Slovene Studies 15/1-2 (1993) 115-138

VP CONSTITUENCY AND CLAUSE STRUCTURE IN SLOVENE

Sandi Kodrič*

1. Introduction

In recent decades syntax has been regarded as one of the core subjects of linguistics. It studies the ways in which words are put together to form phrases, clauses and sentences. Syntax remains an area of several competing theories looking for the simplest formalized explanation: we want to cover the most facts with the fewest postulations, while keeping in mind the range of variation across languages.

Even though syntactic theories agree in various respects, one feature that distinguishes them into *transformational* and *nontransformational* is the issue of levels of syntactic representation. Transformational approaches, such as Government-Binding (GB) theory (Chomsky 1981, 1986), assume a sequence of levels and transformations between them. On the other hand, nontransformational theories, such as Generalized Phrase Structure Grammar (GPSG, Gazdar & alii 1985) and Head-driven Phrase Structure Grammar (HPSG, Pollard & Sag 1987, in press), use only one level of syntactic representation and with it they associate well-formedness conditions of various kinds.[1]

To some extent, most syntactic theories reflect the influence that the structure of English had on the their design. For instance, grammatical relations are in English expressed by linear order whereas many Slavic languages can exploit morphological distinctions. In English, for example, the subject usually precedes the verb and the object usually follows it, whereas in most Slavic

* The research reported here was carried out during my graduate studies at Simon Fraser University, British Columbia, Canada. I am indebted to my supervisor Fred Popowich for his valuable guidance. I was supported by the Jožef Stefan Institute and by the National Science and Engineering Research Council of Canada. For their helpful comments on earlier versions of this paper, I am also grateful to P. McFetridge, J. Calder, W. Roberts, J. Moskovljević, T. Erjavec and the anonymous referees.

1 Yet another approach is that of Lexical Functional Grammar (LFG, Bresnan 1982) which has two syntactic representations that are not transformationally related.

languages these two elements are distinguished by overt case marking (e.g., nominative vs. accusative inflection).

Correspondingly, the ordering of elements within a clause is much more free (less restricted) in Slavic languages than in English. These *free word order* languages present an important problem, particularly in computational linguistics where we aim towards a syntactic description which is sufficiently formalized for subsequent implementation on a computer.

This paper is concerned with nontransformational models of the Slovene clause structure. More precisely, I propose a treatment of the simple clause within *HPSG*, a modern linguistic theory designed with computational application in mind. The paper is organized as follows. I begin, in §2, by examining some factors that influence the ordering of constituents in Slovene clauses. In §3, I go on to investigate empirical evidence for the clause structure. I propose a uniform treatment in which verbs combine with all of their arguments in a single step. I then outline an HPSG model of a simple clause in §4, followed by some concluding remarks.

2. Preliminaries

Phrase structure approaches regard constituency as the central manifestation of syntactic structure.[2] Constituent structure of sentences is traditionally represented by tree diagrams, which are restricted in various ways. For instance, the 'non-tangling' condition means that there should be no crossed branches in the tree: every constituent (node) in the tree represents a contiguous (adjoining) part of the sentence. But there are many phenomena across languages which conflict with this assumption. For example, the relative clause in (1) is not adjacent to the pronoun ('someone') that it modifies.

(1) Someone entered the room whom I had never met before.

The phrase 'someone whom I had 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 1.[3]

---

2 Within dependency approaches, constituency is not fundamentally important (Matthews 1981: 78ff.)

3 Leaves of the tree represent the sequence of words (lexical items) in the sentence. Inner nodes (nonlexical items) are labelled with categorial information, e.g., S — sentence, V — verb, NP — noun phrase, VP — verb phrase.

In order to accommodate such exceptions which are posed by ordering flexibility, we might:

(a) reject word order (WO) as the basis for deriving constituent structure; or

(b) accept that some syntactically determined (underlying) order may be 'scrambled' in the surface form of the sentence; or (c) 'loosen' the hierarchical constituent structure.

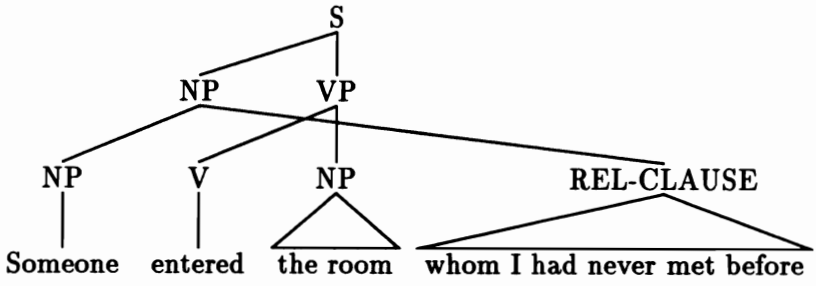

Figure 1: Structure with crossed branches

The first alternative (a) is explored by Reape (1990) and Dowty (1990) for German and English, respectively. But it departs from the standard view on phrase structure, and leads to unclear (theoretical and practical) consequences.

The second option (b) is pursued in most approaches within GB. This theory holds that alternative orderings (i.e., different from the 'base') are derived by 'scrambling.' In this transformation, a constituent is moved from its canonical position and adjoined to some position higher in the tree. More than one constituent may be scrambled, and the same item may be moved several times. Ultimately, the sequence of leaves in the derived (modified) tree corresponds to the surface order in the sentence. Scrambling is widely assumed in the GB treatments, but rarely examined in detail (see Rudin (1985), Haegeman (1991: 539ff.) and the references cited there). For example, while there has been some research into Slovene in the GB paradigm (surveyed in Golden 1990), none of it tackles WO variation.

