, e.g., HEAD stands for SYNSEM | LOC | CAT | HEAD.

[2]Angle brackets indicate that the value of an attribute is a list of objects.

Note that in the HPSG context, labels such as NP[nom] or VP[finite] serve as abbreviations for FSs (in this case they stand for *synsem* objects). Objects with empty SUBCAT lists are called *saturated*, while the others are called *unsaturated*.

The presence of subcategorization information in the lexicon dramatically reduces the number of required *immediate dominance* (ID) schemata which are roughly analogous to traditional phrase structure rules, e.g., S -> NP VP. Only a few rule schemata are employed to specify (in conjunction with universal principles) the immediate constituency, that is, how a 'mother' sign is composed from 'daughter' signs. Phrase structure is manifested by another top-level attribute that all phrasal signs are assumed to have, the DTRS (daughters) attribute. There are various kinds of daughters: heads, complements, adjuncts, fillers etc. In this thesis, we will only be concerned with head, subject, and complement daughters.

ID schemata do not impose the surface order in which the daughter categories have to be realized in the utterance. This is accomplished by separate *linear precedence* (LP) constraints which will be discussed in section 4.5.

FSs are often graphically depicted as *attribute-value matrix* (AVM) diagrams. The attribute-value matrix in Figure 4.1 illustrates the architecture of a phrasal sign (with some attributes omitted for brevity[3]). Sorts of objects are denoted by left subscripts.

In the representation of a sign of sort *phrase* in Figure 4.1, PHON contains the orthographic representation of the phonetic content of the phrase, and DTRS contains its immediate constituency. In this case, the DTRS value includes HEAD-DTR (head daughter) and list-valued COMP-DTRS (complement daughters) as its appropriate features.

The LOC (local) feature of the *synsem* object shows its category (CAT) which roughly includes the part of speech and subcategorization information, along with the semantic content (CONT), which is, in this simplified case, just the referential index.

---

[3]For this reason, FSs will not be shown in full in the figures hereafter. For example, just the SYNSEM | LOC | CAT values may be shown instead of full *synsem* objects.

$$
\text{phrase} \begin{bmatrix} \text{PHON} & \langle \ldots \rangle \\[4pt] \text{SYNSEM} & \text{synsem} \begin{bmatrix} \text{LOC} & \text{local} \begin{bmatrix} \text{CAT} & \text{cat} \begin{bmatrix} \text{HEAD} & \text{head} \\ \text{SUBCAT} & \langle \ldots \rangle \end{bmatrix} \\[4pt] \text{CONT} & \text{parameter} \begin{bmatrix} \text{INDEX} & \text{index} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\[4pt] \text{DTRS} & \text{head-comp-struc} \begin{bmatrix} \text{HEAD-DTR} & \text{sign} \\ \text{COMP-DTRS} & \langle \ldots \rangle \end{bmatrix} \end{bmatrix}
$$
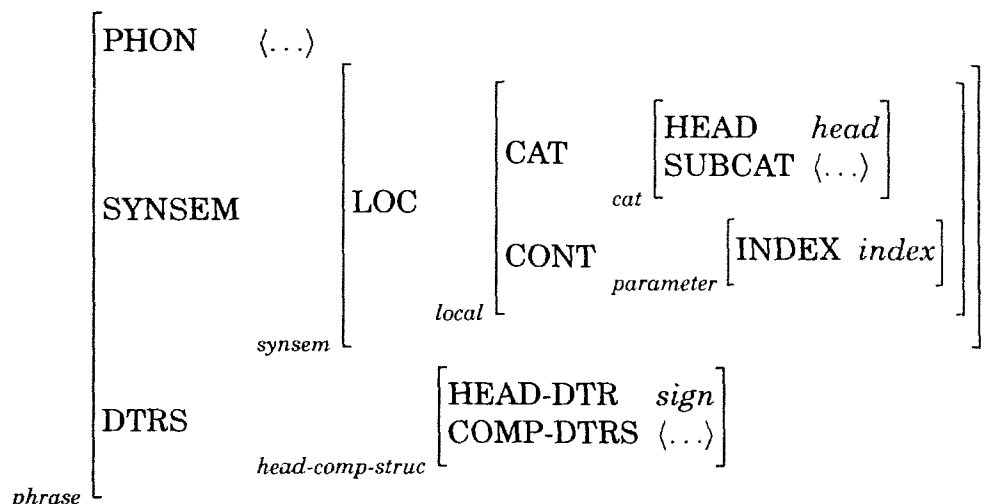
Figure 4.1: Architecture of a phrasal sign

4.3 Lexicon

As mentioned above, HPSG relies crucially on rich lexical information. Within such a lexicon, much of the information is shared by many lexical entries. However, this doesn't mean that this complex information has to be redundantly stipulated in each individual lexical entry. Rather, properties of lexical entries and relationships among them are expressed concisely in terms of the lexical type hierarchy and lexical rules (Pollard and Sag 1987, ch. 8). Using these two mechanisms we can capture linguistic generalizations and avoid redundancy in the lexicon.

Properties shared by a class of words are represented by a 'generic' lexical entry or *lexical type*. A lexical type contains certain attribute values and constraints that hold for all members of that class, so that this common information is stated in a single place once and for all. Lexical types are organized in an inheritance hierarchy. Informally, a class in this hierarchy inherits the properties of all its superclasses and, additionally, it may introduce some new restrictions or information. Consider a fragment of the classification of English verbs according to verb form in Figure 4.2 (Pollard and Sag 1987:202).
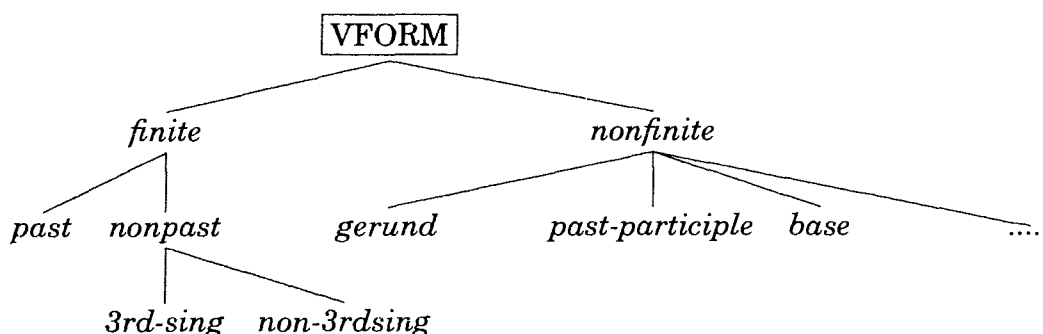
Figure 4.2: A fragment of the classification of English verbs

Words are thus cross-classified on the basis of various distinctions, such as part of speech, verb form, valence etc. For example, we can state that the verb *walks* belongs to classes *intransitive* and *3rd-sing* (see appendix B for some attributes that would have to be encoded in the Slovene lexicon). The amount of idiosyncratic information of individual lexical entries (which can not be predicted from their membership in word classes) is in this way significantly reduced.

Once the information of a general nature has been factored out by a hierarchy of lexical types, there still remains the so-called 'horizontal redundancy.' The lexicon contains groups of words related according to recurring patterns (e.g., inflectional paradigms such as verb conjugation, derivational paradigms such as nominalizations, etc.). HPSG accounts for these relationships by *lexical rules*. These may be interpreted as 'inference rules that derive lexical entries of inflected, derived or compound words from those of simpler words' (Pollard and Sag in press, ch. 1). For instance, we may view the word *cats* as the output of the rule that derives the plural form from a singular noun.

HPSG has another important feature, which is pointed out in the introduction of Pollard and Sag (in press):

> Perhaps most characteristically, in HPSG tree-configurational notions such as c-command and government are not regarded as linguistically significant; instead their role is taken over by the relation of relative obliqueness.

Obliqueness is an ordering of grammatical relations shown in (24).

(24) SUBJECT ≺ PRIMARY OBJECT ≺ SECONDARY OBJECT ≺ OTHER COMPLEMENTS

Obliqueness ordering is motivated cross-linguistically, Sag and Pollard (1989:166) claim, by several classes of generalizations: constituent order, binding, control, and the functioning of lexical rules. Obliqueness is encoded in the SUBCAT list, so that less oblique arguments precede more oblique ones. The subject (nominative) NP precedes the primary object NP in the SUBCAT list of the word *sees* in Figure 4.3.

$$
\begin{bmatrix}
\text{PHON} & \textit{sees} \\[2ex]
\text{SYNSEM} \mid \text{LOC} \mid \text{CAT} &
\begin{bmatrix}
\text{HEAD} & \textit{verb}[\textit{finite}] \\
\text{SUBCAT} & \langle \text{NP}[\textit{nom}], \text{NP}[\textit{acc}] \rangle \\
\text{LEX} & +
\end{bmatrix}_{\textit{cat}}
\end{bmatrix}_{\textit{word}}
$$
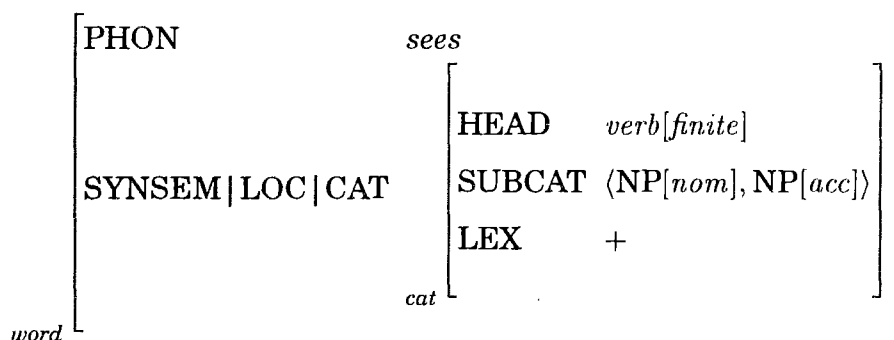
Figure 4.3: Lexical entry for *sees*

In Figure 4.3, notice also the binary[4] feature LEX which is used to distinguish between lexical and nonlexical (phrasal) signs.

4.4 Clause level

I will now sketch a corresponding HPSG analysis of the flat structure that was proposed in chapter 3.

---

[4]We often specify binary-valued features by prefixed + or −, e.g., −LEX instead of [LEX +].

4.4.1 Subject and complements

As established in section 3.3, there is ample evidence for distinguishing the subject from other verb arguments. I will assume the revised, more flexible approach of Pollard and Sag (in press, ch. 9), first proposed by Borsley (1987). Two valence features will be employed: subjects are selected via the (singleton) SUBJ list, and complements by the COMPS list. In addition to the feature SUBCAT $\langle \boxed{1}, \boxed{2}, ... \rangle$, we now obtain features SUBJ $\langle \boxed{1} \rangle$ and COMPS $\langle \boxed{2}, ... \rangle$. The revised approach uses the SUBCAT feature (as the concatenation of SUBJ and COMPS) only for binding (which is not discussed in this thesis).

While the three 'subjecthood' properties (subject-drop, reflexivization, control) generally refer to the same argument, in some exceptional cases they do not coincide. For example, some rare Slovene verbs (e.g., *zebsti* 'to be cold,' 'to feel cold'[5]) can in finite clauses appear in 3rd person, singular, neuter form only. These verbs take a single overt argument which:

- is not in the nominative case (which is the default for subjects);

- cannot be dropped when the verb is finite (25a–25b);

- is not unexpressed when the verb is nonfinite and heads a controlled complement (25c–25d);

- can be the antecedent of a reflexive (25e).

