Implementing Norwegian reflexives in an HPSG grammar

Lars Hellan ⓘD

NTNU, Trondheim

Abstract

The paper reviews basic patterns of reflexive binding in Norwegian, and explores a possible implementation of them in an HPSG grammar using the LKB platform. Norwegian has two reflexive elements, with distinct constraints and corresponding 'anti-binding' effects; they can cooccur but also occur independently. As over-all strategy for resolving reflexive binding we use one resembling the 'slash' procedure for wh-dependencies. Binding constraints are imposed partly through lexical specification, partly through phrasal combination rules. Challenges are noted residing in the possibility for sentences to contain an unbounded number of reflexives.

1   Introduction

Reflexive constructions in the Scandinavian languages obey a number of interacting constraints, involving factors such as linear order, c-command, finite vs. non-finite clausal domains, co-argumenthood, predication (the factor of the anaphor being contained in a phrasal unit predicated of the binder), thematic role hierarchies, and logophoricity. Moreover, the languages have two distinct reflexive elements which can combine, but also occur independently, and which each induces its own distinct 'anti-binding' ("Principle B"-) effects. HPSG being a framework aimed at enabling the integration of many levels of representation in a unified analysis, it is a promising candidate for accommodating the interplay of factors like those mentioned. The present paper explores an account of the phenomena for use in an HPSG grammar implementation using the LKB platform (Copestake 2002). A summary of the main phenomena as instantiated in Norwegian is given in section 2, and section 3 outlines strategies for encoding them in an LKB grammar. As the reflexive patterns are employed pervasively in the language, providing an account of them will have a high priority in a core grammar of Norwegian, and it will be shown that most aspects of the phenomena can be straightforwardly formalized using the strategies chosen. In subsection 3.4, though, we will note some clear challenges to the approach.

2   Empirical background

Like the other Scandinavian languages, Norwegian has two monomorphemic words that are inherently reflexive, in Norwegian taking the forms *selv* 'self', and *seg*, with the genitival form *sin*. *Seg* and *sin* are 3rd person forms. In 1st and 2nd person the corresponding forms coincide with their non-reflexive

counterparts, and for clarity of exposition we therefore largely use examples with the 3rd person forms. *Selv* is a constant form. These words may occur by themselves, but may also combine, in 3rd person as *seg selv*. The first four items listed in Fig 1 below are the NP type items to be called *reflexives*. For convenience, the words mentioned, when regarded as sub-NP items, may be referred to as *reflexive elements*, each such word representing the '+' variant of a binary feature, as informally indicated:

Fig 1.   Four reflexives and one pronominal in Norwegian, as defined through binary features:

*Bare **seg**-reflexive*	*Possessive reflexive*	***Seg selv**-reflexive*	*Pron-**selv**-reflexive*	*Pronominal*
NP	NP	NP	NP	NP
Refl-I +	Refl-I +	Refl-I +	Refl-I -	Refl-I -
Refl-II -	Refl-II -	Refl-II +	Refl-II +	Refl-II -
Poss -	Poss +	Poss -	Poss -	Poss -
seg	***sin***	***seg selv***	***ham selv***	***ham***
'him-/herself'	'his-/her own'	'him-/herself'	himself'	'him'

The reflexive elements (i.e., *seg, sin* and *selv*) are associated with different conditions for wellformedness. The conditions associated with *selv* are fairly similar to those holding for English *self*, concerning mainly co-argumenthood. The conditions associated with *seg* and *sin* (the elements marked 'Refl-I +' in fig.1) will now be reviewed briefly, following in essence Hellan 1988.

2.1   'Short' vs. 'long' distance reflexives

Reflexives consisting solely of the elements *seg/sin* have the possibility of 'long distance' binding, as exemplified in (1): a *sin*-reflexive can be bound 'out of' the genitival position of an NP, and further out of an infinitival VP, as seen in (1b); a bare *seg*-reflexive can be bound out of an infinitival VP, as seen in (1a); neither of these reflexives can be bound across a finite clause boundary, as seen in (1c).