Because scrambling is a transformation between two syntactic representations, it is not available to nontransformational accounts. In the rest of this paper, I therefore examine the third option (c). A more constrained model (i.e., one which *a priori*

excludes crossed branches and transformations) is preferable to a less constrained one: we want to restrict the class of 'possible natural language sentences' as narrowly as possible and rule out as many 'impossible sentences' as we can.[4]

I first exemplify some Slovene data and then identify its WO features which motivate the model elaborated below. Example (2) shows a simple sentence with a nominative subject and an accusative object.

(2) a. Janez piše pismo.
John-NOM writes letter-ACC
'John is writing a letter.'

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

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

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

Neutral subject-verb-object order (2a) is usually preferred in the absence of discourse-related factors, but orderings in (2b - 2d) are grammatical as well, albeit somewhat 'marked.'

As illustrated by (3), Slovene is a *null subject* (subject-drop) language: when the nominative subject of the clause can be recovered from the context, it is usually phonetically unrealized.

(3) Bere knjigo.
reads book-ACC
'He/she is reading a book.'

The above examples demonstrate that the position of the verb is in principle unrestricted: in the sense that it can be initial (3), final (2d) or medial (2a). Also, apart from the placement of the clitic cluster, there are no precedence or adjacency constraints between the verb and any of its arguments (subject, object).[5] More generally, WO in Slavic languages is affected by the following factors (Ard 1975: 95): (a) grammatical rules of the language (relevant for the 'fixed' WO); (b) discourse situation; (c) rhythmic and stylistic principles.

---

4 When implementing a grammar model on a computer, this amounts to restricting the search space, thereby leaving as few hypotheses as possible for the computer to explore.

5 In this paper, I disregard the invariable second position of the clitic cluster; for different non-transformational treatments of Slovene clitics, see Mežnarič (1992) and Kodrič (1993), respectively.

Their mutual interaction is mainly such that (b) and (c) operate only where (a) leaves some latitude. Indeed, Ard (1975) finds that speakers find a sentence 'more unacceptable' if (a) is violated, compared to the violations of (b) or (c). But the relative 'strength'of particular principles is specific to the language in question, and it usually varies among its speakers.

In Slovene, 'fixed' WO comprises phrase-internal structure (e.g., placement of prepositions and conjunctions) and certain positions reserved for elements with a special significance, such as initial for question words and second for the clitic cluster. Apart from that, the ordering is largely governed by discourse principles (Toporišič 1984: 532).

The Prague School tradition of Functional Sentence Perspective (Sgall & alii 1973) holds that a scale of 'communicative dynamism' determines how the utterance is divided into 'topic' part and 'focus' part. 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 WO' languages, it is often signalled by WO. In Slovene, the major constituents of the clause are usually arranged so that the 'intonation center,' which coincides with the focus, falls on the last accented word[6] (Toporišič 1984: 448).

As mentioned above, some orderings are only appropriate in certain restricted contexts. Recall the sentence (2b) (see p. 118). With this order, this sentence presupposes a context where the speaker expects the hearer to be already aware of the existence of the letter and the act of writing.[7] For example, it is acceptable as the answer to the question in (4).

(4) Ali Ana piše pismo.
QUE Ann-NOM writes letter-ACC
'Is Ann writing the letter?'

Hence, (2b) is more accurately translated into English by the cleft sentence in (5).

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

---

6 The same state of affairs is found in Polish, Czech and Bulgarian (Dik 1989: 364).

7 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, as in many other languages, the focus can be conveyed prosodically; the speaker can employ an emphatic stress on the focused word in the utterance. In (6), the stressed (focused) word is capitalized.[8]

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

To summarize, constituent order interacts with the pragmatic situation established by the previous context and with the prosodic structure of the sentence. Apart from the rules of 'fixed' WO, the order of the major constituents of the clause is determined by their communicative prominence: the new (most important) piece of information generally (barring the use of contrastive stress) comes last.

3. Clause structure

It has been suggested that languages with relatively unconstrained WO 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 verb phrase (VP) constituent in the structure, as illustrated by Figures 2 and 3[9] (the term 'VP' stands for a projection of the verb which contains its non-subject arguments and possibly adverbials, but is distinct from the S (sentence) category).

This view of two radically different classes of languages was later refuted, and researchers in the GB paradigm attempted instead to formulate the 'configurationality' parameter of variation among languages (Marácz & Muysken 1989). The criterion considered to be the most important was subject-object asymmetry — the two behave differently (e.g., subject-verb agreement; subject-drop; objects may passivize etc.). It was claimed that the verb and the object form a VP node, but as Speas (1990: 162) points out, this claim 'that [the subject and the object] must be *structurally*

---

8 The position of the stressed element is generally unrestricted, as well.

9 Note that under many recent GB assumptions the structure in Figure 3 is not possible because it is not binary-branching.

distinguishable is a theory-internal prediction made by GB,' and it is not shared by some other theories.[10]

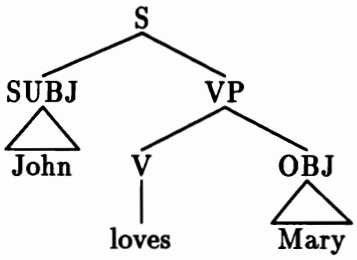

Figure 2: 'Configurational' (hierarchical) structure

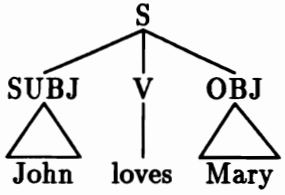

Figure 3: 'Nonconfigurational' (flat) structure

The evidence for the VP constituent in particular has been used as one of the prime diagnostics against the flat clause structure. The argumentation as to whether some construction is a constituent mainly involves distributional and semantic criteria which I now investigate in some detail. Distributional VP constituency 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 (7a), preposed (7b) or replaced by a pro-form (7c).

