**Arbeitspapiere des SFB 340**
Bericht Nr. 132, September 1999

# AN RSRL FORMALIZATION OF SERBO-CROATIAN SECOND POSITION CLITIC PLACEMENT

GERALD PENN

## Contents

1. Introduction .......... 177
2. Clitic Placement in Serbo-Croatian .......... 178
3. Assumptions about Prosody .......... 182
4. The Structure of Domain Lists .......... 183
5. Topological Fields .......... 186
6. Putting it Together: A Small Example .......... 191
7. Appendix 1: Topological Principles .......... 192
8. Appendix 2: Topological Relations .......... 194
References .......... 196

## 1. Introduction

This paper has three main purposes. The first is to generalize the work by Reape 1994, Kathol 1995 and others on domain-based treatments of free verb-phrase-level and clause-level word order to a framework that can express the embedding of word-order domains inside larger domains, and that can deal with multiple kinds of word-order domains in the same

ISSN 0947-6954/99

grammar. The second is to provide a different formalization of word-order domains in the logic of typed feature structures that allows them to account not only for cases of word order that are under-determined by syntactic principles, but also for cases where principles from different levels of linguistic structure interact mutually to constrain the surface realization of a grammatical string. Such is the case in Serbo-Croatian where prosodic constituency plays a role on a par with syntactic constituency in determining the allowable placements of second-position (2P) clitics. The third is to sketch a formal theory of Serbo-Croatian 2P clitic placement using generalized word-order domains in RSRL (Richter, Sailer, and Penn 1999; Richter forthcoming), a relational extension of feature logic for expressing linguistic principles on the feature structures that license utterances, to provide the first analysis of a construction in HPSG that integrates both prosodic and syntactic knowledge.

## 2. Clitic Placement in Serbo-Croatian

Serbo-Croatian is often called a free word-order language; but it is not entirely free. Phrasal constituents generally have a fixed, contiguous order that they must adhere to for a sentence to be grammatical; and the catalog of exceptions, e.g., the free ordering of daughters of verbal projections, the dislocation of prosodically heavy post-modifiers, is, for the most part, unsurprising.

One very interesting exception is the distribution of a class of enclitic particles known as *second-position* (2P) clitics. These clitics attach to a word or phrase at the beginning of their clause. What the host word or words (the "first" position) can be is constrained by several prosodic and syntactic factors, the net effect of which is frequently classified in the literature, e.g., Browne 1974, Zec and Inkelas 1990, Halpern 1995, as one of either *2D placement* (to use Halpern 1995's terminology), where the first element is a syntactic daughter of the clitic's clause, or *2W placement*, where the first element is a prosodic word, possibly disrupting the contiguity of a syntactic constituent:

(1) a. Taj čovek je video Mariju. (2D)
that man cl.3S saw Marija-ACC
*That man saw Mary.*
b. Taj je čovek video Mariju. (2W)

In (1a), the clitic *je* is placed after the first syntactic daughter, *Taj čovek*, which is the subject of the sentence. That NP consists of two prosodic words, *Taj* and *čovek*, so that the 2W placement in (1b) makes the NP discontinuous. All 2P clitics in Serbo-Croatian are capable of both placements, in spite of the fact that some are auxiliaries, some are personal pronouns, and one (*li*) is an interrogative particle. When two or more 2P clitics occur in the same clause, they occur in a cluster in a fixed order, as noted by Browne 1974.

(2) a. Taj pesnik mi je napisao knjigu. (2D)
that poet cl.1S.DAT cl.3S wrote book-ACC
That poet wrote me a book.
b.*Taj pesnik je mi napisao knjigu.
c. Taj mi je pesnik napisao knjigu. (2W)
d.*Taj je mi pesnik napisao knjigu.

The general order is (1) *li*, then (2) the auxiliaries except the third singular *je*, (3) dative pronominals, (4) genitive pronominals, (5) accusative pronominals and the reflexive *se*, and (6) auxiliary *je*. There are no nominative pronominal clitics; Serbo-Croatian is a pro-drop language.

The binary classification of 2P clitic placement is widely accepted by Slavicists as a workable idealization. What is not as widely accepted is the precise role of syntax and prosody in determining placement, even with respect to this classification. Much of the transformational work on this subject has predictably been devoted to determining the position of these clitics at surface structure, which is non-trivial, given sentences such as (3b):

(3) a. U lepi grad je stigao. (2D)
in beautiful city cl.3S arrived
*He has arrived in the beautiful city.*
b. U lepi je grad stigao. (2W)

Serbo-Croatian has another class of clitics which are prepositions proclitic to the first word of their NP objects. In (3), *u* is such a prepositional proclitic. It attaches to *lepi*, so that *u lepi* is the first prosodic word, although it is not traditionally regarded as a syntactic constituent.

There is now at least a general agreement that prosody does in fact play some role in 2W placement. It is, nevertheless, often assumed or concluded, notably by Halpern 1995, which reviews all of the major proposals up to 1995, that the surface position of Serbo-Croatian 2P clitics must be either Comp or a distinguished clitic position adjacent to Comp, such as left-adjoined to IP, with 2W placement being achieved in sentences like (3b) by a prosodic repair operation that dislocates the clitic to its surface-realized position. Several recent papers, e.g., Ćavar 1997, Penn 1998, Penn 1999, have argued against the adequacy of post-syntactic repair by prosody. The last two have done so along the way to making a more general point about the need for a separate, linear level of constituency that can mediate the interaction of the different kinds of constituency or structure at work in determining the linear realization of a phonological string. This point echoes earlier misgivings about analyzing syntactic structure with only a single flavor of constituency, e.g., Zwicky 1986, Blevins 1994, Dowty 1996, and ultimately dates back to Curry 1961's call for a distinction between *tectogrammatical* structure, where semantic interpretations are constructed compositionally, and *phenogrammatical* structure, which is concerned with the overt, linear realization of words in a string.

That call is also what has inspired recent work in HPSG, most notably Kathol 1995, on combining lists of domain objects, abstractions of linear substrings, with topological fields, abstract classes of linear order relative to linearly defined regions, to account for word-order phenomena such as scrambling and verb position in German clauses. In that research program, however, topological fields and domain lists were used to state principles on a syntactically under-determined problem (hence the relatively free order witnessed in German), in a more elegant way than traditional syntactic tree structures plus movement could provide.