(1)   a.   ***Jon*** ba meg snakke om ***seg/ *seg selv***
              Jon asked me to talk about him

b. ***Jon*** ba meg snakke om gramatikken ***sin***
Jon asked me to talk about his grammar

c. *****Jon*** hørte at jeg snakket om ***seg/*** gramatikken ***sin***
Jon heard that I talked about himself/his grammar

So-called 'short distance' binding includes at least the cases in (2):

(2) a. ***Jon*** omtaler ***seg selv***
Jon talks-about himself

b. ***Jon*** vasker ***seg***
Jon washes himself

c. vi fortalte ***Jon*** om ***ham selv***
we told Jon about himself

d. ***Jon*** fortalte oss om ***seg selv***
Jon told us about himself

e. ***Jon*** løp ***seg*** ut av laget
Jon ran himself out of the team

f. ***Jon*** leser boken ***sin***
Jon reads his book

g. ***Jon*** snakker om boken ***sin***
Jon talks about his book

In the cases (a)-(e) we may say that the binding relation obtains between *co-arguments*. In (a)-(c) and (e) this coincides with binder and bindee having a grammatical function related to the same lexical item. In (d), one standardly assumes that although the preposition may be said to have a semantics of its own, this semantics is here used to explicitly highlight one of the roles associated with the verb, so that in a semantic sense, the argument of the preposition is a coargument with the subject of the verb. In (e), such a relation of semantic co-argumenthood is absent, but a syntactic coargument relation holds. Notably, in (b) and (e) only a bare *seg*-anaphor is used, whereas the other cases have a reflexive with *selv*. To a large extent, one can maintain that *selv* is used only when semantic co-argumenthood obtains. The exceptional case is then (b). A generalization covering this case is that verbs expressing actions that are naturally or standardly of a type one performs on oneself, allow for the use of a bare *seg* as bound by a semantic (and syntactic) coargument. (The pattern in (b) also involves verbs whose bare *seg*-object may be obligatory or expletive or both (as in *skamme seg* 'be ashamed') - here semantic co-argumenthood may be seen as altogether absent, accounting for the lack of *selv*.)

In (f) and (g) *sin* is the 'possessor' argument relative to the expressed *possess* relation which has 'the book' as its 'possessed' argument; *sin* is therefore in neither case a coargument of *Jon*, neither syntactically nor semantically. Still

we count it as a 'short distance' anaphor here, as opposed to 'long distance' in (1b).

2.2   The Predication Condition on seg/sin

Crucial to reflexives composed with the elements *seg/sin* is the following condition:

(3)   Predication Condition on *seg/sin*:
       A reflexive *R* composed with the elements *seg/sin* has to be
       contained in a phrasal constituent understood as *predicated* of the
       binder of *R*.

This condition is distinct from a 'subject' condition, in that it licenses a construction like (4a), where the binder is syntactically an object; it is still distinct from a condition of 'any c-commanding item being licensed as binder', since it does not license (4b), where 'his money' is not an expression predicated on 'him':

(4)   a.   Vi gjorde ham glad i **seg** selv
             we made him fond of himself
       b.   ?*Vi fratok ham pengene **sine**
             we took his money from him

It may be noted that this notion of 'predication' is not tied to specific thematic roles for the 'subject'; thus, also subjects in passive sentences can satisfy the predication condition, as in (5):

(5)   Jon ble skutt av naboen **sin**
       Jon was shot by his neighbor

The Predication Condition offers an account of the contrast in (6), under the assumption that in (a), *ut av haven sin* ('out of his garden') is in a sense predicated of *Jon*, whereas in (b), *inne i haven sin/hans* ('inside of his garden') is predicated of the kicking event as such, and not of *Jon*. [1]

---

[1]Accepting this point for a movement performed by (what is expressed through) the object of a verb, it will be reasonable to assume that directionals qualifying a subject are also predicated of the mover, and not the event as such. The case thus provides empirical support for the position taken in Beermann and Hellan (2004), following proposals of, e.g., Jackendoff 1990 as opposed to Kracht 2003.