(25) a. Zebe me.

  Feels cold-3SG,PRES me-ACC

  'I'm cold.'

 b. *Zebe.

  Feels cold-3SG,PRES

---

[5]There is no exact analog in English.

c. Zebsti   me   je   začelo.

feel cold-INF me-ACC AUX-3SG started-PSP,3SG,NEU.

'I started to feel cold.'

d. *Zebsti   je   začelo.

feel cold-INF AUX-3SG started-PSP,3SG,NEU

e. Zebe   ga   v svoji   sobi.

Feels cold-3SG,PRES him-ACC in his-POSS-REFL room.

'He$_i$ feels cold in his$_i$ room.'

Such verbs are similar to meteorological verbs that always take an expletive subject ('it', 'there') in English (e.g., 'to rain')—the difference being that Slovene (like other related Slavic languages) has no expletive pronouns.[6]

HPSG distinguishes between referential and nonreferential (expletive) NPs by the sort of the SYNSEM | LOC | CONT | INDEX value (Pollard and Sag in press, ch. 3). I will assume that the sort *index* in Slovene subsumes subsorts *ref* and *expl*. We can therefore account for these exceptional verbs in the following way:

- the INDEX value of its subject is of sort *expl*;

- the verb's overt argument is a complement but is still least oblique for the purpose of binding of reflexives.

Note that the sort *expl* also encompasses agreement restrictions[7] (3rd person, singular, neuter), thus ruling out other forms (e.g., 1st person) of such verbs (see Pollard and Sag (in press, ch. 2) for the treatment of agreement in HPSG). Expletive subjects will never be overt in Slovene: by virtue of the fact that the lexicon will contain no nominative expletive pronouns (see also section 4.4.3).

---

[6]Empty expletives are attested in other languages as well; for instance, in Italian (Haegeman 1991:303).

[7]The same agreement restrictions are borne by overt expletives in some languages, for instance, 'es' in German.

The SUBJ list of an 'ordinary' verb, on the other hand, contains a referential NP (INDEX value of sort *ref*) as its element for which the following holds (cf. Table 3.1):

(i) it may be dropped in the finite case; and

(ii) it must not be overt in the nonfinite case.

4.4.2 Main verbs and auxiliaries

If the existence of VPs is assumed, it follows directly in HPSG that auxiliaries subcategorize for VPs, and this standard treatment is also adopted by Mežnarič (1992).

As mentioned in section 2.3, the auxiliary and the object clitics have to be ordered with respect to each other (that is, if the objects are realized as pronominal clitics). Mežnarič achieves the locality of ordering rules by a variant of 'argument composition.' In this technique which was in the HPSG context first used by Hinrichs and Nakazawa (1990), arguments of the (embedded) main verb are 'attracted' to the auxiliary. Because of the ordering freedom in Slovene, we would have to 'raise' all complements of the main verb to the subcategorization frame of the auxiliary, as shown in Figure 4.4.

$$
\begin{bmatrix}
\text{HEAD} & aux \\
\text{SUBJ} & \langle \text{NP}[nom] \rangle \\
\text{COMPS} & \langle \text{VP}[\text{HEAD } mainv, \text{COMPS } \boxed{1}\,] \oplus \boxed{1} \rangle
\end{bmatrix}_{category}
$$

Figure 4.4: Argument composition of subcategorization requirements

Note that the number tag in Figure 4.4 represents the entire COMPS list (i.e., all elements) of the embedded verb, and that the sign $\oplus$ denotes the 'append' (concatenation) operation. Therefore, even in this case we effectively end up with a flat clause (the auxiliary, the main verb and its arguments are sisters).

But in the structure without the VP node, the above treatment has no obvious advantage over the alternative in which the main verb subcategorizes for the auxiliary.

In this alternative approach we don't have to resort to argument composition, and I now present some further motivation for it.

First, note that the relation between the auxiliary and the participle is a morphological[8] rather than a syntactic one (Pullum and Wilson 1977). Slovene auxiliaries serve as tense, mood and agreement markers. They resemble inflectional affixes of verbs, as demonstrated by the following example from closely related Serbo-Croatian. Future tense can be realized either by a suffix (26a) or by an (independent) 2P auxiliary *ću* in (26b).

(26) a. Pisaću.

     write-1SG,FUT

     'I will write.'

   b. Ja   ću   pisati.

     I-NOM will-1SG,FUT,CL write-INF

     'I will write.'

Second, it is not at all clear that the notion 'head' can be extended to morphology (Anderson 1992:319). But if the established criteria are applied, they give no overwhelming evidence for the headship of either the base (in this case, the participle) or the inflectional suffix (Bauer 1990:18ff). Yet if the participle is chosen as the head[9] in this case, the construction is simpler than it would be the other way around: we can now propose generalization that the clause is always headed by the main verb. Furthermore, as will be shown below, a single schema will suffice for both simple and compound tenses (and for both finite and nonfinite heads).

Third, the main role of the participle is also supported by the evidence from Czech past tense (P. Kuboň, personal communication): the participle must cooccur with the

---

[8]In this regard, the particular technical realization that I propose here is necessitated by the lack of a morphological component in the current version of HPSG.

[9]The central role of the main verb is also reminiscent of dependency grammar approaches (Mel'čuk 1988).

auxiliary in the 1st and 2nd person, whereas in the 3rd person the same participle has to appear alone (there are no 3rd person auxiliaries). Therefore, if the auxiliary were the head, we would have to assume a null head in the 3rd person which too is undesirable.

Next, consider 'auxiliary strings' of the kind shown in Figure 4.5 (Pollard and Sag 1987:125). Constituency criteria suggest a nested structure of such strings in English
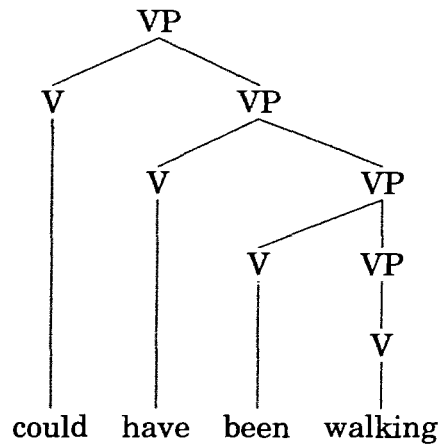
Figure 4.5: Nested structure of the English auxiliary string

whereby the head auxiliary takes a VP as its complement. But this argument does not apply in Slovene in which such auxiliary strings are not attested (in Slovene, modal verbs are not auxiliaries, and there are no perfective or progressive auxiliaries at all).

Finally, there is some motivation for a uniform treatment of pronominal clitics and auxiliaries. They share the following properties:

- invariably, they are elements of the 2P cluster;

- they can not take complements nor can they be modified or conjoined;

- they function exclusively as complements to the main verb (e.g., if a preposition takes a pronoun as its object, then the pronoun must be non-clitic).

Technically, the participle will act as the head taking the auxiliary as its most

oblique complement.[10] Assuming the partition of the sort *verb* into subsorts *mainv* (for main verbs) and *aux* (for auxiliaries), the participle of the intransitive verb can be illustrated by the AVM diagram in Figure 4.6.

$$
cat \begin{bmatrix} \text{HEAD} & mainv[\text{VFORM} \; \boxed{1}] \\ \text{SUBJ} & \langle \boxed{2}\text{NP}[nom] \rangle \\ \text{COMPS} & \langle aux[\text{SUBJ} \; \langle \boxed{2} \rangle , \text{VFORM} \; \boxed{1}] \rangle \end{bmatrix}
$$

Figure 4.6: AVM diagram of the intransitive past participle

The agreement among the participle, the auxiliary, and the subject is obtained by structure-sharing which is notated by pairs of number tags in Figure 4.6. The SUBJ and VFORM values are thus shared between the participle and the auxiliary. The VFORM value encodes the tense information whereas the subject's SYNSEM | LOC | CONT | INDEX value contains the agreement properties. VFORM, INDEX | PER (person) and INDEX | NUM (number) are lexically specified on the auxiliary and INDEX | GEND (gender) on the participle. The VFORM feature of the participle in Slovene is lexically unspecified which (intuitively) makes sense, since it is the same participle that can be combined with different forms of the auxiliary (e.g., present or future).

In this way, the main verb acquires from the auxiliary information that has to be percolated from the verb to the clause node (i.e., properties that may be selected, such as tense, mood etc.).

4.4.3 Immediate dominance schema

Immediate constituency of phrases is in HPSG constrained by a universal principle (27) specifying a disjunction of available ID schemata, from among which each language makes a selection (Pollard and Sag in press, ch. 1).

---

[10]The exact location in the COMPS list is not crucial, since the auxiliary does not interfere with principles that depend on obliqueness (e.g., binding).

(27) Immediate Dominance Principle

The universally available options for a well-formed phrase are: [Schema 1 or Schema 2 or ... etc.]

To account for the flat clause, I use a variant of Schema 3 (Pollard and Sag in press, ch. 9) in (28) which combines a head daughter with a subject daughter and (possibly several) complement daughters. An instance of the application of this schema is depicted in Figure 4.7 (as elaborated below, the subject in this figure is optional).

$$
cat\begin{bmatrix} \text{HEAD} & mainv \\ \text{SUBJ} & \langle\rangle \\ \text{COMPS} & \langle\rangle \\ \text{LEX} & - \end{bmatrix}
$$

S   H   C   C

$$
\boxed{1}
$$

$$
cat\begin{bmatrix} \text{HEAD} & mainv \\ \text{SUBJ} & \langle\boxed{1}\rangle \\ \text{COMPS} & \langle\boxed{2}\ldots\boxed{n}\rangle \\ \text{LEX} & + \end{bmatrix}
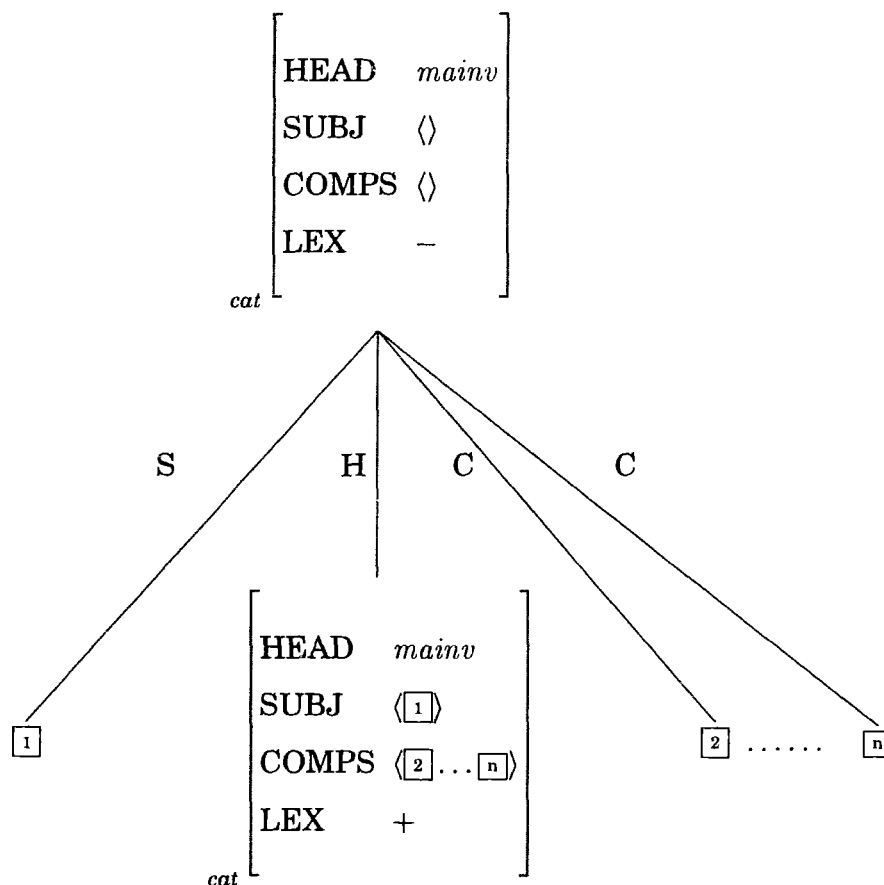$$

$$
\boxed{2} \ldots \ldots \boxed{n}
$$

Figure 4.7: A structure licensed by Schema 3

(28) Schema 3

> ...A phrase with DTRS value of sort *head-subj-comp-struc* in which the head daughter is a lexical sign with the HEAD value of sort *mainv*.