(7) John said he would read the book,
  a. and he did [ ].
  b. and read the book he did.
  c. and so he did.

But for Slovene, analogous tests do not give conclusive results (see also Bolta 1989). First, (8) exemplifies a deletion which is

10 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).

roughly comparable to (7a). We can delete the verb only (8a), but not the VP in (8b). The object clitic *ga* cannot be deleted.

(8) Janez pravi, da bo pojedel zajtrk
John says that will eat breakfast-ACC
a. in res ga bo.
and indeed it-ACC will.
b *in res bo.
and indeed it-ACC will.

Second, VP-ellipsis is another textbook 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 (9a) or just the transitive verb (9b). This too suggests that the verb and the object are not so tightly bound together as they are in English (recall also that in Slovene they often do not appear adjacently in the clause).

(9) a. Janez knjigo piše, Ana pa bere.
John-NOM book-ACC writes Ann-NOM whereas reads
'John is writing a book, whereas Ann is reading one.'
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.'

Third, (10) exemplifies yet another test (somewhat analogous to English VP-preposing): fronting to the position immediately preceding the auxiliary (or the clitics in general). This pre-AUX position has also been taken as providing a test for constituency in Warlpiri (cf. Siewierska 1988: 159): if we disregard extra-clausal constructions, e.g., vocatives and parentheticals, any material that may precede the auxiliary (*je* in this case) forms a constituent.[11] For instance, the preceding material can be a simple noun phrase as in (10a) or a prosodically heavy that-clause as in (10b).

(10) a. Moj sosed je bral knjigo.
my neighbor-NOM AUX read book-ACC
'My neighbor read the book.'
b. Da bo domace moštvo zmagalo, je bilo jasno nikomur.
that will home team win AUX be clear everyone-DAT
'That the home team would win was clear to everyone.'
c. *Bral knjigo je Janez
read book-ACC AUX John-NOM
d. Knjigo je bral Janez.
book-ACC AUX read John-NOM
'John read the book.'

As (10c) shows, the finite verb and its object cannot be fronted (this is also the case in closely related Serbo-Croatian (Rivero

11 This is similar to the situation in Czech (Toman 1986).

1991: 333). Whenever the finite verb is fronted, nothing else may be fronted with it. The pre-AUX test therefore also gives a strong indication against VP constituency.

Semantic constituency criteria, on the other hand, 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 1992: 14). But Miller himself observes that this latter criterion is not reliable in the case of the VP.

In the particular case of Slovene, Bolta (1989: 151) shows that idioms cannot be used to argue for the VP constituent convincingly. But more importantly, the evidence from coreferential interpretation of Slovene pronouns, which crucially led Bolta (1989) to lean towards a structure with the VP, is strictly theory-internal. Assuming the postulations of GB's binding theory, she argues that the subject-object asymmetry in binding can only be explained if the subject *c-commands* the object but not the other way around. In other words, only an analysis with the VP node gives correct predictions (Bolta 1989: 155). However, this argument will not carry over if we adopt another view where binding principles are not based on phrase structure but rather on thematic roles (as in Williams 1994) or on the relative obliqueness of grammatical relations as developed within HPSG. See Pollard & Sag (1992, in press: chap. 6) for extensive argumentation in favor of HPSG's account of binding over that of GB.

We can conclude that the empirical evidence does not support the finite VP constituent. As for nonfinite verbal projections, their distribution is somewhat different. For instance, *control*[12] verbs take complements which are headed by nonfinite verbs (such as the infinitive in (11)) and always have controlled (unexpressed) subjects.

(11) Ana poskuša [brati knjigo]
Ann-NOM tries read-INF book-ACC
'Ann is trying to read the book.'

Unlike finite verbs (10c), nonfinite verbs can form phrases (constituents) according to the pre-AUX test in (12), i.e., the verb and its object may precede the auxiliary.

---

12 The term 'control' refers to the co-indexing relation between the unexpressed (understood) subject of the embedded nonfinite clause and some controller noun phrase in the matrix clause.

(12) [Brati knjigo] je poskušala Ana.
read-INF book-ACC AUX tried Ann-NOM
'Ann was trying to read the book.'

The position of the object is similarly unrestricted as it is in finite clauses. But in contrast to finite verbs, Slovene nonfinite verbs can never take overt subjects (Bolta 1986: 426). There are no comparable constructions to the English one in (13) in which *John* is in some analyses considered to be the subject of the infinitival clause (Radford 1988: 304).

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

I now turn to formulating the structure of the clause. Borsley (1984) argued for an analysis of English where the VPs differ from Ss merely in the feature +/-SUBJ. He proposed that VP be specified as -SUBJ (since it does not contain the subject), and S as +SUBJ (since it includes the subject). It is this observation that will be applied here — but while English VP and S need to be represented by two distinct nodes (as in Figure 2) because of the processes that involve the VP (examples in (7), p. 47), I argue that this is not so in Slovene.