The claim pursued here is that Curry 1961's level of phenogrammatical structure, of which domain lists and topological fields constitute one reasonable formalization, can also serve as the mediating level of structure in which prosodic, syntactic and other constraints must be reconciled in order for a string to be grammatical. In simpler terms, the domain objects of Kathol 1995 and the assignment of topological fields to them can be constrained by principles from not only tectogrammar but also prosodic structure and discourse structure. Grammatical strings must have an assignment of domain objects and topological fields that is simultaneously consistent with the principles from each of these sources, although each source can have its own additional abstract representation of structure, such as *synsem* structures in the case of syntax.

This application of domain lists is particularly useful in making sense of the cases where the 2W vs. 2D classification of clitic placement breaks down. There are certain syntactically defined combinations of words in which clitics cannot intervene, called *fortresses* by Halpern 1995, even if that position would be 2W, i.e., immediately following the first prosodic word. One of them is postnominal genitive modifiers:[1]

(4) a. Roditelji uspešnih studenata su se razišli.
parents successful-GEN students-GEN cl.3P cl.REFL dispersed
*The parents of the successful students dispersed.*
b.*Roditelji su se uspešnih studenata razišli.

*Roditelji* is the first prosodic word; but the clitic cluster, *su se* cannot intervene between it and its genitive plural postmodifier, *uspešnih studenata.*

Another case is the grammatical placement of a clitic within the daughter of a verbal projection in a position other than 2W:[2]

(5) a. Na čiji povratak iz Afrike je Ivan čekao? (2D)
on whose return from Africa cl.3S Ivan waited
*On whose return from Africa was Ivan waiting?*
b. Na čiji je povratak iz Afrike Ivan čekao? (2W)
c. Na čiji povratak je iz Afrike Ivan čekao?

(5c) is grammatical, even though *iz Afrike* modifies *Na čiji povratak.*[3] 2P clitics, however, cannot cause discontinuities within the modifier phrases.

(6) a. U radnji u lepom gradu je kupio čamac. (2D)
in shop in beautiful city cl.3S bought boat
*He bought the boat in (a) shop in (the) beautiful city.*
b. U radnji je u lepom gradu kupio čamac. (2W)
c.*U radnji u lepom je gradu kupio čamac.

A more accurate generalization of the distribution of 2P clitics is a restriction on interrupting prosodic words plus a syntactic "one-level-deep rule,"[4] that clitics may only intervene one phrasal level deep within the daughters of verbal projections. That, of course, prompts the question as to what the precise phrasal structure of these daughters is; and the

[1]The example is taken from Progovac 1996.

[2]The example was provided by Damir Ćavar, p.c.

[3]Halpern 1995 claims that similar examples of 2P clitics separating nouns from their PP post-modifiers are ungrammatical, but cites only adjunct PPs. These are often awkward because of an attachment ambiguity, not because of a syntactic prohibition.

[4]Wayles Browne, p.c.

answer is not entirely straightforward. For example, with a related discontinuous construction, it does not appear that sequences of prenominal adjectives behave as though nested in separate NP or $\bar{\mathrm{N}}$ projections in this respect.[5] For the purposes of this paper, however, it will suffice to accept this generalization at face value.

## 3. Assumptions about Prosody

For expository reasons, I assume that signs have the feature geometry shown in Figure 1. They have an extra feature, PROSODY, appropriate

$$\begin{bmatrix} \text{sign} & \\ \textsc{synsem} & \begin{bmatrix} \text{synsem} \\ \textsc{local}\ \ \text{local} \\ \textsc{nonlocal}\ \ \text{nonlocal} \end{bmatrix} \\ \textsc{qstore}\ \ \text{set(quantifier)} & \\ \textsc{qretr}\ \ \text{list(quantifier)} & \\ \textsc{dtrs}\ \ \text{con\_struc} & \\ \textsc{dom}\ \ \text{list(dom\_obj)} & \\ \textsc{prosody}\ \ \text{list(prosodic\_word)} & \end{bmatrix}$$

Figure 1. Part of the structure of *sign*.

to a list of prosodic words, which represent basic prosodic constituency. For every syntactic sign, the value of this feature contains the prosodic words that correspond to the yield of that sign. Because I assume that all prosodic constituents can ultimately be derived from subconstituents that are no larger than syntactic words, this list is well-defined, i.e., it is always possible to find constituents small enough to cover the yield of the current syntactic sign exactly. I will not delve any further into prosody here than the level of syntactic words themselves.

While it is clearly outside the scope of this paper to formulate an entire theory of prosodic structure, a few points deserve to be made about it here. First, it is entirely unsatisfying to represent prosodic structure as a disconnected list of prosodic words to be carried around in an otherwise syntactic derivation. Ultimately, the notion of sign in HPSG must be changed to allow parallel derivations of prosody, syntax and discourse that mutually constrain each other through a common DOM list. This

[5] Zlatić 1997, Penn 1998 and Penn 1999 discuss this discontinuous construction in more detail.

would involve, among other things, adding respective DTRS features for each derivation, with appropriate *constituent-structure* objects containing not more signs, but independent syntactic, i.e., *synsem*, or prosodic structures, respectively.[6] The representation of prosodic structure here is only adopted for the purposes of illustrating a more general point.

Second, I follow Inkelas 1989 in assuming that clitics are characterized by their possession of prosodic subcategorization requirements. Within *prosodic_word*, I therefore assume the existence of a PSUBCAT feature, whose value is a list of suitable frames of prosodic information that satisfy these requirements. The only contribution to a theory of prosody that I am, at present, prepared to provide is a Prosodic Subcategorization Principle that discharges these requirements to form larger prosodic words,[7] in the case of Serbo-Croatian clitics, or prosodic phrases in some other languages, much like its syntactic counterpart. A *prosodic_word* will also need to bear a domain list of its own, PDOM.

Third, there must be a principle that requires the list of domain objects in the DOM feature of the sign itself to be the result of appending the lists of domain objects in the list of prosodic words. The fact that these are the same is what requires the results of prosodic and syntactic constraints on a grammatical utterance to be consistent.

## 4. The Structure of Domain Lists

To control the freedom of word-order variation, Reape 1994 and Kathol 1995 use *compaction*, the combination of domain objects into a single domain object whose phonology has a fixed order and must be contiguously realized in the domain lists of mother signs. Compaction is also used in the present analysis to enforce contiguity among the daughters of verbal projections, and, as a special case of that, to control the placement of 2P clitics within a clause.