Some comments are in order here. First, the schema includes a parochial condition that the head be a main verb (and not an auxiliary). Following Pollard and Sag (in press, ch. 9), I assume that the sort *phrase* incorporates [COMPS $\langle\rangle$] specification. In the figures below, the label S will abbreviate *phrase*[−LEX, HEAD *mainv*].

Second, I further assume that this is the only schema that combines a verbal head with its complements. Since the schema licenses a lexical head daughter and a nonlexical mother, a verb can have at most one projection. Therefore, all of the complements (if any) must combine with the head verb at the same time. Both finite and nonfinite verbs are treated by this schema.

Third, the schema makes no reference to the SUBJ attribute and the HEAD–SUBJ schema (which combines a VP and a subject in English) is not employed for Slovene. The Valence Principle ensures that the mother's SUBJ list is empty if and only if a subject has been combined with the head verb (by this schema). When the subject of the finite clause is phonetically unrealized (null), then the mother will contain a non-empty SUBJ value.[11] Figure 4.8 shows the construction of the sentence (17a) by Schema 3 in which the mother sign retains a non-empty SUBJ list. Recall that the COMPS list on the mother, on the other hand, is always empty due to the constraints on the sort *phrase*.

Next, I will assume that determining when the subject can be overt (see Table 3.1) is the task of an additional language-specific principle (29) which is associated with Schema 3 (since this is the only schema involving a subject daughter).

---

[11]As Jo Calder (p.c.) points out to me, other treatments for null subjects are possible. We could also have a lexical rule that would remove the subject from the subcategorization requirements of finite verbs.

$$S[\text{SUBJ } \langle NP[nom]\rangle]$$

H   C

mainv[SUBJ $\langle NP[nom]\rangle$,   NP[*acc*]
COMPS $\langle NP[acc]\rangle$]

Bere   knjigo.
reads-3SG   book-ACC
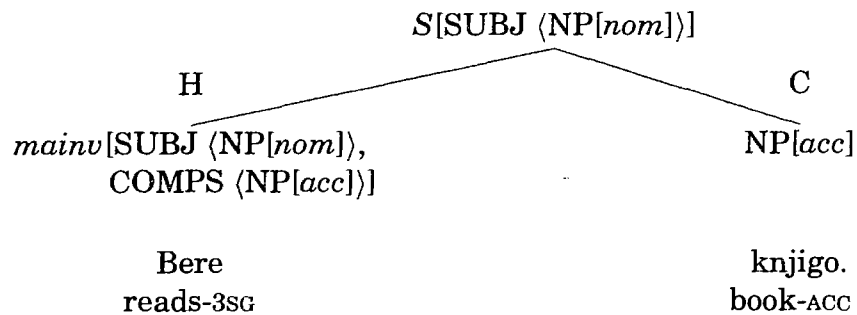
Figure 4.8: Sentence with a null subject

(29) Subject Realization Principle (language-specific)

    a. If the head verb is nonfinite, then the subject cannot be overt. The constraint will be imposed that the SUBJ list of the mother must be non-empty (unexpressed subject).

    b. If the head verb is finite, then

        (i) if its subject is specified as expletive (see section 4.4.1), then the mother's SUBJ list may be non-empty;

       (ii) otherwise it depends on the context whether or not a null (referential) subject is allowed (i.e., a non-empty SUBJ list of the mother).

However, deciding exactly in what circumstances the (referential) subject of a finite clause can be recovered from the context is a hard problem (somewhat similar to anaphora resolution). Obviously, we would require representation of the previous discourse and an appropriate decision algorithm. I will not have anything more to say about this issue here. What is important to note here is that the unrealized subject (of any kind) is not a part of the constituent structure (i.e., DTRS). This treatment of the unrealized subject in Slovene in consistent with the one proposed for the unexpressed subject in English (Pollard and Sag in press, ch. 3). There is no empty constituent for the unexpressed subject—it is only present in the subcategorization frame.

Finally, the controlled (unexpressed) subject is a consequence of coindexing which takes place in control verbs. Recall the sentence (19) repeated here as (30).

(30) Ana   poskuša [brati   knjigo].

     Ann-NOM tries   read-INF book-ACC

     'Ann is trying to read the book.'

The SYNSEM | LOC | CAT value of the control verb *poskuša* 'tries' is depicted in Figure 4.9.

$$
\begin{bmatrix}
\text{HEAD} & mainv \\
\text{SUBJ} & \langle \text{NP}[nom, \text{INDEX } \boxed{1}] \rangle \\
\text{COMPS} & \left\langle
\begin{bmatrix}
\text{HEAD} & mainv[nonfinite] \\
\text{SUBJ} & \langle \text{NP}[\text{INDEX } \boxed{1}] \rangle \\
\text{COMPS} & \langle \rangle \\
\text{LEX} & -
\end{bmatrix}_{cat}
\right\rangle
\end{bmatrix}_{cat}
$$
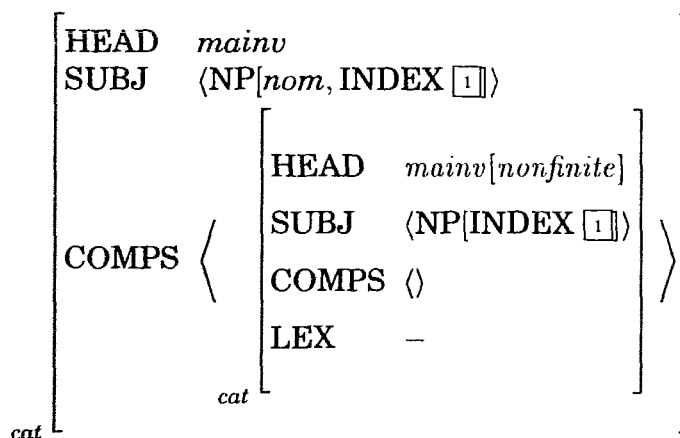
Figure 4.9: Control verb

The control verb subcategorizes for a nonfinite verb phrase complement which is required to have a single noun phrase in its SUBJ list (just the category part of the *synsem* object is shown). The referential indices of the embedded verb's subject and the control verb's subject are structure-shared: the two subjects are coindexed.[12] Figure 4.10 shows the construction of the sentence (30) in which Schema 3 is applied twice.

4.5 Linear precedence constraints

The following standard restrictions on the rules of constituent order are listed in Siewierska (1988:244):

---

[12]Note that this structure-sharing also correctly predicts that the controller of an expletive subject must also be expletive (i.e., non-overt in Slovene) in (25c).

$$
\text{S}[\textit{fin}, \text{SUBJ} \langle \rangle]
$$

S   H   C

$$
\boxed{1}\text{NP}[\textit{nom}]
\qquad
\text{cat}\begin{bmatrix} \text{HEAD} & \textit{mainv}[\textit{fin}] \\ \text{SUBJ} & \langle \boxed{1}\text{NP}[\text{INDEX } \boxed{5}] \rangle \\ \text{COMPS} & \langle \boxed{2} \rangle \end{bmatrix}
\qquad
\boxed{2}\ \text{cat}\begin{bmatrix} \text{HEAD} & \textit{mainv}[\textit{nonfin}] \\ \text{SUBJ} & \langle \boxed{4}\text{INDEX } \boxed{5} \rangle \\ \text{COMPS} & \langle \rangle \end{bmatrix}
$$

H   C

$$
\text{cat}\begin{bmatrix} \text{HEAD} & \textit{mainv}[\textit{nonfin}] \\ \text{SUBJ} & \langle \boxed{4}\text{NP} \rangle \\ \text{COMPS} & \langle \boxed{3} \rangle \end{bmatrix}
\qquad
\boxed{3}\text{NP}[\textit{acc}]
$$

Ana   poskuša   brati   knjigo.
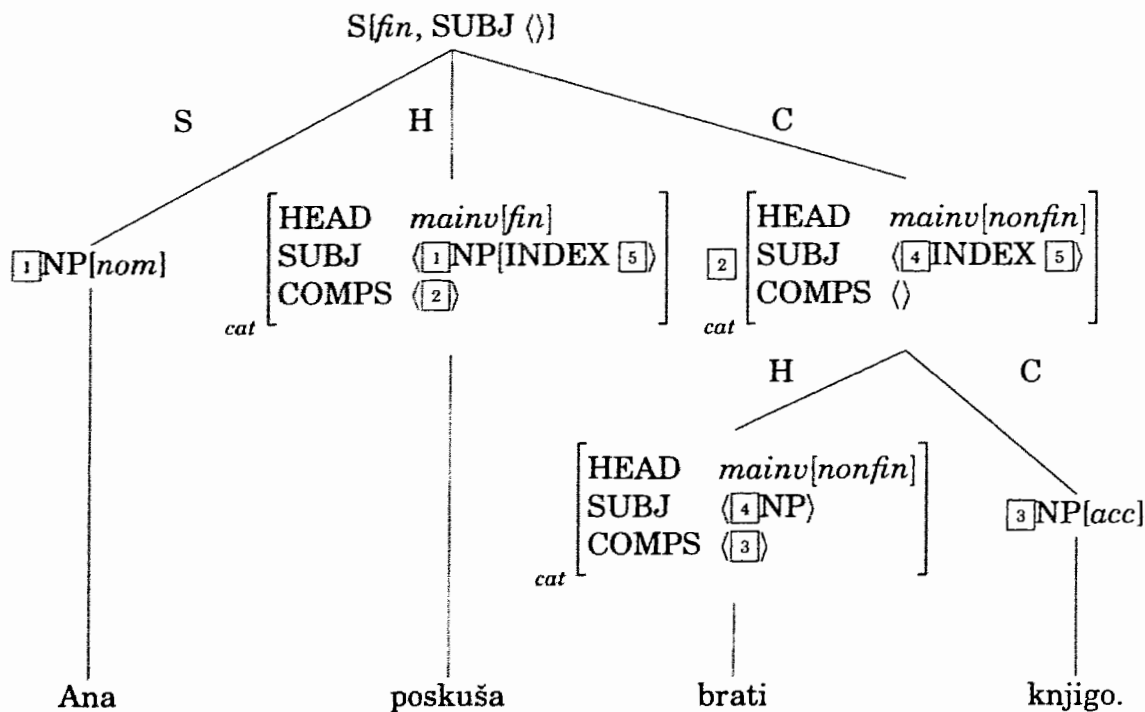
Figure 4.10: Sentence with a controlled complement

- they apply only to two sister categories;

- initial, second and final are the only constituent locations recognized;

- they may specify precedence, but not adjacency;

- information concerning the internal structure of phrases is not allowed.