The keynote here is the lack of 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 with respect to the subject (see Table I):

label	verb	overt subject
S	finite	+/−
VP	nonfinite	−

Table 1: S and VP in Slovene

• a finite verb takes either an overt subject (thus forming what we may provisionally[13] call S[+SUBJ]) or a null subject (forming S[-SUBJ]);

• a nonfinite verb can never take an overt subject (its projection is labelled 'VP').

Clearly, once we accept that the verb and the object do not form a syntactic unit, we do not have to account for the discontinuity in cases where they fail to appear adjacently in the clause as, for instance, in the sentence (2d) (see p. 118 above).

This sentence would be represented by the 'flat' tree in Figure 4 rather than by the 'hierarchical'one in Figure 5 which contains a discontinuous VP and crossed branches.

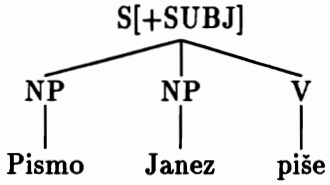

Figure 4: Representation without the VP

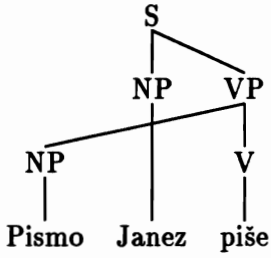

Figure 5: Representation with the VP

Having surveyed the data and proposed the clause structure, I will next introduce the relevant details of HPSG and then outline the model of the Slovene clause within this framework. HPSG is a suitable theory for present purposes because:

• it separates immediate constituency (what 'daughter' nodes may form a 'mother' node) from linear order (in what order the

[13] The labels S and VP serve here merely as abbreviations of structures to be made precise in the following section.

'daughters' have to appear), thus allowing for a simpler and more concise description of WO;[14]

- it is not confined to binary branching trees (as recent versions of GB are);

- it does not share GB's assumption of the *Projection Principle* (Radford 1988: 552). For example, this principle would imply that whenever the semantic representation is 'logically dyadic,' the syntactic representation must also have exactly two parts.

4. The HPSG Account of the Data

Head-driven Phrase Structure Grammar is an integrated theory of natural language syntax and semantics. The theory is *non-derivational* in that it does not employ transformational operations or movements between distinct levels of syntactic structure. It is *declarative* in the sense that an interaction of universal and language-specific constraints on linguistic expressions determines their admissibility in a particular language, independently of the order that they are applied in.

Linguistic entities are modelled by *sorted feature structures* (FSs). Informally, FSs are collections of attribute-value pairs. The sort of a particular FS determines what attributes (features) are appropriate for it. The finite set of sort symbols corresponds to the world of grammatical objects that we are modelling. The value of an attribute may be an embedded FS (possibly atomic, i.e., with no appropriate attributes) or a list of FSs.[15] Substructures in a given FS may be shared, that is, distinct attributes may have one and the same object as their value.

FSs are graphically depicted as *attribute-value matrix* diagrams which vertically summarize attribute names (in upper case) along with their values (to the right of them). Each FS is enclosed in square brackets with a small subscript on the left denoting its sort. Atomic values are shown in italicized lower case and lists of objects are enclosed in angle brackets. The diagram in Figure 6

---

14 In the transformational paradigm, too, it has been suggested on several occasions that the 'linearization rules' should be separated from 'constituency rules' (Radford 1988: 277). See Lightfoot (1979) for one such proposal.

15 Technically, a list can also be viewed as a special case of FS.

illustrates the architecture of a *sign* — the information associated with a word or a phrase.

A sign is organized in several top-level attributes which are internally structured, bringing together (under the same attribute) information that forms a natural class. Top-level attributes include the sign's phonology (PHON) which contains the orthographic representation of the phonetic content, and the sign's syntax-semantics (SYNSEM). The LOC (local) feature of the *synsem* object includes the categorial information (CAT) along with the semantic content (CONT). In this simplified case, the content is just the referential index which is used as a pointer; e.g., it can refer to some other phrase in the structure. A binary feature LEX is used to distinguish between lexical signs (entries from the lexicon), and nonlexical signs which represent phrasal projections.

HPSG relies crucially on rich lexical information. Within such a lexicon, much of the information is shared by many lexical entries. But this does not 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 & Sag 1987, chap. 8). Using these two mechanisms, we can capture linguistic generalizations and avoid redundancy in the lexicon.