[6] This will mean that signs themselves are not recursive structures. With certain subordinate clauses, one could perhaps make a case for treating them as embedded signs, as in the case of German *denn* clauses and verb-second *weil* clauses (Frank Richter, p.c.); but it is not at all beyond question that a theory of "sub-utterance events" is warranted in the general case.

[7] Inkelas 1989 also claims that Serbo-Croatian clitics, in contrast to those of some other languages, subcategorize for prosodic words to create new prosodic words, not prosodic phrases.

Reape 1994 and Kathol 1995 use a representation of domain lists in which compaction is encoded by placing a *sign* or *dom(ain)_obj(ect)* on the list that contains a list of compacted structures from a daughter instead of the structures themselves. Daughter domain lists and mother domain lists are constrained by relations that enforce the appropriate compaction or shuffling of structures. Extending this to a formulation that allows prosodic and syntactic constraints to interact is no easy matter. Figure 2, for example, shows potential syntactic and prosodic compaction require-

⟨[u lepi],grad]⟩
Prosody: prosodic words compact
⟨¡[u],[lepi grad]⟩
Syntax: non-initial NPs compact

FIGURE 2. A prosodic and syntactic compaction on a non-initial PP.

ments on a non-sentence-initial PP that must combine. A non-initial NP normally must compact; but, on the other hand, the preposition *u* is a proclitic, and so it compacts with the adjective, *lepi*, to form a prosodic word. Unfortunately, the two resulting lists are inconsistent in feature logic if the compacted intervals are represented as single feature structures of any kind, including those of Reape 1994 and Kathol 1995, so many more relations would be necessary. In order not only to compact, but to merge compactions from different sources in a convenient way, we need a different representation.

One representation is shown in Figure 3. The type *rf* is a topological field marker, and will be introduced shortly. Here, the domain list contains one domain object for each word, but compaction is represented by the structure-sharing of a C(OMPACTION) feature. Using this distributed representation of phenogrammatical constituency, the process of adding compaction information to a list of domain objects becomes monotonic. In this case, a consistent domain list would compact the entire non-initial PP, as shown in the figure.

Every *dom_obj* has a PHON(OLOGY), to represent the phonological realization of its word. The order in which several phonological strings are realized is determined by the order in which their domain objects occur in the list itself. To enforce compaction, there is a principle on every domain list that no two domain objects with the same C value can have a domain

$$\left\langle \begin{bmatrix} \text{dom\_obj} \\ \textsc{phon}\ \text{u} \\ \textsc{c}\ \boxed{0}\ \text{rf} \end{bmatrix}, \begin{bmatrix} \text{dom\_obj} \\ \textsc{phon}\ \text{lepi} \\ \textsc{c}\ \boxed{0} \end{bmatrix}, \begin{bmatrix} \text{dom\_obj} \\ \textsc{phon}\ \text{grad} \\ \textsc{c}\ \text{rf} \end{bmatrix} \right\rangle$$

Prosody: prosodic words compact

$$\left\langle \begin{bmatrix} \text{dom\_obj} \\ \textsc{phon}\ \text{u} \\ \textsc{c}\ \text{rf} \end{bmatrix}, \begin{bmatrix} \text{dom\_obj} \\ \textsc{phon}\ \text{lepi} \\ \textsc{c}\ \boxed{0}\ \text{rf} \end{bmatrix}, \begin{bmatrix} \text{dom\_obj} \\ \textsc{phon}\ \text{grad} \\ \textsc{c}\ \boxed{0} \end{bmatrix} \right\rangle$$

Syntax: non-initial NPs compact

$$\left\langle \begin{bmatrix} \text{dom\_obj} \\ \textsc{phon}\ \text{u} \\ \textsc{c}\ \boxed{0}\ \text{rf} \end{bmatrix}, \begin{bmatrix} \text{dom\_obj} \\ \textsc{phon}\ \text{lepi} \\ \textsc{c}\ \boxed{0} \end{bmatrix}, \begin{bmatrix} \text{dom\_obj} \\ \textsc{phon}\ \text{grad} \\ \textsc{c}\ \boxed{0} \end{bmatrix} \right\rangle$$

Consistent domain list

FIGURE 3. A distributed representation of compaction.

object between them with a different C value (see the topological principle, Planarity, below). The equivalent of Reape 1994's and Kathol 1995's compacted domain objects is now a sequence of domain- objects with a shared feature, which cannot, and should not, bear *sign*- or *synsem*-typed values, because phenogrammatical constituents in this account sometimes correspond to prosodic words that cannot be assigned syntactic categories or certain other syntactic properties in a meaningful way. Penn 1999 argues for the elimination of this information from domain objects at more length.

Using this representation, an example of the distribution of different domain lists in a sign is given in Figure 4. Normal syntactic constituency

$$\begin{bmatrix} \text{phrase} & \\ \textsc{synsem} & \textsc{pp} \\ \textsc{dom} & \left\langle \boxed{1}\begin{bmatrix} \text{dom\_obj} \\ \textsc{phon}\ \text{u} \\ \textsc{c}\ \boxed{0}\ \text{rf} \end{bmatrix}, \boxed{2}\begin{bmatrix} \text{dom\_obj} \\ \textsc{phon}\ \text{lepi} \\ \textsc{c}\ \boxed{0} \end{bmatrix}, \boxed{3}\begin{bmatrix} \text{dom\_obj} \\ \textsc{phon}\ \text{grad} \\ \textsc{c}\ \boxed{0} \end{bmatrix} \right\rangle \\ \textsc{prosody} & \left\langle \begin{bmatrix} \text{prosodic\_word} \\ \textsc{pdom}\ \langle \boxed{1},\boxed{2} \rangle \end{bmatrix}, \begin{bmatrix} \text{prosodic\_word} \\ \textsc{pdom}\ \langle \boxed{3} \rangle \end{bmatrix} \right\rangle \\ \textsc{dtrs} & \begin{bmatrix} \textsc{hdtr} & \begin{bmatrix} \textsc{synsem}\ \textsc{p} \\ \textsc{dom}\ \langle \boxed{1} \rangle \end{bmatrix} \\ \textsc{cdtrs} & \left\langle \begin{bmatrix} \textsc{synsem}\ \textsc{np} \\ \textsc{dom}\ \langle \boxed{2},\boxed{3} \rangle \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

FIGURE 4. A *sign* object with distributed compaction.

is indicated in the DTRS value; prosodic constituency is indicated in the PROSODY value. All of the different representations use the same domain objects to express constraints on word order. The combination of those constraints is, thus, reflected in the DOM list of a *sign*.