All of them will be observed below. I am assuming the standard definition of a language-specific Constituent Ordering Principle, as formulated by Pollard and Sag (1987:169).

(31) Constituent Ordering Principle

$$
\textit{phrasal-sign}\begin{bmatrix} \ \end{bmatrix} \Rightarrow \begin{bmatrix} \text{PHON} & \textit{order-constituents}(\boxed{1}) \\ \text{DTRS} & \boxed{1} \end{bmatrix}
$$

This principle includes a functional constraint *order-constituents* which is satisfied by a disjunction of permutations of daughter constituents consistent with all of the LP constraints. All grammatical permutations will be licensed regardless of their degrees of acceptability (not just the unmarked ones).

The only 'syntactic' requirement in main declarative clauses is the second position of clitics, after the constituent or lexical item in the clause-initial (topic) position. This requirement is formulated by Mežnarič (1992:9) as follows.

(32) a. Something in the sentence must be topicalized and it cannot be the clitic.

b. Clitics always follow the topicalized constituent.

I will employ two binary features: TOPIC and 2P-CL. Provisionally, I am assuming that these features belong to objects of sort *category*, i.e. they abbreviate SYNSEM | LOC | CAT | TOPIC and SYNSEM | LOC | CAT | 2P-CL.

The purpose of the TOPIC feature is, much like the TOP feature of Uszkoreit (1987:114) and the TOPIC feature of Mežnarič (1992), to identify the topicalized element in the clause. The 2P-CL[13] feature (appropriate for words only), on the other hand, distinguishes between 2P clitics and other lexical signs (this could also be accomplished by the sortal hierarchy).

Assuming that clitics are lexically specified as [+2P-CL, −TOPIC] and that all other signs are specified as [−2P-CL] and unspecified for the TOPIC feature, (32) can be expressed by the following two rules (other rules would be needed to restrict internal order of the cluster which is ignored here).

(33) Linear Precedence Rules (language-specific)

LP-1. [+TOPIC] < [ ]

LP-2. [+2P-CL] < [−TOPIC, −2P-CL]

---

[13]I am using a more precise name 2P-CL rather than CL of Mežnarič (1992) since there are other clitics which are not elements of the 2P cluster.

Note that the present TOPIC feature differs from the one employed by Mežnarič. In her version, it is a head feature and the topicalized element is licensed by a Filler-Head rule (Mežnarič 1992:39). She treats this fronting as a long-distance dependency which is introduced by a phonetically null trace (gap) and mediated by the (non-local) SLASH feature (but she gives no further details on where in the clause the gap originates).

In this analysis, on the other hand, TOPIC is not a head feature. Its value is instantiated by the Constituent Ordering Principle, according to the rules in (33). Their domain of locality is the entire clause (since the verb and its arguments are sisters) and the topicalized element is enforced as a consequence of LP constraints. Non-local features are not used here and no gap is assumed in the clause.

Recall also that the NP object of the verb *bere* 'reads' is in Figure 4.8 simply specified as NP[acc]. If the object is realized as a clitic, it will appear in the second position by virtue of linear precedence constraints—there is no need for a separate mechanism (for instance, a new CLTS feature which is employed by Monachesi (1993) for Italian object clitics).

It can be observed that the rules in (33) only impose order on the clause level, since the 2P elements lack phrasal projections and can only be subcategorized for by the main verb. Thus it would be desirable to express this fact explicitly and restrict their domain. This could be achieved in the sortal hierarchy (so that the TOPIC feature would be appropriate only for those signs which represent constituents that may appear in the clause-initial position) but this matter needs to be further investigated.

To rule out pragmatically unacceptable permutations, we would require an appropriate representation of the previous discourse. A fine grained interaction of discourse and stylistics subtleties will be left for further research, perhaps along the lines of Uszkoreit (1986), by incorporating additional pragmatic and stylistic rules of various weights. Markedness would then be determined according to the total weight of violated rules.

4.6 Analysis

Let us now look at the analysis step by step. Apart from Schema 3 (28) we will refer to the Valence Principle in (34).

(34) Valence Principle

In a headed phrase, for each valence feature F, the F value of a head daughter is the concatenation of the phrase's F value with the list of SYNSEM values of the F-DTRS value.

Informally, the Valence Principle 'checks off' the realized subcategorization require-ments. Valence features on which this principle operates include attributes SUBJ and COMPS. Consider again the sentence (3a) repeated here as (35).

(35) Janez   je   bral   knjigo.

John-NOM AUX-3SG read-SG,PSP,MASC book-ACC

'John read a book.'

Lexical entries that correspond to (35) are shown in Figure 4.11 (simplified for exposi-tory purposes).

First, let us examine the immediate dominance part of the construction of the sentence. Since the verb *bral* is a lexical sign with the HEAD value of sort *mainv*, it can serve as the head daughter in Schema 3 (28). This ID schema will license a phrase of sort *head-subj-comp-struc* (see Figure 4.7). The Valence Principle (34) ensures that all subcategorization requirements of the head verb are met. The nominative NP[14] *Janez* is assigned the subject role, while the accusative NP and the auxiliary serve as primary and secondary object, respectively.

Consider now the LP rules in (33). They constrain the surface order of daughter signs of the phrase which is expressed in the PHON list. Rule LP-1 will ensure that just one of the sister signs bears the [+TOPIC] specification, the one representing the

---

[14]Bare nouns are also NPs (there are no obligatory determiners in Slovene).

$$
\begin{bmatrix}
\text{PHON} & Janez \\
\text{CAT} & \begin{bmatrix} \text{HEAD} & noun[nom] \\ \text{SUBJ} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix}_{cat}
\end{bmatrix}_{word}
\qquad
\begin{bmatrix}
\text{PHON} & je \\
\text{CAT} & \begin{bmatrix} \text{HEAD} & aux \\ \text{SUBJ} & \langle synsem \rangle \\ \text{COMPS} & \langle\rangle \\ \text{2P-CL} & + \end{bmatrix}_{cat}
\end{bmatrix}_{word}
$$

$$
\begin{bmatrix}
\text{PHON} & bral \\
\text{CAT} & \begin{bmatrix} \text{HEAD} & mainv \\ \text{SUBJ} & \langle \boxed{1}NP[nom] \rangle \\ \text{COMPS} & \langle NP[acc], aux[\text{SUBJ } \langle \boxed{1} \rangle] \rangle \\ \text{LEX} & + \end{bmatrix}_{cat}
\end{bmatrix}_{word}
\qquad
\begin{bmatrix}
\text{PHON} & knjigo \\
\text{CAT} & \begin{bmatrix} \text{HEAD} & noun[acc] \\ \text{SUBJ} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix}_{cat}
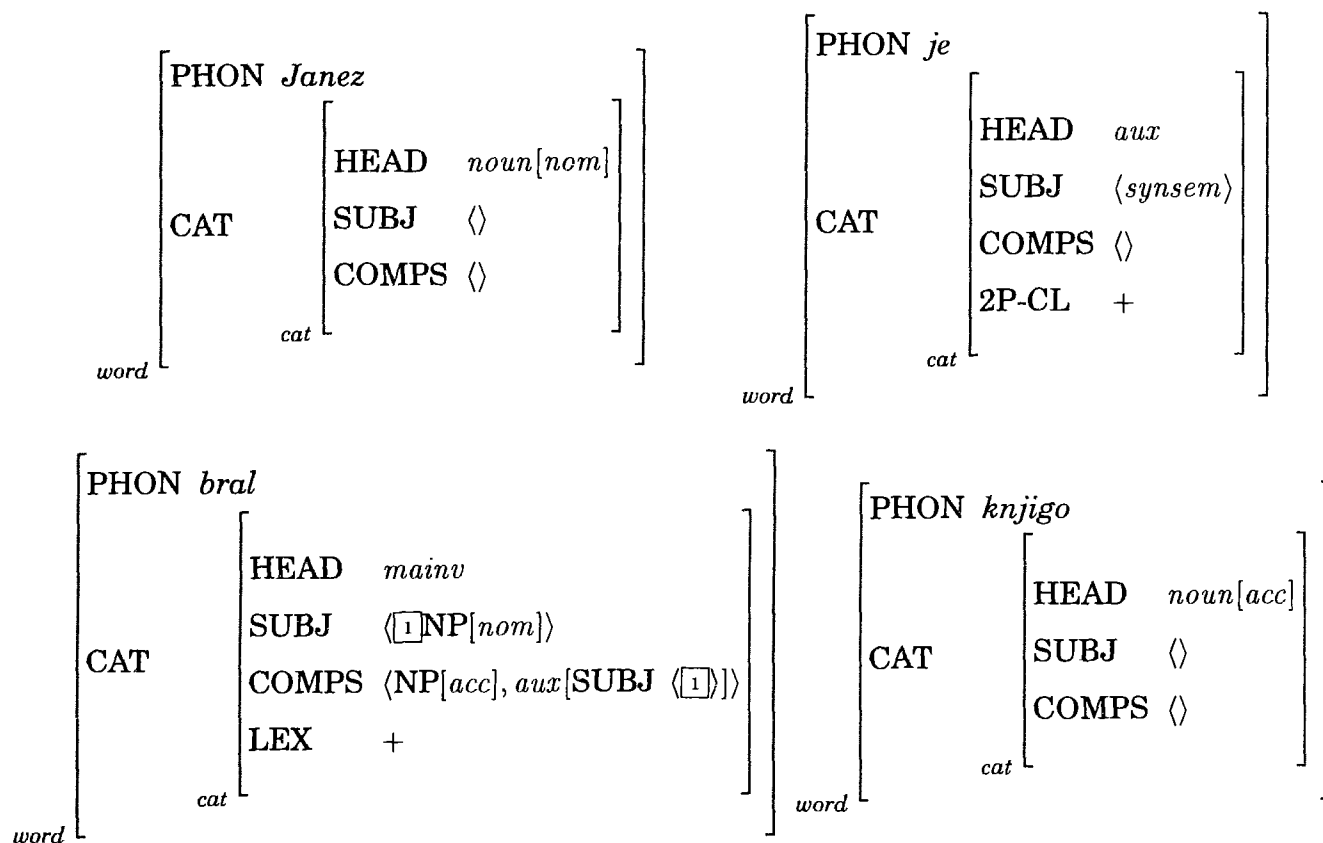\end{bmatrix}_{word}
$$

Figure 4.11: Sample lexical entries

constituent that precedes all the others in the surface string. In this case, the topic position is occupied by the subject *Janez*.

Rule LP-2, on the other hand, requires that any clitic must precede any other non-clitic (except for the topic constituent). Since the auxiliary is marked [+2P-CL] and there are no other 2P elements in (35), it must be positioned immediately after the subject in the topic position.