(6) a. Vi sparket Jon ut av haven **sin**
         we kicked Jon out of his garden
     b. Vi sparket Jon inne i haven ***sin**/ hans
          we kicked Jon inside of his garden

In (a), thus, Jon follows the trajectory expressed by 'out of his garden', and ends up in a state describable as 'Jon be outside of his garden'. Hence *Jon* here fulfills the Predication Condition holding of *sin*, validating the binding constellation in (a), whereas in (b), this condition is violated as far as a binding relation between *Jon* and the PP containing *sin* is concerned.

2.3   'Anti-binding' effects

Languages to varying extents grammaticalize a tendency to, for each anaphoric item operative in a certain domain, excluding other potential anaphors from that domain. For English, which only has one reflexive element (*self*), this tendency is observed in what has become encoded as the 'Principle B' of the Chomsky 1981 Binding Theory. For Norwegian, which has the *two* reflexive elements *seg/sin* and *selv*, one would expect this tendency to materialize in two anti-binding effects, one for each reflexive element; and this one can observe: the *selv*-reflexive appears to induce an anti-binding constraint on bare *seg*, to the effect that such a reflexive can not be used within a coargument domain, and *seg/sin* induces a constraint to the effect that within a constellation where the Predication Constraint is satisfied, a reflexive with a *ham* as first element is excluded.   These effects are exemplified in (7) and (8), respectively (all examples with a binding interpretation presupposed):

(7) a. ***Jon** omtaler **seg***
         Jon talks-about himself
     b. *vi fortalte   ***Jon** om **ham***
         we told   Jon about himself
     c. ***Jon** fortalte oss om **seg***
         Jon told us about himself

(8) a. ***Jon** omtaler **ham selv***
         Jon talks-about himself
     b. ***Jon** vasker **ham***
         Jon washes himself
     c. ***Jon** fortalte oss om **ham selv***
         Jon told us about himself
     d. ***Jon** løp **ham** ut av laget*
         Jon ran himself out of the team

e. *__Jon__ leser boken __hans__
Jon reads his book

f. *__Jon__ snakker om boken __hans__
Jon talks about his book

g. *Vi gjorde __ham__ glad i __ham__ selv
we made him fond of himself

For long distance patterns and the construction in (6a), both the *seg/sin* form and the *ham/hans* form are possible:

(9) a. __Jon__ ba meg snakke om __seg/ ham__
Jon asked me to talk about him

b. __Jon__ ba meg snakke om gramatikken __sin/ hans__
Jon asked me to talk about his grammar

c. Vi sparket __Jon__ ut av haven __sin/ hans__
we kicked Jon out of his garden

As far as syntactic determinants go, the alternations in (9) are free, however, a consistent observation throughout the literature is that the reflexive options have a subject-centered point of view expressed, in opposition to the options using a pronominal. With that factor taken into account, one could say that an anti-binding effect is at play here as well: in the 'logophoric' domain, only *seg/sin* can be used. An analogous reasoning can be applied concerning the status of the ungrammaticality of (7a), recalling the grammaticality of (2b), repeated:

(2b) __Jon__ vasker __seg__
Jon washes himself

As noted above, the pattern in (2b) obtains for activities that are naturally, although not always necessarily, self-directed. Let's call such self-directed two-place relations lexically 'tamed', and the object simply 'tamed': the anti-binding effect observed in (7a), which is one induced by the availability of *selv*, is then one that obtains relative to *non*-tamed second arguments: *selv* is only used with non-tamed objects, and these are therefore also the ones that are 'defended' against other encodings. (For 'tamed' objects, like the *seg* in (2b), the anti-binding effect geared to the predication factor is demonstrated in (8b).)

The phenomena now mentioned constitute the more 'robust' patterns of reflexives in Norwegian, and are those that will be considered in the next section. Among important areas we have not touched on are 'reconstruction' effects related to wh-fronting of constituents containing reflexives, and possible more subtle effects of thematic role dominance, both topics which would naturally have been included in a more extensive overview.