## 5. Topological Fields

Following Kathol 1995, I also assume the existence of *topological fields*. Topological fields are often defined as classes of domain objects over which linear precedence constraints are stated. This is not quite correct. They are classes of domain objects within a particular *region* over which linear precedence constraints are stated for that region. The ones that Kathol 1995 uses are for the region of a clause consisting of a verb and its complements and adjuncts. To illustrate how multiple regions interact, we shall consider this one plus the region of 2P clitics, with the topology given by Browne 1974. One would need yet another set of topological fields, for example, to describe the order of domain objects inside an NP. When regions are embedded inside other regions, as with the clitic region, NPs, PPs or embedded clauses in the verbal case, a new assignment of fields must be considered. By convention, membership in one of these classes is indicated with a TOPO attribute on domain objects appropriate to one of a language-specific collection of topological field objects for a particular region. Often, these regions are defined syntactically, but that does not need to be the case in general.

In the case of Kathol 1995's treatment of compaction, the compacting relation can assign a new field to the compacted structure in a higher region, if it requires one. For the distributed representation used here, a simple change of "assignment" is not possible because the same domain objects (as feature structures) are used in every region.

In the present analysis, instead of using another attribute, we can use the C attribute itself, with structure-sharing indicating compaction as before, and with the type of the C value bearing the topological field, such as *rf* in Figure 3. We can also explicitly represent the region in which the topological field is defined so that the field itself may persist in the domain object. To do this, every topological field except one must have an appropriate feature, R(EGION), which is also appropriate to a topological field. The one exception is the topological field, *matrix*, which has no appropriate features. In theory, one could use topological fields to order sentences within a larger discourse-defined region, in which case *matrix* would also have an R feature. As a result, every domain object has a string of topological field assignments along a path of R features, ending in *matrix*. For every sentence, there is only one matrix region, so this *matrix* object occurs on every C/R path of every domain object in a sentence (see the topological principle, Matrix Compaction, below). Figure 5 shows the domain list for (7).

$$\left\langle \begin{bmatrix} \text{dom\_obj} \\ \textsc{phon}\ \text{Ivan} \\ \textsc{c}\ \begin{bmatrix} \text{pre-cf} \\ \textsc{r}\ \boxed{0}\ \text{matrix} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{dom\_obj} \\ \textsc{phon}\ \text{mi} \\ \textsc{c}\ \begin{bmatrix} \text{dat-cf} \\ \textsc{r}\ \boxed{1} \begin{bmatrix} \text{cf} \\ \textsc{r}\ \boxed{0} \end{bmatrix} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{dom\_obj} \\ \textsc{phon}\ \text{je} \\ \textsc{c}\ \begin{bmatrix} \text{je-cf} \\ \textsc{r}\ \boxed{1} \end{bmatrix} \end{bmatrix}, \right.$$

$$\left. \begin{bmatrix} \text{dom\_obj} \\ \textsc{phon}\ \text{rekao} \\ \textsc{c}\ \begin{bmatrix} \text{rf} \\ \textsc{r}\ \boxed{0} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{dom\_obj} \\ \textsc{phon}\ \text{da} \\ \textsc{c}\ \begin{bmatrix} \text{pre-cf} \\ \textsc{r}\ \boxed{2} \begin{bmatrix} \text{ef} \\ \textsc{r}\ \boxed{0} \end{bmatrix} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{dom\_obj} \\ \textsc{phon}\ \text{voli} \\ \textsc{c}\ \begin{bmatrix} \text{rf} \\ \textsc{r}\ \boxed{2} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{dom\_obj} \\ \textsc{phon}\ \text{Mariju} \\ \textsc{c}\ \begin{bmatrix} \text{rf} \\ \textsc{r}\ \boxed{2} \end{bmatrix} \end{bmatrix} \right\rangle$$

FIGURE 5. A simple sentence with an embedded clause.

(7) Ivan mi je rekao da voli Mariju.
Ivan cl.1S.DAT cl.3S told that loves Marija-ACC
*Ivan told me that he loves Mary.*

The topological fields themselves will be introduced shortly. For clarity, we can represent these structures as graphs, as shown in Figure 6. These

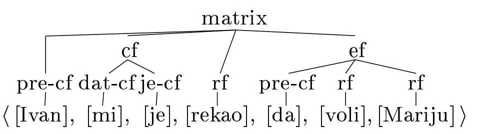


FIGURE 6. A graphical representation of a domain list.

look like trees; and they are. C values encode the derivations of linear, phenogrammatical constituents, just as phrase structure trees encode the derivations of syntactic constituents.

We need one more definition to state linear precedence constraints with topological fields. The natural question when stating linear precedence constraints is which topological field in the string of fields for a domain object should be used. Again, this does not have an absolute answer, because topological fields do not necessarily specify absolute positions or orderings in a domain list. Given two objects on a domain list, $D_1$ and $D_2$, the *relative topological fields* of $D_1$ and $D_2$ are the types of the deepest non-identical structures in their C values whose R values are identical. If their C values themselves are the same object, then their relative topological fields are the type of that object, i.e., the same field. Of course, we may also speak of the topological field of an object relative to a region. This is its topological field assignment relative to the other objects of that region.

Linear precedence is then enforced by requiring every $D_1$ and $D_2$ to be ordered in accordance with the ordering of their relative topological fields. In Figure 5 or 6, the relative topological fields of *da* and *voli* are *pre-cf* and *rf*, respectively. As will be seen shortly, *pre-cf* objects must appear before *rf* objects; so *da* must precede *voli*. The relative topological fields of *rekao* and *da* are *rf* and *ef*, respectively. *rf* objects precede *ef* objects; so *rekao* precedes *da*. Note that if we had used *da*'s *pre-cf* assignment, the reverse would have been predicted.