The rest of the elements are not constrained by LP rules. Again, note that this is just one of the admissible orderings. Any ordering of (35) that has the auxiliary in the second position is accepted. In this way, we obtain the 'mother sign' (shown as a FS in Figure 4.12 and as a tree in Figure 4.13) that represents the entire clause: a saturated

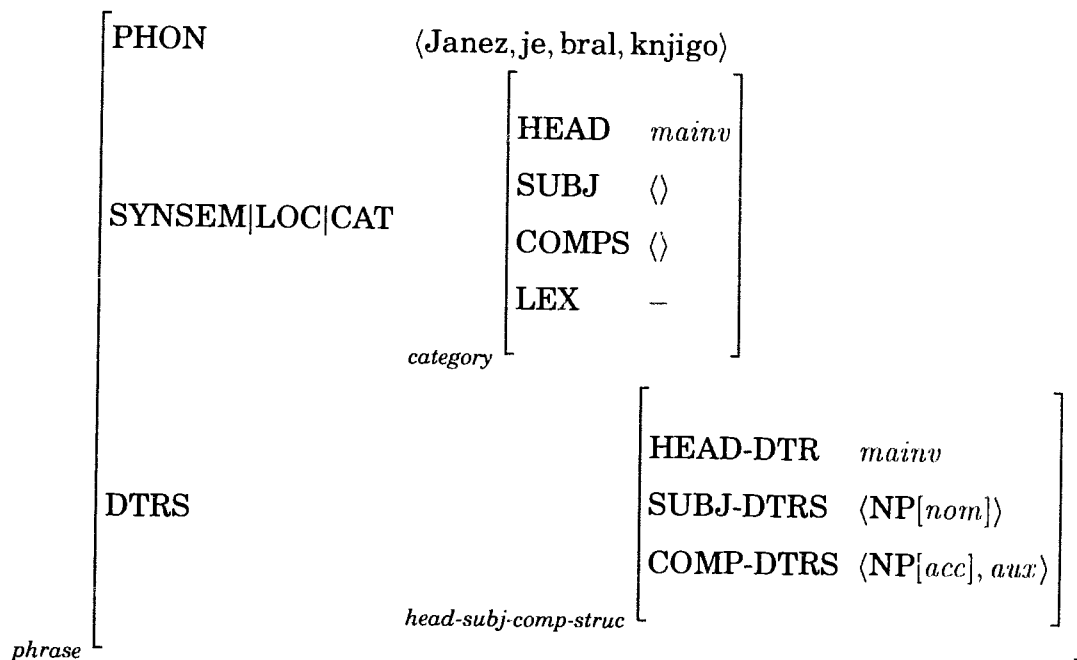phrasal projection of the verb which satisfies all ID and LP constraints.

$$
phrase \begin{bmatrix}
\text{PHON} & \langle \text{Janez}, \text{je}, \text{bral}, \text{knjigo} \rangle \\
\\
\text{SYNSEM|LOC|CAT} & category \begin{bmatrix}
\text{HEAD} & mainv \\
\text{SUBJ} & \langle \rangle \\
\text{COMPS} & \langle \rangle \\
\text{LEX} & -
\end{bmatrix} \\
\\
\text{DTRS} & head\text{-}subj\text{-}comp\text{-}struc \begin{bmatrix}
\text{HEAD-DTR} & mainv \\
\text{SUBJ-DTRS} & \langle \text{NP}[nom] \rangle \\
\text{COMP-DTRS} & \langle \text{NP}[acc], aux \rangle
\end{bmatrix}
\end{bmatrix}
$$

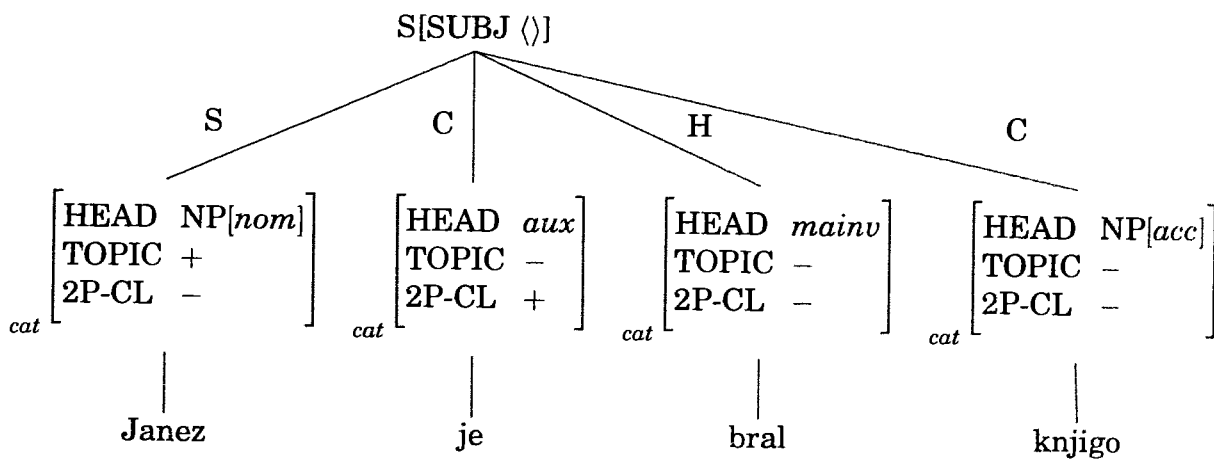Figure 4.12: Well-formed sign representing the clause

Figure 4.13: Clause construction as a tree

4.7 Summary

The flat clause structure that was proposed above gives a simpler account of syntax compared to the structure with separate VP and S nodes. The present analysis fits the empirical facts, and has the following advantages:

- none of the major constituents is discontinuous in the surface string;

- for elements that have to be rigidly ordered with respect to each other (e.g., auxiliary and the pronominals in the 2P cluster), this can be accomplished by standard LP constraints, and without employing 'argument composition';

- the fronted constituent is brought about as a consequence of LP constraints, and not by 'extraction' from the VP (implying the presence of an empty category).

This account can be easily extended to fully specify the internal order of the 2P cluster by simply adding appropriate LP constraints. We can also cover types of clauses with fixed initial positions. For example, relative clauses are always introduced by a relative pronoun, and complementized clauses by a complementizer. On this view, such words will be lexically specified [+TOPIC], and the rule LP-1 will ensure their clause-initial position.

To extend this analysis to sentences with adverbials, we would have to assume that they too are sisters of the main verb. This assumption usually creates a new problem: how to ensure a proper composition of semantic contribution without relying on the hierarchical syntactic structure. For example, in (36) the interpretation depends on the surface order of adverbials.[15]

(36) a. Frequently, the baby cried for an hour.

b. For an hour, the baby cried frequently.

This matter was investigated recently by Kasper (to appear) in the domain of German *Mittelfeld*. He suggests that a compositional treatment in the HPSG framework

---

[15]Similar examples from German are due to J. Nerbonne.

is feasible, albeit somewhat complicated. Unfortunately, the scope of modification of Slovene adverbials and its relation to surface order are still largely unexplored (but see Davis (1989) for some related investigations).

Chapter 5

Computational ramifications

Characteristics of a natural language are in modern theories stated in terms of various kinds of underlying constraints to allow for a more succinct linguistic description. When designing a computational implementation of such a modular framework, important practical questions arise:

- how can this declarative specification of interacting constraints be realized procedurally; and

- what is the computational complexity (time and space requirements) of parsing and generation.

In this chapter, I first consider some theoretical consequences of the flattened syntactic structure that was described above. I then show how an existing parser for HPSG can be adapted to process grammars of this sort.

5.1 ID/LP grammars

In traditional phrase structure rules, immediate dominance is correlated with linear order. The right-hand side of the rule specifies the total order of daughter constituents. For example, in (37) the constituent C must precede D, while the position of B is unrestricted.

49

(37) Context-free rules

$A \rightarrow BCD$

$A \rightarrow CBD$

$A \rightarrow CDB$

In the case of languages with a high degree of ordering freedom, spelling out linearizations in this way would require a large number of rules, 'leading to grammars that overlook key generalizations' (Shieber 1984:135). In order to avoid this, Gazdar and Pullum (1982) introduced an extension of Generalized Phrase Structure Grammar in which the two components are separated: the grammar consists of a set of ID rules and a set of LP rules.

Informally, an ID rule of the form $M \rightarrow D_1, \ldots, D_n$ states that a mother constituent M may immediately dominate a multiset of daughter constituents $D_1, \ldots, D_n$ without specifying the order of daughters. An LP rule $A < B$ applies to any ID rule that introduces siblings of categories A and B, and it states that in such case, B must not precede A. For example, the set of context-free rules in (37) is equivalent to the combination of two rules in (38).

(38) a. ID rule: $A \rightarrow B, C, D$

    b. LP rule: $C < D$

Using this concept, we can express grammatical rules more concisely and capture generalizations about linear order. Because of its advantages, most subsequent variants of GPSG (and later also HPSG) adopted some version of the ID/LP notation.

5.2 Complexity

Parsing is a task of recovering the syntactic structure of the input sentence. ID/LP grammars have been used in parsing in two ways. The first option is to expand (compile) the ID/LP grammar into the object grammar in a more familiar format. Every

ID/LP grammar can be converted into an equivalent context-free grammar (Shieber 1984:145). We could then use standard CFG parsing algorithms, for example that of Earley (1970). However, a realistic grammar describing a fragment of natural language would expand into a huge[1] CFG. Even though the time complexity of the Earley parsing algorithm is in $O(|G|^2 n^3)$ where $|G|$ is the size of the grammar and n the input length, the effects of the grammar size would dominate complexity for a relevant range of input lengths (Berwick and Weinberg 1982).

The other alternative is direct interpretation. Shieber (1984) generalized Earley's algorithm to parse an ID/LP grammar directly, circumventing the initial expansion stage. As Barton (1985) proves, the problem of ID/LP parsing is $\mathcal{NP}$-complete and the worst-case runtime of Shieber's algorithm is exponential in grammar size. But even so it is, on the average, still faster (because of its more concise data representation) than Earley's algorithm on the object grammar.

Barton (1985:212) demonstrates that the best case for Shieber's algorithm represents the case of strong constraints, and that the algorithm performs worse if the constraints are weakened. Languages with a high degree of ordering freedom possess fewer LP constraints and therefore present a computationally harder problem for this algorithm.

5.3 Parsing efficiency

From the computational perspective, parsing is a search problem. Regardless of the internal data representation and searching strategy, parsing complexity is related to the number of hypotheses that the parser has to keep track of and eventually explore.

The amount of work that the parser has to do in turn depends on how much disambiguation information it gets. If the parser is given enough disambiguation information early in the process, it can prune the search space and thus decrease time requirements.

---

[1]Barton (1985:205) estimates the size to 'from hundreds or thousands up to trillions of rules.'

Barton (1985:207ff) identifies two properties of natural language grammars that are a source of inefficiency in Shieber's algorithm because they lead to combinatorial explosion:

- lexical ambiguity: lexical entries of the same surface form may belong to different categories (for example, the English word 'like' may be viewed as a verb, a noun, an adjective or a preposition);

- empty categories: in most linguistic theories, grammars may include some form of categories that are part of the derivation but are empty on the surface (i.e., phonetically).

In the case of lexical ambiguity, the parser cannot assign to the lexical entry its categorial identity unequivocally—all possibilities must be considered. Even worse is the situation with empty categories: they must be hypothesized in various positions in the sentence.

It is commonly assumed that natural languages are efficiently parsable, and that natural language grammars are to a large extent shaped by considerations of processing ease (Hawkins 1990). Therefore, the question of efficiency matters beyond pure practical concerns: any plausible grammar model should allow for efficient processing.

As it is often observed, languages that do not rely heavily on word order to convey syntactic information, make up for weakened structural constraints elsewhere, often by rich morphological information. For example, Slovene nouns inflect for number and case ($3 \times 6 = 18$ forms), adjectives for gender, number and case, and verbs for person, gender and number. Lexical items are distinguished on the surface and lexical ambiguity is significantly lower than, for instance, in English.

Among the verb and its arguments, where most of the ordering flexibility is exhibited, (overtly marked) morphological distinctions again play an important disambiguating role. For instance, subjects are generally nominative, primary objects mostly accusative and so forth. In HPSG, case assignment is specified in the lexicon. The

parser can use this information in order to assign grammatical roles to nominal categories deterministically.[2]

As usual, there is a trade-off between generality and efficiency. A more specialized parser can exploit additional information and thus achieve greater efficiency than a more general one. Grammars of languages like Slovene contain the following relevant features:

(39) a. strict adjacency and precedence constraints between subconstituents on the phrase level;

    b. ordering variation on the clause level;

    c. lexically available case assignment information.