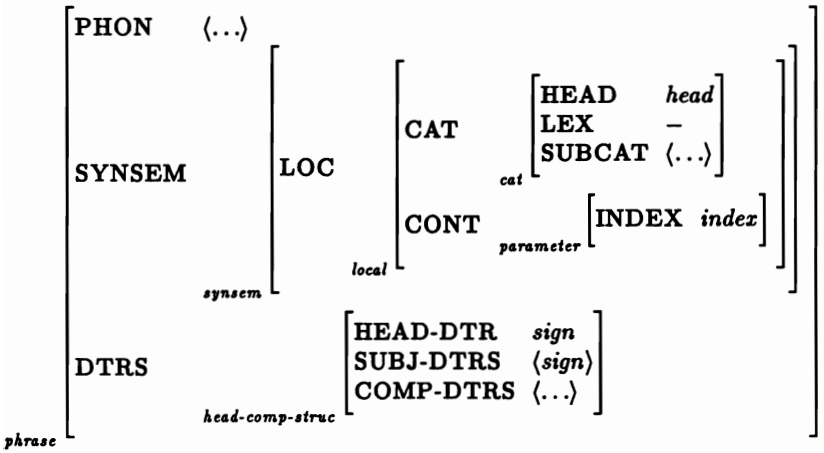

Figure 6: Architecture of a phrasal sign

A fundamental assumption of HPSG is that some signs act as *heads* and share certain features with their phrasal projections. These features are joined in the SYNSEM|LOC|CAT|HEAD[16] attribute. In the lexicon, heads are specified with their *subcategorization* requirements (SUBCAT) which indicate, what constituents the heads have to combine with (subcategorize for) to form phrasal projections. 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) immediate constituency.

Phrase structure is manifested by another top-level attribute that all phrasal signs are assumed to have, the DTRS (daughters) attribute. In Figure 6, the DTRS value includes the head daughter (HEAD-DTR) along with list-valued subject (SUBJ-DTRS) and complement (COMP-DTRS) daughters. We will not be concerned with other types of daughters in this paper. But unlike phrase structure rules, ID schemata do not impose the surface order in which the daughter constituents have to be realized in the utterance: this is accomplished by *linear precedence* constraints (Pollard & Sag 1987: 87).

As mentioned above, an important part of HPSG is the notion of *obliqueness* which is an ordering of grammatical relations shown in (14).

(14) SUBJECT < PRIMARY OBJECT < SECONDARY OBJECT < OTHER COMPLEMENTS

It is motivated cross-linguistically by several classes of generalizations: constituent order, binding, control, and the functioning of lexical rules (Sag & Pollard 1989: 166). Obliqueness is encoded in the SUBCAT list: less oblique arguments precede more oblique ones.

I will now sketch an HPSG analysis of the flat structure that was proposed in §3. With respect to subcategorization, I am assuming the revised approach, first proposed by Borsley (1987):

---

16 Such a sequence of attributes is termed a *path* and is used to refer to embedded attributes. We often use just the abbreviated feature names instead of full paths, e.g., HEAD stands for SYNSEM|LOC|CAT|HEAD. Also for brevity, FSs are not shown in full in the figures hereafter. For example, just the SYNSEM|LOC|CAT values may be shown instead of full *synsem* objects.

subjects are selected via the (singleton) SUBJ list, and complements by the COMPS list. In addition to the feature SUBCAT < 1, 2,... > we now obtain features SUBJ <1> and COMPS <2> (pairs of number tags indicate structure-sharing). In this way, the subject and the object are distinguished *lexically*, rather than *structurally*.

The SUBJ list of an 'ordinary'verb — one that subcategorizes for a nominative subject — contains a referential NP as its element for which the following holds[17] (cf. Table 1):

- (i) it may be dropped in the finite case; and
- (ii) it must not be overt in the nonfinite case.

Immediate constituency of phrases is in HPSG constrained by a universal principle specifying a disjunction of available ID schemata, from among which each language makes a selection. The flat clause will be accounted for by the following schema which combines a head daughter with a subject daughter and (possibly several) complement daughters. The schema in (15), a variant of Schema 3 (Pollard & Sag, in press, chap. 9) states that one of the options for a well-formed phrase is:

(15) **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*.

An instance of the application of this schema is depicted in Figure 7. It constructs a phrasal mother node from a head (H) daughter, a subject (S) daughter (as elaborated below, the subject in this figure is optional), and several complement (C) daughters.

Some comments are in order here. First, the schema includes a language-specific parochial condition that the head be a main verb, i.e., of sort *mainv* (and not an auxiliary).[18] Following Pollard & Sag (in press, chap. 9) I assume that the sort *phrase* incorporates [COMPS< >] specification. In the figures below, the labels S

17 The so-called 'logical subjects' (Toporišič 1984: 477) which are not in the nominative case would be regarded as least oblique complements. Their corresponding verbs, such as *zebsti* 'to be cold,' whose logical subject is in the genitive, or *sanjati se* 'to dream,' which takes a dative logical subject, would subcategorize for an expletive (nonreferential) subject. See also Orešnik 1992: 111ff..

18 A brief proposal on how auxiliaries might be treated within such a model is given in Kodrič (1993).

and VP from Table 1 will both correspond to *phrase* [LEX -, HEAD *mainv*] — note that in the HPSG context, labels such as NP[nom] serve as abbreviations for FSs.