In the graph notation, the relative topological fields of domain objects are the highest projections of those domain objects that are distinct and have an identical parent, or, if their C values are the same, their common pre-terminal field. Viewed in this way, it can easily be seen that the topological fields plus compaction of Kathol 1995 are just a different way of writing an ID/LP grammar over phenogrammatical constituents. Because every domain object in Kathol 1995 has a SYNSEM value, they are also equivalent to a liberation-style ID/LP grammar over tectogrammatical constituents where intermediate verbal projections are liberated. With partial compaction, they are strictly stronger than ID/LP with liberation, even over phenogrammatical constituents, because one daughter can liberate while the rest of the local tree stays intact.

In general, this device is more powerful than classical ID/LP in terms of strong generative capacity, relative to the same set of topological field assignments. If one were to add different regions with their own topologies, then it would be possible to encode linear precedence principles relative only to a particular region. In classical ID/LP, on the other hand, every linear precedence rule applies to every immediate dominance rule. Of course, one can always expand the set of categories in an ID/LP grammar to compensate for this.

In the present analysis, the use of topological fields is also equivalent to a phenogrammatical ID/LP grammar; but the correspondence with tectogrammatical ID/LP does not exist because some phenogrammatical constituents are formed not on the basis of tectogrammatical compaction, but rather on the basis of prosodic compaction.

The structure of domain objects given here is proposed as a language-universal construct. The language-specific components of linear precedence are the specification of regions, each with a topology, the assignment of topological fields, and the compaction constraints that form phenogrammatical constituents. In more configurational languages, compaction occurs so prolifically among syntactic constituents that, when we restrict our attention to these languages, we see a much more direct correspondence between syntactic constituents and domain objects. This is why, in the case of German and English, domain objects with *synsem*- or *sign*-typed substructures have worked to the extent that they do.

The two regions considered in this paper for Serbo-Croatian, one for clauses, and one for 2P clitic clusters, are shown in Tables 1 and 2. These regions predict the placement of second-position clitics and embedded clauses within clauses, and the relative positions of clitics in clitic clusters, as described on p. 179. Each table shows the possible names for the

| Compacts to | matrix ∨ ef | | | | |
|---|---|---|---|---|---|
| Field | pre-cf | cf | post-cf | rf | ef |
| Occupancy | 1 | $\leq 1$ | $\leq 1$ | — | $\leq 1$ |

TABLE 1. Topology for clauses.

| Compacts to | cf | | | | | |
|---|---|---|---|---|---|---|
| Field | li-cf | aux-cf | dat-cf | gen-cf | acc-cf | je-cf |
| Occupancy | $\leq 1$ | $\leq 1$ | $\leq 1$ | $\leq 1$ | $\leq 1$ | $\leq 1$ |

TABLE 2. Topology for clitic clusters.

region within larger regions, i.e., what the appropriate R(EGION) values are for the topological field types, the names of the topological fields for the region, and what the occupancy requirements (per instance of region) are for each field, if any. Some fields, such as *pre-cf* are required to have exactly one object. This requirement on *pre-cf* is what makes 2P clitics "second-position." Others can require at most one or at least one object. Note that these are not requirements on the number of words that can occur in a field, but on the number of actual structures of the type corresponding to that field within a given region. The *pre-cf* field, for example, can be filled by two words, such as *u lepi*, but because they compact, the C values of the *dom_obj* objects for *u* and *lepi* are the same *pre-cf* object, which is unique for that region. *cf* can be filled by a cluster of clitics, but those clitics belong to a single *cf* region; so there is at most one *cf*-typed structure in the next higher clause region.

The topological fields defined for these regions are, for the most part, self-explanatory. The field, *post-cf*, is necessary to hold the remnant of a syntactic constituent that has been made discontinuous by 2W clitic placement. When a 2P clitic is placed after the first prosodic word, the rest of the syntactic constituent to which the first prosodic word belongs is constrained to occur either immediately after the 2P clitic(s), i.e., in the position marked by *post-cf*, or in certain other locations whose significance and availability are not completely understood. The other locations are, therefore, omitted from the present analysis.[8] *ef* is for embedded clauses. Note that embedded clauses can contain embedded clauses, so the clausal region is potentially recursive. The field, *rf*, is the remainder field, for everything else.

The parochial principles governing the clausal topological fields and their assignment are formalized in RSRL in Appendix 1, assuming the definitions of some basic relations, which are then defined in Appendix 2.

The principles, Matrix Compaction, Planarity and Relevance, ensure the integrity of the C values. Matrix Compaction guarantees the acyclicity of R paths, since *matrix* has no appropriate features,[9] and their convergence on a single object per utterance. Without Planarity, constituents in a multiply-fillable topological field (such as *rf*) could be made discontinuous by other occupants of that field. It is what also allows us to draw C values as planar graphs, as in Figure 6. Relevance is the price of using a constraint-based rather than deductive approach: without it, the other constraints would not preclude the existence of arbitrarily long R paths whose structures did not correspond to any particular region, since some of them, such as clauses, are necessarily recursive. The Relevance principle is also open to modification whenever new regions are added because those

[8]Some other locations are briefly discussed in Ćavar 1997. Seemingly the most common one, clause-final position, has been discussed at much greater length, e.g., Halpern 1995, Ćavar 1997, Zlatić 1997, Penn 1999.

[9]If discourse-level regions are added on top of *matrix*, then *matrix* would have an R feature, and another field would need to take its place in this constraint.

new regions may then license the existence of new structures. As it stands, it says that topological field structures are licensed either by the lexicon, by embedded clause compaction, by a clitic cluster (*cf* structures), or by an unembedded sign (*matrix* structures).

The other principles essentially encode the information presented in Tables 1 and 2. They do not constrain the internal structure of NPs or PPs, nor, crucially, the compaction of prepositional proclitics to the first prosodic word of their hosts. We eventually need to define topological regions for NPs and PPs in order to account for those facts. One also needs the usual extensions to the ID Schemata that shuffle domain lists of daughters together.