In the final section of this chapter, I will outline how we can adapt a parser to process grammars of this sort. As we will see below, it only takes relatively minor modifications to an existing chart parser for HPSG.

5.4 Chart parser

Chart parsing has its origins in the works of M. Kay and R. Kaplan (see a textbook introduction to this technique in Gazdar and Mellish (1989) and references therein). It has been popular in the natural language processing community since the early 1980s for its several advantages:

- various parsing algorithms are applicable (it does not commit us to either top-down or bottom-up approach);

- the method is neutral with respect to search strategies (breadth-first, depth-first);

- direction of processing may be arbitrary (from left to right, right to left, from some head word in the middle of the sentence towards both ends);

---

[2]Except for some rare verbs that subcategorize for two nominative or two accusative arguments.

- in the event of an ungrammatical sentence, well-formed substrings can be used to locate the error and determine its type using heuristic procedures.

The idea is to exploit a data structure called a *chart* to encode the current state of the parsing process. A chart is a graph which comprises inactive and active edges spanning parts of the sentence (each word lies between two vertices). Inactive edges represent confirmed parsing hypotheses (complete well-formed constituents), while the active ones represent hypotheses that still need to be confirmed (incomplete constituents). The chart provides storage for intermediate results so that each hypothesis is verified only once (no parsing subtask has to be repeated).

5.4.1 Parsing process

The HPSG-PL system described by Kodrič, Popowich, and Vogel (1992) utilizes a bottom-up chart parser driven by subcategorization requirements. This parser operates with a chart in which each edge contains the following data:

1. a pair of vertices (source, destination) which tells what part of the sentence the edge covers;

2. a FS which represents the spanned part of the sentence; and

3. a list of expectations (which correspond to subcategorization requirements[3]).

Inactive edges have empty expectations and active edges have non-empty expectations. An inactive edge is said to 'meet an expectation' of some active edge if and only if:

(a) the two edges cover adjacent substrings of the sentence; and

(b) the sign of the inactive edge satisfies that expectation of the active edge (in practice, the two FSs must unify and all relevant constraints must be satisfied).

---

[3]I am simplifying since we are not concerned with other types of daughters here (e.g., adjuncts).

New edges are introduced to the chart through a waiting list or *an agenda*. Bottom-up parsing starts with the initialization stage in which the words from the sentence are located in the lexicon and the corresponding lexical signs are added to the agenda. Figure 5.1 and Table 5.1 depict an initialized chart of the sentence (30). Edges and vertices are numbered (vertices are shown in circles).
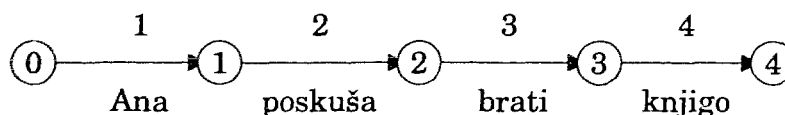

Figure 5.1: Initialized chart

edge number	vertices (S,D)	HEAD value	expectations
1	(0,1)	*noun*[nom]	⟨⟩
2	(1,2)	*verb*[finite]	⟨⟩
3	(2,3)	*verb*[nonfinite]	⟨⟩
4	(3,4)	*noun*[acc]	⟨⟩

Table 5.1: Initialized chart (edges)

After the chart has been initialized, the following steps are repeated until all possibilities are exhausted (i.e., the agenda is empty):

(40)   1. The next (current) edge is taken from the agenda.

2. If the current edge is inactive and its sign can serve as a head daughter (according to some grammar schema), then the parser creates a new active edge representing its mother sign with appropriate expectations (the new edge spans the same substring as the current inactive edge). The new edge is then placed on the agenda.

3. If the current edge 'meets' the first expectation of some edge from the chart or vice versa (i.e., one of the edges is active and the other one inactive),

then the two edges are combined: a new edge is created which spans the two constituent edges and in which the satisfied expectation is cancelled (removed from the list). The new edges is then added to the agenda.

4. The current edge is then added to the chart.

After this procedure is finished, any inactive edge that spans the entire sentence represents a successful parse.

5.4.2 Modifications

Even though the grammar contains a flattened clause structure, the fundamental assumption on which the above parsing algorithm relies is still obeyed: every mother sign is composed of adjacent daughter signs. In other words, the edges representing the head and its arguments, taken together, form a contiguous substring of the sentence.

The parsing process in (40) would be altered along the following lines so as to include the properties in (39).

(41) a. Whenever the phrasal head is not a verb (i.e., if it is a noun or a preposition), the parser can determine whether its complement immediately follows or immediately precedes the head.

b. When an active edge representing the verb is created (by Schema 3), it is placed at the *bottom* of the agenda rather than on top of it. In this way, we will postpone the application of this schema.

c. Once the active edge representing the head verb is encountered on the agenda, most of its arguments (possibly all) are already available in the chart. The parser now looks for the verb's expectations (taking into account case assignment information).

The revised parsing procedure reads as follows (modifications are italicized):

(42) 1. The next (current) edge is taken from the agenda.

2. If the current edge is inactive and its sign can serve as a head daughter (according to some grammar schema), then the parser creates a new active edge representing its mother sign with appropriate expectations (the new edge spans the same substring as the current inactive edge). The new edge is then placed *at the bottom of the agenda if it represents a verb and on top of the agenda otherwise.*

3. If the current edge *is inactive or its sign's head is not a verb, then*

   (a) if the current edge 'meets' the first expectation of some edge from the chart or vice versa (i.e., one of the edges is active and the other one inactive), then the two edges are combined: a new edge is created which spans the two constituent edges and in which the satisfied expectation is cancelled (removed from the list). The new edge is then placed on the agenda.

   otherwise (the current edge is active and its sign's head is a verb) proceed as follows:

   (b) *Successively locate edges that 'meet' the current edge's expectations (not necessarily all of them are already available in the chart) and combine them with the current edge. A new edge*[4] *is created which spans the verb and all its arguments that have been found (the corresponding satisfied expectations are removed). The new edge is then placed at the bottom of the agenda.*

4. The current edge is then added to the chart.

If all of its arguments have already been parsed when the verb is encountered on the agenda, the inactive edge spanning the clause will be created in a single step—no intermediate active edges are created in such case which will improve efficiency (active edges which do not contribute to the parse represent 'wasted activity'). A chart parsing example is described in appendix A.

---

[4]The new edge will be inactive if all verb's arguments have been found and active otherwise.

In summary, even though the constraints in the grammar are weakened by adopting a flattened clause structure with fewer LP rules, this need not necessarily induce significant parsing difficulties since other disambiguating information is available. A standard chart parser can be easily modified to handle such grammars.

Chapter 6

Discussion and conclusions

In this final chapter, I sum up the analysis developed in this thesis and compare it to some other approaches which address problems posed by flexible word order.

On the grounds of word order and VP constituency in particular, I have argued that an analysis in terms of flat structure is appropriate for the Slovene clause. An HPSG model was sketched which is simpler than the previous analysis of Mežnarič (1992)[1] in that:

- it does not employ any additional devices, such as 'argument composition'; and

- word order variation is treated as a local phenomenon rather than a long-distance dependency.

The clause is constructed by a variant of ID Schema 3 which accounts for both finite and nonfinite verbs and for both simple and compound tenses. To cover other kinds of phrases, we would use Schema 2 (a preposition and its object, a deverbative noun and its complements), Schema 4 (complementized clauses), Schema 5 (adjuncts of various kinds), and so on. Schema 1 (HEAD–SUBJ) is not needed for Slovene.

The proposed method obeys the locality of ordering constraints and the standard relation between word order and phrase structure which is advantageous from the

---

[1]Apart from that one, I am not aware of any other nontransformational (GPSG or HPSG) accounts dealing with the clause structure of any South Slavic language.

practical point of view. For example, standard parsing algorithms may be used to process the grammar. Also, we do not assume any empty constituents (unrealized subjects are only present in the subcategorizations frames but not in the constituent structure) which is important for parsing efficiency.

Among the limitations, it has to be noted that this analysis 'overgenerates' in that it licenses all grammatical orderings making no distinctions between marked and unmarked ones. But since the 'preferred' order depends heavily on the pragmatic situation, one cannot make that distinction unless a sufficiently formalized representation of the context (i.e., the previous sentences) is available. Furthermore, as we have seen in section 2.4, information on prosodic structure is important as well.

6.1   Related work

As I mentioned in chapter 2, various proposals exist for the treatment of word order variation. GB's philosophy of binary branching trees with structurally defined positions in the sentence holds that alternative orderings (i.e., different from the 'base') are derived via 'scrambling.'

In this process, a constituent is moved from its canonical position (leaving behind an empty trace) and adjoined to some position higher in the tree. More than one constituent may be scrambled in the same clause, and the same item may be moved several times in succession, as (simplified) shown in Figure 6.1 taken from Borsley and Rivero (to appear). Ultimately, the terminal yield of the non-empty positions corresponds to the surface order.

Scrambling is widely assumed in the GB treatments, but rarely examined in detail (see Haegeman (1991:539ff) and references cited there). For example, while there has been some research of Slovene in the GB paradigm (surveyed in Golden 1990), none of it tackles word order variation. It is often seen as a superficial property of little theoretical interest, and most attempts to deal with word order in Slavic languages are usually restricted to some particular kind of movement (cf. Rivero 1991, Borsley

```
              IP
             /  \
          ⌐NP⌐   IP
           ▲    /  \
           ¦   I    VP
            \      /  \
             \    t    VP
              \   ▲   /  \
               \  ¦  V    t
                \  \
                 \  \____/
                  _____/
```
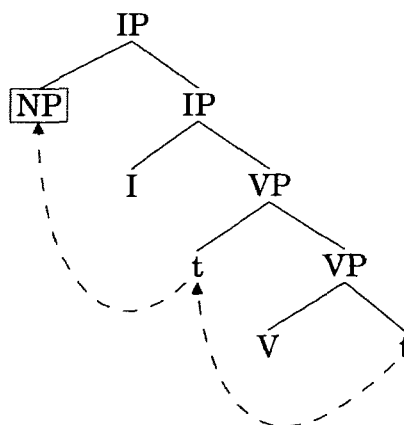
Figure 6.1: Scrambling of a constituent in GB

and Rivero to appear).

Since scrambling entails transformations between distinct syntactic representations, it is not available to nontransformational accounts. In the monostratal frameworks, on the other hand, different fundamental theoretic assumptions[2] lead to other treatments in cases where the standard view on phrase structure and word order creates problems either from the theoretical or the practical point of view.

One direction that is being explored is a modified concept of LP constraints. Engelkamp et al. (1992) discuss the German verbal complex, where well established independently motivated evidence excludes the possibility of a flat constituent structure. They argue that in such constructions ordinary LP rules cannot provide an adequate description. They incorporate a mechanism into standard feature unification formalisms which allows the LP constraints to be stated between nonsibling constituents. The domain of locality is extended by encoding LP constraints directly in feature structures.

This method is not directly applicable to Slovene as it was motivated by a different, more restricted type of word order flexibility. Even though we could account for the

---

[2]For instance, HPSG does not employ structurally distinguished positions or tree-configurational notions, such as c-command.

internal order of the clitic cluster by adopting this method and the hierarchical structure with a VP node, we would not gain much: the problem of discontinuity between the verb and the object would still remain.