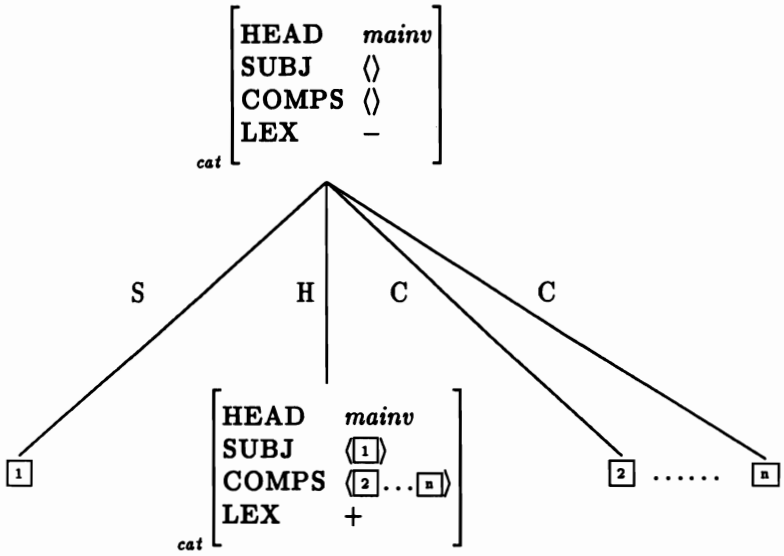

Figure 7: A structure licensed by Schema 3

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 imposes no explicit constraints on the SUBJ attribute. The *Valence Principle*, whose task is to cancel the realized subcategorization requirements, 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.[19] Figure 8 shows the construction of the sentence (3) by Schema 3 in which the mother sign retains a non-empty SUBJ list (recall that the COMPS list on the

19 Subject-verb agreement, as worked out by Pollard & Sag (in press, chap. 2) is not affected by this assumption. Both overt and null subjects are treated in the same fashion.

mother, on the other hand, is always empty due to the above-mentioned constraint on the sort phrase).

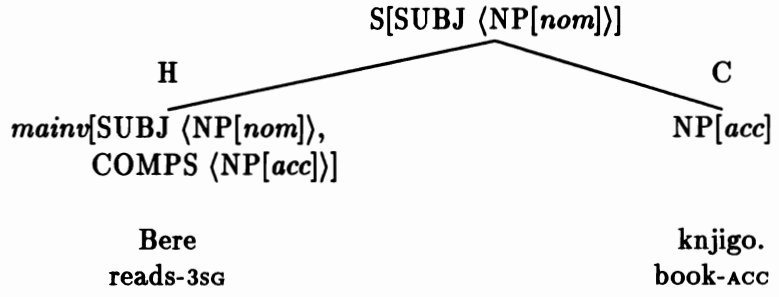

Figure 8: Sentence with a null subject

Next, I will assume that determining when the subject can be overt (see Table 1) is the task of an additional principle (16) which is associated with Schema 3 (since this is the only schema involving a subject daughter.)[20]

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). There is no empty constituent for the unexpressed subject — it is only present in the subcategorization frame.[21]

(16) Subject Realization Principle (language-specific)

a. A nonfinite head verb can not take an overt subject: the SUBJ list of the mother must be non-empty (representing the unexpressed subject)
b. If the head verb is finite, then
(i) if its subject is specified as expletive (see fn. 17), then the mother's SUBJ list must be non-empty;[22]
(ii) otherwise it depends on the context[23] whether or not a null (referential) subject is allowed (i.e., a non-empty SUBJ list of the mother).

20 The HEAD-SUBJ schema (which combines a VP with the subject in English) would not be needed for Slovene.

21 This treatment of the unrealized subject in Slovene is consistent with the one proposed for the unexpressed subject in English (Pollard & Sag, in press, chap. 3). It is also somewhat similar to the account of unrealized subjects based on semantic roles proposed by Orešnik (1992: 67 ff.)

22 Because, as noted by Bolta (1989: 149), Slovene expletives are always phonetically empty.

23 Deciding exactly in what circumstances the subject can be recovered from the context is a difficult problem (somewhat similar to anaphora resolu-

Finally, the controlled (unexpressed) subject is a consequence of co-indexing which takes place in control verbs. The categorial information of the control verb *poskusa* 'tries' is depicted in Figure 9.

The control verb subcategorizes for a nonfinite verb phrase complement which is required to have a single noun phrase in its SUBJ list. The referential indices of the embedded verb's subject and the control verb's subject are structure-shared: the two subjects are co-indexed.

Note also that the co-indexing between the two subjects accounts for the well-known observation which holds for Slovene as well: in the so-called 'raising constructions,' the controller of an expletive subject can only be an expletive NP (Bolta 1986: 428). Consider, for example, the expletive subject associated with the verb *deževati* 'to rain' in (18) below:

$$\textit{cat}\left[\begin{array}{ll}\textbf{HEAD} & \textit{mainv}[\textit{finite}] \\ \textbf{SUBJ} & \langle \text{NP}[\textit{nom}, \text{INDEX } \boxed{1}] \rangle \\ \textbf{COMPS} & \left\langle \textit{cat}\left[\begin{array}{ll}\textbf{HEAD} & \textit{mainv}[\textit{nonfin}] \\ \textbf{SUBJ} & \langle \text{NP}[\text{INDEX } \boxed{1}] \rangle \\ \textbf{COMPS} & \langle\rangle \\ \textbf{LEX} & - \end{array}\right] \right\rangle \end{array}\right]$$

Figure 9: Control verb

(18) a. Začelo je deževati.
started-PSP,SG,NEU AUX-3SG rain-INF
'It started to rain.'
b. *Začela je dezevati.
started-PSP,SG,FEM AUX-3SG rain-INF

It conveys (via its controller) its agreement requirements (third person, singular, neuter) which are contained in the INDEX attribute on the raising verb *začelo*. That is why having the expletive on the verb's SUBJ list is better than, for instance, assuming an empty SUBJ list in such cases. In addition, a uniform view is

---

tion). We would require the representation of the previous discourse and an appropriate decision algorithm.

preserved: each verb has at least one syntactic dependent (either referential or expletive).