## 6. Putting it Together: A Small Example

Figure 7 shows a syntactic derivation for (3b), along with its compaction

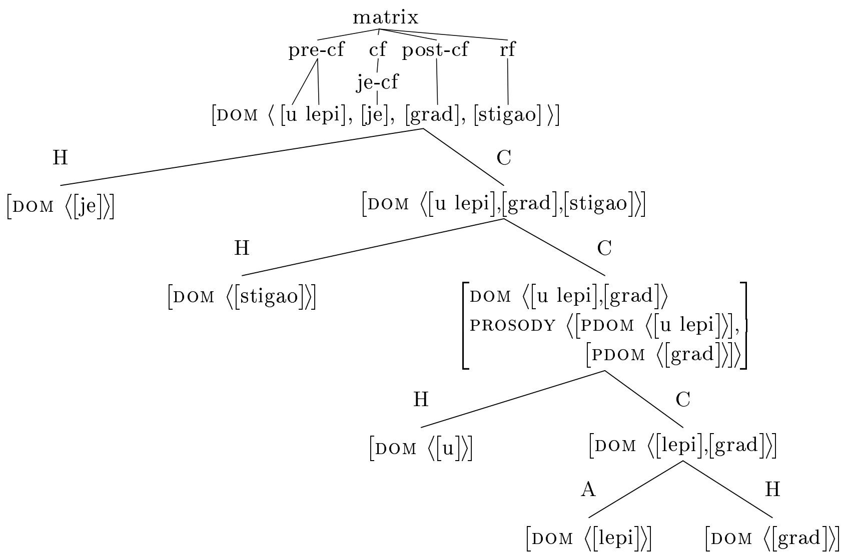


FIGURE 7. A derivation of (3b).

tree at the top. For brevity, only the DOM and, where significant, PROSODY values are shown. In domain lists, compaction is depicted in the more traditional way with bracketings for convenience. Of course, in reality, C values are used everywhere, and a compaction on two objects is visible in every list where both of those domain objects occur. As with all linearization-based analyses, the yield of the syntactic tree is not the phonology of the utterance. That information is only available from the DOM list. Here, the PP itself does not compact in our regions; but the PP's region requires a prosodic compaction in the clausal region.

## 7. Appendix 1: Topological Principles

In the principles and relations formulated here, conjunction has tighter precedence than disjunction, quantification has tighter precedence than conjunction, and implication has tighter precedence than quantification. $x\exists y.p$ means that $y$ is a component of $x$, i.e., accessible from $x$ by a (finite) path. It is a shorthand for $\exists y.(exist(x,y)\ \&\ p)$, where $\exists$ is RSRL's component quantification operator (Richter, Sailer, and Penn 1999). $x\forall y.p$ abbreviates $\neg(x\exists y.\neg p)$. $x \sim \tau$ is a type description. : denotes the object that a constraint describes.

1. **Topological Order:** In any domain list, domain objects are ordered by relative topological field in the order shown in Tables 1 and 2.
   $\textit{nelist(dom-obj)} \rightarrow \exists r.(region(:,r)\ \&\ topo\_order(:,r))$
2. **Field Existence:** Among the relative topological fields of every complete clause's domain list, there is at least one feature structure of type *pre-cf*.
   $(\textit{unembedded-sign} \vee \text{DTRS}{:}\textit{head-mark-struc}) \rightarrow$
   $\quad \exists ds.\exists r.(\text{DOM}{:}ds\ \&\ region(ds,r)\ \&$
   $\qquad ds\exists d.\exists f.(topo\_field(d,r,f)\ \&\ f \sim \textit{pre-cf}))$
3. **Field Uniqueness:** Among the relative topological fields of every complete clause's domain list, there is at most one feature structure of type *pre-cf*.[10]
   $(\textit{unembedded-sign} \vee \text{DTRS}{:}\textit{head-mark-struc}) \rightarrow$
   $\quad \exists ds.\exists r.(\text{DOM}{:}ds\ \&\ region(ds,r)\ \&$
   $\qquad ds\forall d_1.ds\forall d_2.\forall f_1.\forall f_2.$
   $\qquad\quad (d_1 \neq d_2\ \&\ topo\_field(d_1,r,f_1)\ \&$
   $\qquad\quad topo\_field(d_2,r,f_2)\ \&\ f_1 \sim \textit{pre-cf}\ \&\ f_2 \sim \textit{pre-cf})$
   $\qquad\quad \rightarrow f_1 = f_2)$

[10] Likewise for *cf, post-cf, ef, li-cf, aux-cf, dat-cf, gen-cf, acc-cf*, and *je-cf*. The formalizations for these are obtained by replacing occurrences of *pre-cf* in the one provided here with each of these fields in turn.

4. **2P Clitic Compaction:** In the domain list of any prosodic word, its initial domain object cannot have a field *cf*, relative to any clause; and if there is a domain object with field *cf* relative to a clause, then it compacts in the next higher region with the preceding *pre-cf* object, relative to that clause.
   $\mathit{prosodic\text{-}word} \rightarrow$
   $\exists ds.\exists c_1.(\textsc{pdom}{:}ds\ \&\ ds\textsc{first}{:}\textsc{c}{:}c_1\ \&$
   $c_1 \forall f_1 \sim \mathit{topo\_field}.\neg(f_1 \sim cf)\ \&$
   $\exists dsrest.(ds\textsc{rest}{:}dsrest\ \&$
   $dsrest \forall d \sim \mathit{dom\_obj}.\exists c.(d\textsc{c}{:}c\ \&$
   $c \forall f.f \sim cf \rightarrow c_1 \exists f_1'.(f_1' \sim \mathit{pre\text{-}cf}\ \&\ f_1'\textsc{r}{=}\ f\textsc{r}))))$
5. **Embedded Clause Compaction:** In the domain list of any head-marker phrase, there is a unique object of type *ef* that is the R value of the relative topological fields of every pair of domain objects in that domain list.
   $\textsc{dtrs}{:}\mathit{head\text{-}mark\text{-}struc} \rightarrow\ \exists ds.(\textsc{dom}{:}ds\ \&\ \exists r \sim \mathit{ef}.\mathit{region}(ds, r))$
6. **NP/PP Compaction:** For every NP and PP, either it compacts in its clause to *rf* (non-initial), compacts in its clause to *pre-cf* (2D placement), or does not compact, with its first prosodic word compacting to *pre-cf* and the rest compacting to *post-cf* (2W placement).
   $\textsc{synsem}{:}\textsc{local}{:}\textsc{cat}{:}(\textsc{head}{:}(\mathit{noun} \vee \mathit{prep})\ \&\ \textsc{subcat}{:}\langle\rangle) \rightarrow$
   $\exists ds.\exists r.(\textsc{dom}{:}ds\ \&\ \mathit{region}(ds, r)\ \&$
   $(r \sim \mathit{pre\text{-}cf}$
   $\vee r \sim \mathit{rf}$
   $\vee (r \sim \mathit{ef} \vee r \sim \mathit{matrix}\ )\ \&$
   $\exists ps.\exists pr.\exists r.(\textsc{prosody}{:}\textsc{first}{:}\textsc{pdom}{:}ps\ \&$
   $\mathit{region}(ps, pr)\ \&\ pr \sim \mathit{pre\text{-}cf}\ \&\ pr\textsc{r}{:}r\ \&$
   $\exists restr.(restr \sim \mathit{post\text{-}cf}\ \&\ restr\textsc{r}{:}r\ \&$
   $ds \forall d \sim \mathit{dom\_obj}.(ps \forall d'.d' \neq d) \rightarrow exist(d, restr)))))$
7. **Matrix Compaction:** There is a unique *matrix* object that appears in every C value of every domain object in any domain list.
   $\mathit{dom\text{-}obj} \rightarrow \exists y.y \sim matrix$
   $\mathit{unembedded\text{-}sign} \rightarrow \forall y \forall z.(y \sim matrix\ \&\ z \sim matrix) \rightarrow y = z$
8. **Planarity:** Given three domain objects on a domain list, the shortest R path at whose value the first (by order in the list) has a substructure of the second's C value must be a (not necessarily proper) prefix of the shortest path at whose value the first has a substructure of the third's, i.e., the graphical representations as in Figure 6 must be planar.$\forall d_1.\forall d_2.\forall d_3.\forall r.\forall f.$
   $(\mathit{nelist(dom\text{-}obj)}\ \&\ \textsc{first}{:}d_1\ \&\ \textsc{rest}{:}\textsc{first}{:}d_2\ \&$