Avgustinova and Oliva (1990) suggest an approach reconciling declarative phrase structure tradition with the Functional Sentence Perspective, bearing in mind in particular the properties of Slavic languages. This idea is elaborated by Oliva (1992), who develops an HPSG model that allows ordering constraints to hold for a list of constituents (i.e., such a list is the domain of locality). Lists are encoded in a binary branching sortal hierarchy (using the attributes *first* and *rest*).

On his account, the burden of constraining the surface order is shifted to the sortal hierarchy. Consider the structure of a sample German sentence in Figure 6.2 taken from Oliva (1992:189).
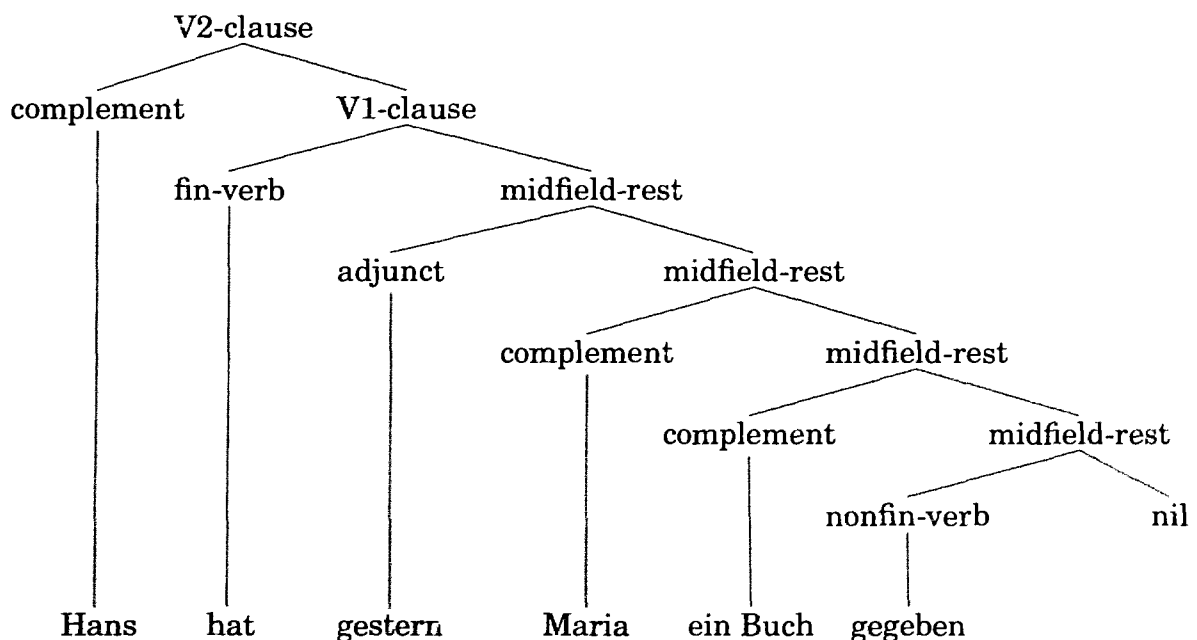
Figure 6.2: Binary branching structure (on the basis of surface order)

This approach seems to have adequate descriptive power to model the word order variation. But it makes it difficult to capture linguistic generalizations about word

order concisely, since this separate sortal hierarchy is specified solely on the basis of linear order rather than in terms of constituent structure. For example, a node in a structure like the one in Figure 6.2 may represent a substring of the sentence which is not necessarily a constituent.

Reape (1990) explores a more radical idea of separating the 'composition' part of syntactic structure from its actual realization in the utterance. The surface string is not derived as the terminal yield of the syntactic tree representing the sentence. Rather, it is derived by a recursive process (in a bottom-up way) from 'word order domains' associated with subconstituents.

Linear order is determined (locally) within these domains which can be formed compositionally via the 'sequence union' operation. This operation merges the elements of subdomains while preserving all previously imposed ordering constraints. Additionally, the domain of a daughter constituent may be the same as the domain of its mother. It is this property that allows the elements of the daughter's domain to be interleaved with the other elements of the mother's domain.

This method provides an elegant and principled way of accounting for a broad range of word order phenomena. However, in the particular case of Slovene, it seems that we don't need this additional expressive power. In the absence of any convincing evidence for the hierarchical clause structure, a more conservative approach (i.e., one which does not depart from standard assumptions) seems preferable to a more powerful one. Further investigation will have to show whether a wider coverage of word order phenomena would force us to adopt a more powerful approach.

6.2 Concluding remarks

In this thesis, I have presented the first nontransformational model of the Slovene clause which includes the treatment of both finite and nonfinite verbs and both simple and compound tenses. It has been shown how the notoriously difficult problems related to word order can be elegantly accounted for by flattening the clause structure in the

constraint-based framework of HPSG.

What is common to the above alternative approaches is that each of them is forced to do away with some standard assumptions. The analysis presented in this thesis suggests that a significant degree of variation can be accommodated within standard frameworks. Since most of the ordering flexibility in Slovene is exhibited on the clause level, this analysis can be straightforwardly extended to cover a wider range of syntactic structure.

The grammar containing a flat clause can be processed by a chart parser without any major changes. The standard parsing algorithm requires only minor changes and I have shown how they can be incorporated.

The approach adopted here should also be applicable to related Slavic languages, even though some modifications would be necessary, particularly in the treatment of clitics. For instance, in Serbo-Croatian we would have to account for the alternation between the 'syntactic' and the 'phonological' position of the clitic cluster.

Results of the work presented here provide a basis for future research. It remains to be seen how we can extend this model to include adverbials and incorporate semantic phenomena (e.g., scope, coreferential interpretation, and so forth) with the goal to set up a more complete formal description of the Slovene grammar that could serve as a foundation for computational implementation.

Appendix A

Parsing example

This appendix exemplifies the chart parsing algorithm in more detail.

Figure A.1 depicts the chart after the sentence (30) has been parsed by the algorithm outlined in chapter 5. Edges and vertices are numbered (vertex numbers are shown in circles). Inactive edges are represented by solid lines and active edges by dashed lines.
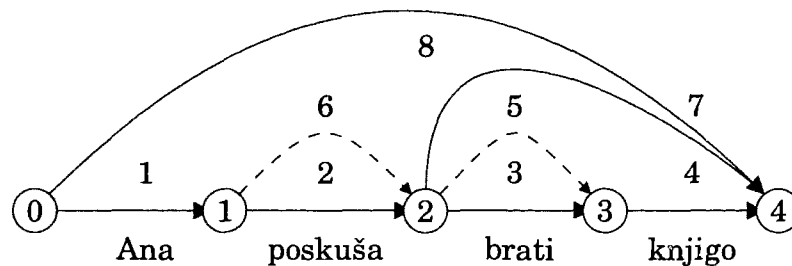

Figure A.1: Sample chart

Edge numbers reflect the sequence in which the edges have been generated (but not the sequence in which they have been added to the chart). It has to be borne in mind that this order depends on the chosen control strategy and on the organization of the agenda.

Table A.1 contains the information associated with the edges from the chart in Figure A.1. The third column in the table shows the set of edges from which that

edge no.	vertices (S,D)	const. edges	HEAD value	expectations
1	(0,1)	{}	*noun*[nom]	⟨⟩
2	(1,2)	{}	*verb*[finite]	⟨⟩
3	(2,3)	{}	*verb*[nonfinite]	⟨⟩
4	(3,4)	{}	*noun*[acc]	⟨⟩
5	(2,3)	{3}	*verb*[nonfinite]	⟨C:NP[acc]⟩
6	(1,2)	{2}	*verb*[finite]	⟨S:NP[nom], C:VP[nonfin]⟩
7	(2,4)	{5,4}	*verb*[nonfinite]	⟨⟩
8	(0,4)	{6,7,1}	*verb*[finite]	⟨⟩

Table A.1: Set of edges

particular edge has been created. If this set is empty, the edge was created during the initialization and represents a lexical entry. The column 'HEAD value' contains the SYNSEM | LOC | CAT | HEAD value of the sign that represents the constituent that the edge spans. The list of expectations is empty if the edge is inactive and non-empty otherwise (recall that labels NP and VP abbreviate FSs).

Table A.2 illustrates the sequence of steps as they are carried out during the parsing of the sentence (30) according to the procedure described in section 5.4.2.

step	action	agenda	chart contents
1	push	1	{}
2	push	21	{}
3	push	321	{}
4	push	4321	{}
5	pop	321	{4}
6	gen-act	215	{4,3}
7	gen-act	156	{4,3,2}
8	pop	56	{4,3,2,1}
9	gen-inact	6	{4,3,2,1,7,5}
10	gen-inact		{4,3,2,1,7,5,8,6}

Table A.2: Parsing process

The second column in the table tells what action is being performed in that step. The actions can be described as follows.

push: A word is located in the lexicon, an inactive edge is created and put on top of the current agenda.

pop: The current edge (the one on top of the agenda) is moved into the chart.

gen-act: A new active edge is created using the current edge as the head daughter. The new edge is placed at the bottom of the agenda and the current edge is added to the chart.

gen-inact: A new inactive edge is created using the current edge and the appropriate existing edges. The new edge and the current edge are added to the chart.

The third column and the fourth column in Table A.2 contain the list of edges on the agenda and the contents of the chart, respectively (after the step has been completed). Note that the agenda is ordered whereas the chart is not.

Let us summarize the exemplified parsing process. Initally, the agenda and the chart are empty. The initialization stage comprises steps 1–4. The expectations for the two verbs are created in the steps 6 and 7. These hypotheses are then verified: the controlled complement is found in step 9, and the finite verb is combined with both its arguments in step 10. After the process is completed, the chart contains an inactive edge spanning the entire sentence: a successful parse has been found.

Notice that in this particular example both arguments of the finite verb are already available when the verb is encountered on the agenda. Correspondingly, an inactive edge (number 8) is created. If that were not the case, an active edge (containing the expectations for the 'missing' argument) would be created and put on the agenda—in such a case the verb would not be combined with its arguments in a single step.

Appendix B

Attributes and values

In this appendix, I summarize some syntactic properties that would have to be included in a full-fledged formal grammar of a fragment of Slovene. Figure B.1 shows a sortal hierarchy of parts of speech (SYNSEM | LOC | CAT | HEAD values).
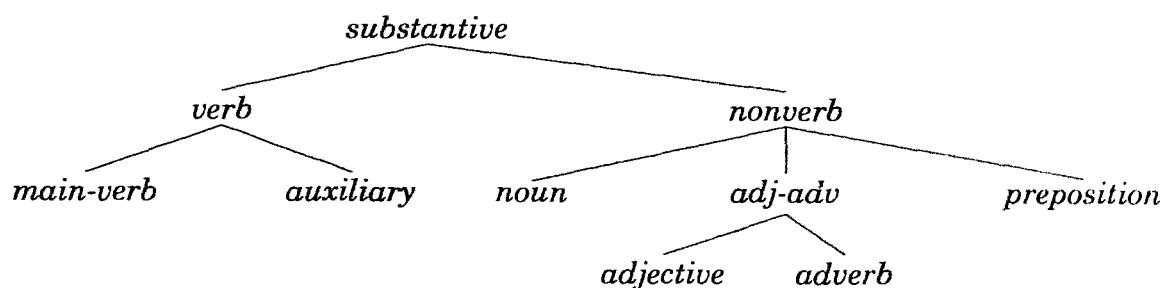
Figure B.1: Partial part of speech hierarchy

However, given the current stage of the HPSG theory and the fact that by far the most research has been done for English, some of the attributes that we need for Slovene do not have a definitive place in the sign structure as yet.