Let us illustrate the construction of a sentence with an example. Recall the sentence (11) (page 123 above). Lexical entries that correspond to (11) are shown in Figure 10 (which has been simplified for expository purposes).

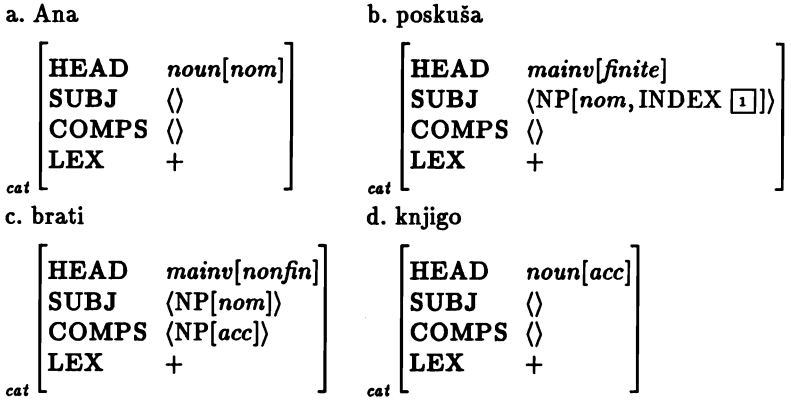

Figure 10: Sample lexical entries

First, consider the controlled complement. Since the verb *brati* is a lexical sign with the HEAD value of sort *mainv*, it can serve as the head daughter in Schema 3 (15). This schema will license a phrase of sort *head-subj-comp-struc* (see Figure 7). The Valence Principle ensures that the verb combines with its complement — the accusative NP *knjigo* — forming a nonfinite verb phrase *brati knjigo*. The Subject Realization Principle determines that an overt subject must not be present.

Second, a similar process constructs the entire sentence in a single step by taking the head finite verb *poskuša* and both its arguments: the nominative NP *Ana* is assigned the subject role and the nonfinite phrase *brati knjigo* serves as primary object. As a consequence of structure sharing, the phrase *Ana* is also interpreted as the understood subject of the embedded verb. We thus obtain the 'mother sign,' (shown as a FS in Figure 11, next page). This represents the entire sentence: a phrasal projection of the main verb which conforms to all universal and language-specific constraints.

A tree diagram of the sentence (12), in which Schema 3 is applied twice, is shown in Figure 12 (next page).

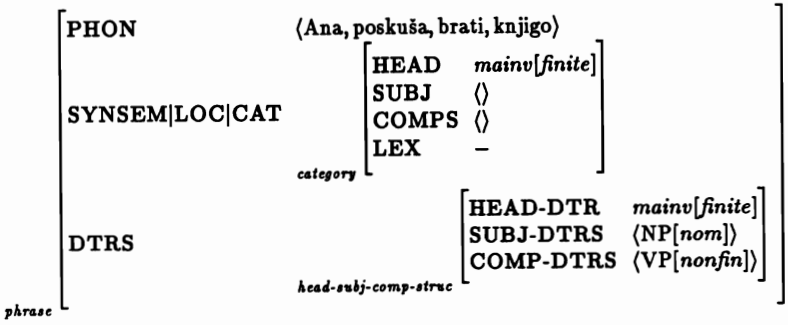

Figure 11: Well-formed sign representing the sentence

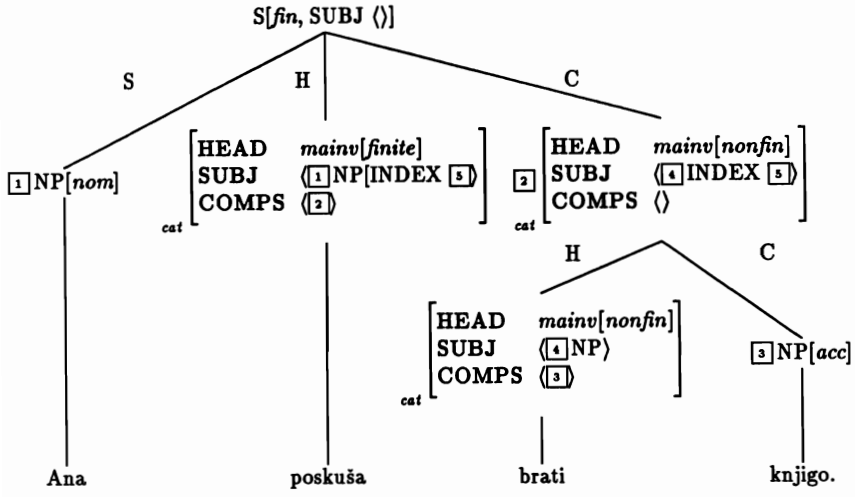

Figure 12: Sentence with a controlled complement

Linear order among the daughter nodes must be specified separately. Any ordering which is not excluded by some rule is admissible. All orderings will derive from the same ID rules and will not have any syntactic consequences. While I do not put forward a specific set of linear precedence constraints (or, more appropriately, tendencies) among the elements of the Slovene clause, note that we can obey the standard restrictions on ordering rules within the present approach (Siewierska 1988: 214):

- 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.

In this way, WO variation is viewed as a local phenomenon. Namely, the domain of locality in a 'flat' structure is the entire clause (since the verb and its arguments are sisters). To rule out unacceptable permutations, we would require an appropriate representation of the previous discourse. A fine-grained interaction of the 'fixed' WO, 'free' WO, and stylistics subtleties will be left for further research, perhaps along the lines of Uszkoreit (1986), by incorporating precedence rules of various weights (referring to notions like topic, focus, presupposition, etc., as well as prosodic factors). Markedness would then be determined according to the total weight of violated rules.