3   Implementing the patterns

The LKB formalism is a rather restrictive version of typed feature formalism, disallowing negative constraints, disjunctive constraints and quantificational constraints (i.e., constraints of the form 'for some...' or 'for all...'). Assemblies of items are construed as lists, not sets. Possible loss of expressive power under such a formalism is compensated for by gain in efficiency (cf., e.g., Flickinger (2000)). Among the growing family of LKB based grammars (referred to as 'Matrix grammars') related to the 'HPSG Grammar Matrix' (Bender et al. 2000), is a grammar for Norwegian, *NorSource* (cf. //www.ling. hf.ntnu.no/forskning/norsource/), which constitutes a background and actual testbed for proposals under discussion, without however being explicitly invoked in the present exposition.

Anaphora has so far only to a small extent been reflected in the LKB grammars available; for some languages, this is in part because reflexivity is arguably a largely pragmatic phenomenon. In Norwegian, as the overview will have shown, reflexives and their anti-binding effects are firmly situated in the core grammar itself, and only marginally related to pragmatics; hence, they should be covered by a Norwegian grammar. As most of the modules present in the Norwegian anaphora system can be found in many other languages as well, the efforts going into this task hopefully may be of relevance also beyond Norwegian.

3.1   Determining the 'locus' for anaphora resolution

Manning and Sag (1998), extending the design of Pollard and Sag (1994), propose ARG-ST as an appropriate construct for imposing binding conditions: this is a list, for each predicate expressed, of those of its arguments that are syntactically realized, ordered according to an 'obliqueness' hierarchy. An anaphor, according to these proposals, is acceptable under two circumstances: if it occurs non-initially on its ARG-ST list, its binder must be an item preceding it on that list; and if initial, its interpretation is free. This account does not seem very relevant for the phenomena we have seen: constructions like (1a,b), repeated as (10a,b), and (2f,g), repeated as (10c,d),

(10)   a.   ***Jon*** ba meg snakke om ***seg***
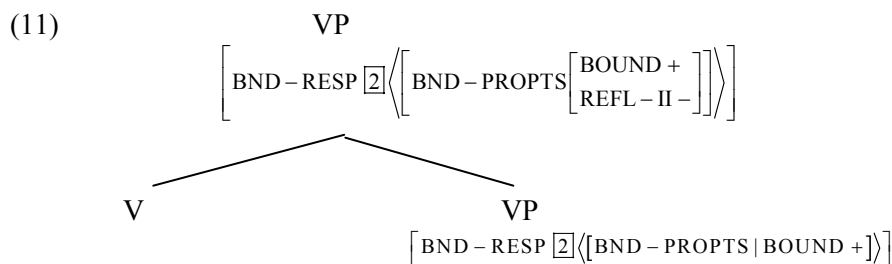   Jon asked me to talk about him
         b.   ***Jon*** ba meg snakke om gramatikken ***sin***
                  Jon asked me to talk about his grammar
         c.   ***Jon*** leser boken ***sin***
                  Jon reads his book

d.　　***Jon*** snakker om boken ***sin***
　　　　　Jon talks about his book

show that grammatically necessary antecedents may appear outside of the local domain represented by a single ARG-ST list; (10c,d) also show that items initial on an ARG-ST list (as a possessor argument perhaps is) are by no means necessarily free in their interpretation.[2]