For example, English has relatively few inflected forms, and the 'verb form' hierarchy (Figure 4.2, p. 29) includes both tense (*past* / *nonpast*) and agreement (*3rdsing* / *non-3rdsing*) distinctions. On the other hand, inflected forms of Slovene verbs would have

to be classified along several dimensions with tense and agreement kept apart. Figure B.2 exemplifies a classification of verbs with respect to tense.
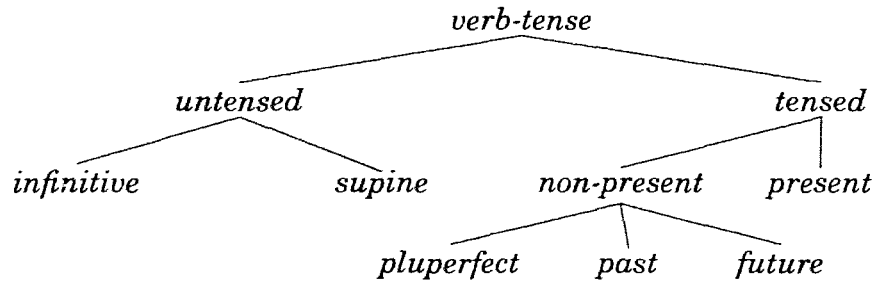

Figure B.2: Part of the classification of Slovene verbs

Finally, Table B.1 indicates some features that are associated with Slovene words, and Table B.2 lists their possible values.

category	attributes
verb	tense, mood, person, number, gender
noun	case, number, gender
adjective	case, number, gender, definiteness, degree
adverb	degree

Table B.1: Syntactically relevant attributes of Slovene words

attribute	values
mood	indicative, imperative, conditional, optative
person	first, second, third
number	singular, dual, plural
gender	masculine, feminine, neuter
case	nominative, genitive, dative, accusative, locative, instrumental
definiteness	definite, indefinite
degree	positive, comparative, superlative

Table B.2: Attributes and their possible values

Bibliography

Anderson, S. R. (1992). *A-Morphous morphology*. Cambridge: Cambridge University Press.

Anderson, S. R. (1993). Wackernagel's revenge: clitics, morphology, and the syntax of second position. *Language 69*(1), 68–98.

Ard, W. J. (1975). *Raising and Word Order in Diachronic Syntax*. Ph. D. thesis, UCLA.

Avgustinova, T. and K. Oliva (1990). Syntactic description of free word order languages. In *Proceedings of the 13th International Conference on Computational Linguistics*, Volume 3, Helsinki, pp. 311–313.

Barton, Jr., G. E. (1985). On the complexity of ID/LP parsing. *Computational Linguistics 11*(4), 205–218.

Bauer, L. (1990). Be-heading the word. *Journal of Linguistics 26*(1), 1–31.

Bennett, D. C. (1987). Word-order change in progress: the case of Slovene and Serbo-Croat and its relevance for Germanic. *Journal of Linguistics 23*(2), 269–287.

Berwick, R. C. and A. S. Weinberg (1982). Parsing efficiency, computational complexity and the evaluation of grammatical theories. *Linguistic Inquiry 13*(2), 165–191.

Bolta, M. (1986). Slovenski glagoli z nedoločniškim dopolnilom v TP-pristopu. *Slavistična revija 36*(4), 419–432. Ljubljana.

Bolta, M. (1989). Besednozvezna zgradba stavka v vezalno-navezovalni teoriji. *Znanstvena revija 2*, 147–158. Maribor.

Bolta, M. (1990). Slovenski in angleški povratni zaimki v vezalno-navezovalni teoriji. *Anthropos 22*(3–4), 333–349. Ljubljana.

Borsley, R. D. (1984). On the nonexistence of VP's. In W. de Geest and Y. Putseys (Eds.), *Sentential Complementation*, pp. 55–65. Dordrecht: Foris Publications.

Borsley, R. D. (1987). Subjects and complements in HPSG. Report CSLI-87-107, Center for the Study of Language and Information, Stanford.

Borsley, R. D. and M.-L. Rivero (to appear). Clitic auxiliaries and incorporation in Polish. *Natural Language & Linguistic Theory*.

Bunt, H. (1990). Disco-trees and their use in DPSG. In *Proceedings of the Symposium on Discontinuous Constituency*, Tilburg University, pp. 7–19. Institute for Language Technology and Artificial Intelligence.

Carpenter, B. (1992). *The Logic of Typed Feature Structures*. New York: Cambridge University Press.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris Publications.

Chomsky, N. (1986). *Barriers*. Cambridge: MIT Press.

Davis, M. G. (1989). *Aspects of Adverbial Placement in English and Slovene*. München: Sagner.

Dik, S. C. (1989). *The Theory of Functional Grammar*. Dordrecht: Foris Publications.

Dowty, D. R. (1990). Toward a minimalist theory of syntactic structure. In *Proceedings of the Symposium on Discontinuous Constituency*, Tilburg University, pp. 34–72. Institute for Language Technology and Artificial Intelligence.

Earley, J. (1970). An efficient context-free parsing algorithm. *Communications of the ACM 13*(2), 94–102.

Engelkamp, J., G. Erbach, and H. Uszkoreit (1992). Handling linear precedence constraints by unification. In *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*, Newark, DE, pp. 201–208.

Gazdar, G., E. Klein, G. K. Pullum, and I. A. Sag (1985). *Generalized Phrase Structure Grammar*. Oxford: Basil Blackwell.

Gazdar, G. and C. S. Mellish (1989). *Natural language processing in Prolog*. Wokingham: Addison-Wesley Publishing.

Gazdar, G. and G. K. Pullum (1982). *Generalized Phrase Structure Grammar: a theoretical synopsis*. Bloomington: Indiana University Linguistics Club.

Golden, M. (1990). Transformational-generative approaches to Slovene syntax. *Slovene Studies 12*(2), 205–215.

Grewendorf, G. and W. Sternfeld (Eds.) (1990). *Scrambling and Barriers*. Amsterdam: John Benjamins Publishing.

Gunji, T. (1987). *Japanese Phrase Structure Grammar: a unification-based approach*. Dordrecht: Reidel.

Haegeman, L. M. V. (1991). *Introduction to Government and Binding Theory*. Oxford: Basil Blackwell.

Hajičová, E., P. Sgall, and H. Skoumalová (1993). Identifying topic and focus by an automatic procedure. In *Proceedings of the 6th Conference of the European Chapter of the Association for Computational Linguistics*, Utrecht, pp. 178–182.

Hale, K. (1983). Warlpiri and the grammar of non-configurational languages. *Natural Language & Linguistic Theory 1*(1), 5–47.

Halpern, A. L. (1992). *Topics in the placement and morphology of clitics*. Ph. D. thesis, Stanford University. To be published by CSLI Publications.

Hawkins, J. A. (1983). *Word Order Universals*. New York: Academic Press.

Hawkins, J. A. (1990). A parsing theory of word order universals. *Linguistic Inquiry 21*(2), 223–261.

Hinrichs, E. and T. Nakazawa (1990). Subcategorization and VP structure in German. In S. Hughes and J. Salmons (Eds.), *Proceedings of the Third Symposium on Germanic Linguistics*. John Benjamins Publishing. To appear.

Hudson, R. A. (1984). *Word Grammar*. Oxford: Basil Blackwell.

Jackendoff, R. S. (1977). X̄ *Syntax: A Study of Phrase Structure*. Cambridge: MIT Press.

Kaisse, E. M. (1985). *Connected speech*. London: Academic Press.

Kasper, R. (to appear). Adjuncts in the Mittelfeld. In J. Nerbonne, K. Netter, and C. Pollard (Eds.), *German in Head-driven Phrase Structure Grammar*. Stanford: CSLI Publications. In press.

Kodrič, S., F. Popowich, and C. Vogel (1992). The HPSG-PL system—Version 1.2. Technical report CMPT 92–08, Simon Fraser University.

Laughren, M. (1989). The configurationality parameter and Warlpiri. In L. Marácz and P. Muysken (Eds.), *Configurationality: The Typology of Asymmetries*, pp. 319–353. Dordrecht: Foris Publications.

Marácz, L. and P. Muysken (Eds.) (1989). *Configurationality: The Typology of Asymmetries*. Dordrecht: Foris Publications.

Mel'čuk, I. A. (1988). *Dependency syntax: theory and practice*. Albany: SUNY Press.

Mežnarič, J. (1992). Clitics and Wackernagel position in Slovenian. Master's thesis, University of Edinburgh.

Miller, P. H. (1992). *Clitics and Constituents in Phrase Structure Grammar*. New York: Garland.

Monachesi, P. (1993). Object clitics and clitic climbing in Italian HPSG grammar. In *Proceedings of the 6th Conference of the European Chapter of the Association for*

Computational Linguistics, Utrecht, pp. 437–442.

Oliva, K. (1992). The proper treatment of word order in HPSG. In *Proceedings of the 14th International Conference on Computational Linguistics*, Nantes, pp. 184–190.

Orešnik, J. (1985). O naslonskem nizu v knjižni slovenščini. *Jezik in slovstvo 31*(6), 213–215. Ljubljana.

Perlmutter, D. M. (Ed.) (1983). *Studies in Relational Grammar I*. Chicago: University of Chicago Press.

Pollard, C. J. and I. A. Sag (1987). *Information-Based Syntax and Semantics, Volume 1: Fundamentals*. Stanford: CSLI Publications.

Pollard, C. J. and I. A. Sag (in press). *Head-driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Pullum, G. and D. Wilson (1977). Autonomous syntax and the analysis of auxiliaries. *Language 53*(4), 741–788.

Radford, A. (1988). *Transformational Grammar*. Cambridge: Cambridge University Press.

Reape, M. (1990). Getting things in order. In *Proceedings of the Symposium on Discontinuous Constituency*, Tilburg University, pp. 125–137. Institute for Language Technology and Artificial Intelligence.

Rivero, M.-L. (1991). Long head movement and negation: Serbo-Croatian vs. Slovak and Czech. *Linguistic Review 8*(2–4), 319–351.

Sag, I. A. and C. J. Pollard (1989). Subcategorization and head-driven phrase structure. In M. R. Baltin and A. S. Kroch (Eds.), *Alternative Conceptions of Phrase Structure*, pp. 139–181. Chicago: University of Chicago Press.

Sgall, P., E. Hajičová, and E. Benešová (1973). *Topic, Focus and Generative Semantics*. Taunus: Kronberg.

Shieber, S. M. (1984). Direct parsing of ID/LP grammars. *Linguistics and Philosophy* 7(2), 135–154.

Siewierska, A. (1988). *Word Order Rules*. New York: Croom Helm.

Speas, M. (1990). *Phrase Structure in Natural Language*. Dordrecht: Kluwer.

Toporišič, J. (1982). *Nova slovenska skladnja*. Ljubljana: Državna založba Slovenije.

Toporišič, J. (1984). *Slovenska slovnica* (2nd ed.). Maribor: Založba Obzorja.

Uszkoreit, H. (1986). Constraints on order. *Linguistics 24*, 883–906.

Uszkoreit, H. (1987). *Word Order and Constituent Structure*. Stanford: CSLI Publications.

Wackernagel, J. (1892). Über ein Gesetz der indogermanischen Wortstellung. *Indogermanische Forschungen 1*, 333–436.

Yokoyama, O. T. (1986). *Discourse and Word Order*. Amsterdam: John Benjamins Publishing.