5. Conclusion

In summary, 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 nontransformational model of the Slovene clause. The flat clause structure gives a simpler account of syntax compared to the structure with separate VP and S nodes. This analysis fits the presented empirical facts which might pose problems for analyses under recent GB assumptions.

In HPSG, the Slovene clause is constructed by a variant of ID Schema 3 which accounts for both finite and nonfinite verbs and for the differences in their distributions. To cover other kinds of phrases, we would use other Schemata.[24]

---

[24] Namely: 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. — 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. In the following sentences, for example, the interpretation depends on the surface order of adverbials: *Frequently, the baby cried for an hour./For an hour, the baby cried frequently.* This matter was investigated recently by

The method proposed in this paper obeys the locality of ordering constraints and the standard relation between WO and phrase structure which is advantageous from the practical (computational) point of view. First, none of the major constituents is discontinuous in the sentence. Second, we do not assume any empty constituents (unrealized subjects are only present in the subcategorization frames but not in the constituent structure) which is also important for parsing efficiency. As a result, standard parsing algorithms may be used to process the grammar.

It has been shown that we can accommodate a significant degree of WO variation without having to appeal to transformations. 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. Results reported here are the basis of an ongoing research with the goal to set up a more complete formal description of the Slovene grammar that could serve as a foundation for computational implementation.

Inštitut Jožef Stefana, Univerza v Ljubljani

BIBLIOGRAPHY

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

Bolta, M. (1986). "Slovenski glagoli z nedolocniskim dopolnilom v TP-pristopu," *Slavistična revija* 36/4: 419-432.

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

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

-------. (1987). "Subjects and complements in HPSG," = Report CSLI-87-107, Center for the Study of Language and Information, Stanford CA.

Bresnan, J. (Ed.) (1982). *The Mental Representation of Grammatical Relations*. Cambridge MA: MIT Press.

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

------. (1986). *Barriers*. Cambridge MA: MIT Press.

Kasper (to appear) in the domain of German 'Mittelfeld'. He suggests that a compo-sitional treatment in the HPSG framework 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).

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.

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

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

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

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

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

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

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 CA: CSLI Publications.

Kodrič, S. (1993). *On the Constituent Structure of Slovene*. M.Sc. thesis, Simon Fraser University.

Lightfoot, D. (1979). *Principles of Diachronic Syntax*. Cambridge: CUP.

Marácz, L. & P. Muysken (eds.) (1989). *Configurationality: The Typology of Asymmetries*. Dordrecht: Foris.

Matthews, P.H. (1981). *Syntax*. Cambridge: CUP.

Mežnarič, J. (1992). *Clitics and Wackernagel position in Slovenian*. M.Sc. thesis, University of Edinburgh.

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

Orešnik, J. (1992). *Udelezenske vloge v slovenščini*. Ljubljana: SAZU.

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

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

------- & -------. (1992). Anaphors in English and the scope of binding theory. *Linguistic Inquiry* 23/2, 261-303.

------- & -------. (in press). *Head-driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

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

Reape, M. (1990). "Getting things in order," 125-37 in *Proceedings of the Symposium on Discontinuous Constituency*, Tilburg University. 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.

Rudin, C. (1985). *Aspects of Bulgarian Syntax: Complementizers and WH Constructions*. Columbus OH: Slavica.

Sag, I.A. and C.J. Pollard (1989). "Subcategorization and head-driven phrase structure," 139-81 in M.R. Baltin and A.S. Kroch (Eds.), *Alternative Conceptions of Phrase Structure*. Chicago IL: U. of Chicago Press.

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

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

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

Toman, J. (1986). "Cliticization from NPs in Czech and comparable phenomena in French and Italian," 123-45 in H. Borer (ed.), *The Syntax of Pronominal Clitics*. Orlando: Academic Press.
Toporišič, J. (1984). *Slovenska slovnica* (2nd ed.) Maribor: Obzorja.
Uszkoreit, H. (1986). "Constraints on order," *Linguistics* 24: 883-906.
Williams, E. (1994). *Thematic structure in syntax*. Cambridge MA: MIT Press.

POVZETEK

SESTAVNIK GLAGOLSKE ZVEZE IN STAVČNA ZGRADBA V SLOVENŠČINI

Slovenski jezik dovoljuje precej svobodnejši besedni red v stavkih, ki ga ima angleščina. Da bi take jezike lahko računalniško obravnavali, potrebujemo formalni slovnicni model, ki je ustrezen tako z jezikoslovnega kot tudi z računalniškega vidika. Razprava utemeljuje predlog, da je slovenski stavek primerneje opisati z nekoliko 'sploščeno' (nekonfiguracijsko) skladenjsko zgradbo, kot pa s tradicionalno hierarhično (konfiguracijsko). Na podlagi preverjanj je opuščena predpostavka o obveznem obstoju klasičnega sestavnika glagolske zveze (VP). Tako dobimo zgradbo, v kateri se glagolska oblika (osebna ali neosebna) povezuje z dopolnili (osebek, predmeti) v enem samem koraku. Nekaj vidikov tovrstne analize stavka je predstavljenih v okviru nepretvorbene frazne slovnice HPSG (Head-driven Phrase Structure Grammar).