$\text{REST:REST:FIRST:}d_3$ & $region(:, r)$ & $topo\_field(d_1, r, f)$ &
$topo\_field(d_3, r, f))$
$\rightarrow topo\_field(d_2, r, f)$

9. **Relevance:** In every domain list, for every domain object in that list, for every object, $X$, in its C value, there is another domain object such that their relative topological fields' R values are both $X$.
$sign \rightarrow$
$\forall f \sim field.(\exists s \sim word.s\exists d \sim dom\_obj.d\text{C}{:}f$
$\vee \exists s \sim phrase.\exists ds.(s\text{DTRS}{\sim}head\text{-}mark\text{-}struc$ &
$s\text{DOM}{:}ds$ & $region(ds, f)$ &
$\forall d.\exists c.(d\text{C}{:}c$ & $(c\exists f'.f' = f) \rightarrow ds\exists d'.d' = d))$
$\vee f \sim cf$
$\vee f \sim matrix)$

## 8. Appendix 2: Topological Relations

In relation definitions, a variable not explicitly quantified on the left hand side of the bi-implication is assumed to be universally quantified at the beginning of the formula.

- **rel_topo_fields**$(x,y,f_x,f_y)$: the relative topological fields of domain objects, $x$ and $y$, are $f_x$ and $f_y$, respectively.[11]
$rel\_topo\_fields(x, y, f_x, f_y) \leftrightarrow$
$x \sim dom\text{-}obj$ & $y \sim dom\text{-}obj$ &
$\exists c_x.\exists c_y.(x\text{C}{:}c_x$ & $y\text{C}{:}c_y$ &
$(c_x = c_y$ & $f_x = c_x$ & $f_y = c_y$
$\vee c_x \neq c_y$ & $c_x\exists f'_x.f_x = f'_x$ & $c_y\exists f'_y.f_y = f'_y$ & $f_x \neq f_y$ &
$\exists r.(f_x\text{R}{:}r$ & $f_y\text{R}{:}r)))$

- **topo_field**$(d,r,f)$: the relative topological field of domain object, $d$, with respect to region, $r$, is $f$.
$topo\_field(d, r, f) \leftrightarrow$
$d \sim dom\text{-}obj$ &
$\exists c.(d\text{C}{:}c$ & $(c = r$ & $f = c$
$\vee c \neq r$ & $c\exists f'.(f = f')$ & $f\text{R}{:}r))$

[11]This relation is not used by the principles. It is only provided to formalize the notion of relative topological fields given in the text.

- **region**$(ds,r)$: $r$ is the smallest region that contains the domain objects on the domain list, $ds$.
$region(ds,r) \leftrightarrow$
$ds \sim \mathit{nelist(dom\text{-}obj)}\ \&$
$\exists dsrest.\exists d.\exists c.(ds\mathrm{REST}{:}dsrest\ \&\ ds\mathrm{FIRST}{:}d\ \&\ d\mathrm{C}{:}c\ \&$
$region\_recurse(dsrest,d,c,r))$
$region\_recurse(ds,d_1,r_1,r) \leftrightarrow$
$(ds \sim elist\ \&\ r_1 = r)$
$\vee\exists d_2.\exists dsrest.\exists c_2.(ds\mathrm{FIRST}{:}d_2\ \&\ ds\mathrm{REST}{:}dsrest\ \&\ d_2\mathrm{C}{:}c_2\ \&$
$c_2\exists r_2.((r_2 = r_1$
$\vee c_2\forall x.x \neq r_1\ \&$
$r_1\exists f_1.c_2\exists f_2.(f_1 \neq f_2\ \&\ f_1\mathrm{R}{:}r_2\ \&\ f_2\mathrm{R}{:}r_2))\ \&$
$region\_recurse(dsrest,d_2,r_2,r)))$

- **topo_order**$(ds,r)$: the domain list, $ds$, respects the topological ordering defined on region, $r$.
$topo\_order(ds,r) \leftrightarrow$
$ds \sim elist$
$\vee\exists dsrest.\exists d.\exists f.(ds\mathrm{REST}{:}dsrest\ \&$
$topo\_order(dsrest,r)\ \&$
$ds\mathrm{FIRST}{:}d\ \&$
$topo\_field(d,r,f)\ \&$
$(f \sim cf \rightarrow dsrest\forall d_2 \sim dom\text{-}obj.$
$\exists d_2.\exists f_2.(topo\_field(d_2,r,f_2)\ \&\ \neg(f_2 \sim pre\text{-}cf)))\ \&$
$(f \sim post\text{-}cf \rightarrow dsrest\forall d_2 \sim dom\text{-}obj.$
$\exists d_2.\exists f_2.(topo\_field(d_2,r,f_2)\ \&$
$\neg(f_2 \sim pre\text{-}cf)\ \&\ \neg(f_2 \sim cf)))\ \&$
$(f \sim rf \rightarrow dsrest\forall d_2 \sim dom\text{-}obj.$
$\exists d_2.\exists f_2.(topo\_field(d_2,r,f_2)\ \&$
$\neg(f_2 \sim pre\text{-}cf)\ \&\ \neg(f_2 \sim cf)\ \&\ \neg(f_2 \sim post\text{-}cf)))\ \&$
$(f \sim ef \rightarrow dsrest\forall d_2 \sim dom\text{-}obj.$
$\exists d_2.\exists f_2.(topo\_field(d_2,r,f_2)\ \&$
$\neg(f_2 \sim pre\text{-}cf)\ \&\ \neg(f_2 \sim cf)\ \&\ \neg(f_2 \sim post\text{-}cf)\ \&$
$\neg(f_2 \sim rf))))$

- **exist**$(x,y)$: $y$ is a component of $x$.[12] $exist(x,y) \leftrightarrow$
$x \sim field\ \&\ y \sim field\ \&$

[12] There is no known, single method of expressing this over an arbitrary RSRL signature. What are presented here are just the necessary cases for the present analysis relative to its assumed signature.

$(x = y$
$\vee \exists z.(x \neq y \;\&\; x\textsc{R}z \;\&\; exist(z, y)))$
$exist(ds, d) \leftrightarrow$
$ds \sim \textit{nelist(dom-obj)} \;\&\; d \sim \textit{dom-obj} \;\&$
$(ds\textsc{first}{:}d$
$\vee \exists dsrest.(ds\textsc{rest}{:}dsrest \;\&\; exist(dsrest, d)))$
$exist(ps, p) \leftrightarrow$
$ps \sim \textit{nelist(prosodic-sign)} \;\&\; p \sim \textit{prosodic-sign} \;\&$
$(ps\textsc{first}{:}p$
$\vee \exists psrest.(ps\textsc{rest}{:}psrest \;\&\; exist(psrest, p)))$
$exist(s, f) \leftrightarrow$
$s \sim \textit{word} \;\&\; f \sim \textit{field} \;\&$
$\exists ds.(s\textsc{dom}{:}ds \;\&$
$ds \exists d.\exists c.(d\textsc{c}{:}c \;\&\; c \exists f'.f' = f)$
$\vee \exists ps.(s\textsc{prosody}{:}ps \;\&$
$ps \exists p.\exists ds.(p\textsc{pdom}{:}ds \;\&$
$ds \exists d.\exists c.(d\textsc{c}{:}c \;\&\; c \exists f'.f' = f))))$

## References


Blevins, J. (1994). Derived constituent order in unbounded dependency constructions. *Journal of Linguistics 30*, 349–409.

Browne, W. (1974). On the problem of enclitic placement in Serbo-Croatian. In R. Brecht and C. Chvany (Eds.), *Slavic Transformational Syntax*, Michigan Slavic Materials 10. Ann Arbor: Department of Slavic Languages and Literatures, University of Michigan.

Ćavar, D. (1997). On cliticization in Croatian: Syntax or prosody? In *Proceedings of the Workshop on the Syntax, Morphology and Phonology of Clitics. ZAS Papers in Linguistics, 6 (October)*, pp. 51–65. Universität Potsdam.

Curry, H. B. (1961). Some logical aspects of grammatical structure. In *Structure of Language and its Mathematical Aspects: Proceedings of the Twelfth Symposium in Applied Mathematics*, pp. 56–68. American Mathematical Society.

Dowty, D. (1996). Toward a minimalist theory of syntactic structure. In H. Bunt and A. van Horck (Eds.), *Discontinuous Constituency.* Walter de Gruyter.

Halpern, A. (1995). *On the Placement and Morphology of Clitics.* Stanford: CSLI.

Inkelas, S. (1989). *Prosodic Constituency in the Lexicon.* Ph. D. thesis, Stanford University.

Kathol, A. (1995). *Linearization-Based German Syntax.* Ph. D. thesis, Ohio State University.

Penn, G. (1998). On the plausibility of purely structural multiple WH-fronting. In *Proceedings of the Second European Conference on Formal Description of Slavic Languages.* Slavica Publishers.

Penn, G. (1999). Linearization and WH-extraction in HPSG: Evidence from Serbo-Croatian. In R. Borsley and A. Przepiórkowski (Eds.), *Slavic in HPSG*, pp. 149–182. Palo Alto: CSLI Publications.

Progovac, L. (1996). Clitics in Serbian/Croatian: Comp as the second position. In A. Halpern and A. Zwicky (Eds.), *Approaching Second: Second Position Clitics and Related Phenomena.* Stanford: CSLI Publications.

Reape, M. (1994). Domain union and word order variation in German. In J. Nerbonne, K. Netter, and C. Pollard (Eds.), *German in Head-Driven Phrase Structure Grammar*, CSLI Lecture Notes. CSLI Publications.

Richter, F. (forthcoming). *A Mathematical Formalism for Linguistic Theories with an Application in Head-Driven Phrase Structure Grammar and a Fragment of German.* Ph. D. thesis, Universität Tübingen. Working Title. Version: February 9th, 1998.

Richter, F., M. Sailer, and G. Penn (1999). A formal interpretation of relations and quantification in HPSG. This volume.

Zec, D. and S. Inkelas (1990). Prosodically constrained syntax. In S. Inkelas and D. Zec (Eds.), *The Phonology-Syntax Connection.* Chicago: University of Chicago Press.

Zlatić, L. (1997). *The Structure of the Serbian Noun Phrase.* Ph. D. thesis, University of Texas at Austin.

Zwicky, A. (1986). Concatenation and liberation. In *Papers from the 22nd Regional Meeting of the Chicago Linguistic Society*, pp. 65–74. Chicago Linguistic Society.

Gerald Penn

Seminar für Sprachwissenschaft
SFB 340, B8
Kleine Wilhelmstr. 113
72074 Tübingen
Germany

*E-mail address*: `gpenn@sfs.nphil.uni-tuebingen.de`
*Homepage*: `http://www.sfs.nphil.uni-tuebingen.de/~gpenn/`