　　　The general strategy for anaphora resolution to be envisaged here is somewhat traditional relative to an HPSG/GPSG setting (resembling, for instance, the 'slash' strategy for wh-dependencies; it may also be seen as compatible with proposals in Branco 2001): Each anaphor, as encountered by the phrasal combination rules, is 'stored' with its critical information: a reference index, its agreement features, and its feature characterization in terms of the possibilities given in fig. 1. When, in a later (higher) combination rule, a putative binder is encountered, then, given match in agreement features and acceptance of the 'REFL-I/REFL-II' constellation relative to the putative binder, the indices of the putative binder and the reflexive are identified and the 'store' containing the anaphor information is emptied. By technical assumption, no non-emptied store can be present at the final combinatorial stage. We now consider how to implement these conceived moves one by one.
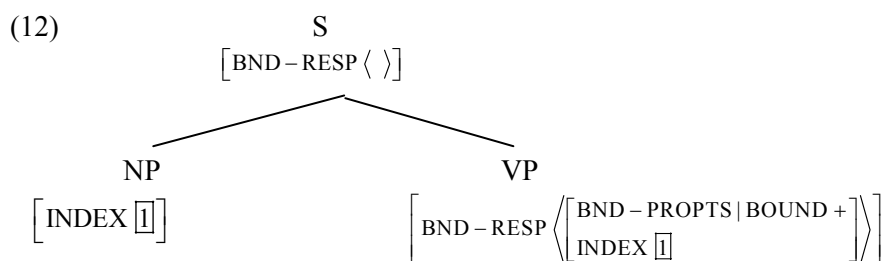
　　　Given strict locality of combinatory rules, to enforce that no *selv*-anaphor can be long distance bound (cf. (1a)), a rule combining a (non-auxiliary) V with a VP will have to impose as a general restriction that any reflexive to be carried up in the 'store' from this point on has to be a non-*selv* reflexive - i.e., one with the feature 'REFL-II -'. I.e., we may envisage a combination rule of the following form:

(11)　　　　　　　　　VP

$$\left[\text{BND} - \text{RESP} \; \boxed{2} \left\langle \left[\text{BND} - \text{PROPTS} \begin{bmatrix} \text{BOUND} + \\ \text{REFL} - \text{II} - \end{bmatrix} \right] \right\rangle \right]$$

　　　　　V　　　　　　　　　　　　VP

$$\left[\text{BND} - \text{RESP} \; \boxed{2} \left\langle \left[\text{BND} - \text{PROPTS} \,|\, \text{BOUND} + \right] \right\rangle \right]$$

'BND-RESP' here introduces the package of information about the reflexive for which the structure has a 'binding responsibility'; it is stated as a singleton list, to allow expression of the possibility that there be no item under such a responsibility (and also that there be more than one - cf. 3.4 below). 'BOUND +'

---

[2] Although no mention will be made of ARG-ST or similar devices in what follows, we are not precluding that it may be relevant at some level of analysis, for instance, for incorporation of an account of role dominance (cf. end of section 2).
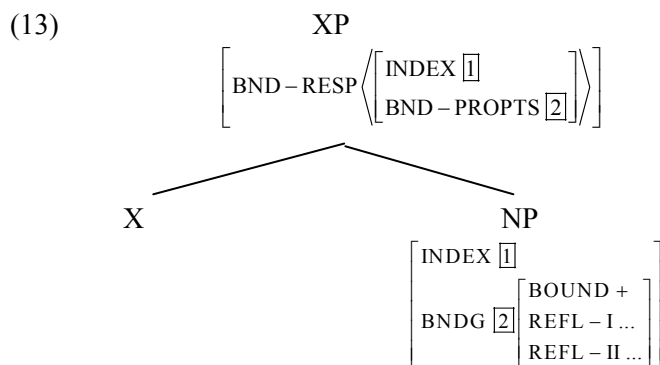
indicates that the reflexive must be bound. A structure where a *selv*-anaphor has not been bound by the point where the rule (11) applies, will fail unification. Hence, for the rule which technically resolves the anaphor (such as (12) below), it is, in this respect, enough that the structure is wellformed up to that point; for the rule in question, what remains is to identify indices and check for agreement. If we assume, along with Pollard and Sag op.cit., that agreement features are actually part of the referential index, the operation of the 'binding' rule will be essentially as indicated in (12):

(12)

$$
S \\
\left[\text{BND} - \text{RESP} \langle \ \rangle \right]
$$

NP
$$
\left[\text{INDEX} \ \boxed{1} \right]
$$

VP
$$
\left[\text{BND} - \text{RESP} \left\langle \begin{array}{l} \text{BND} - \text{PROPTS} \,|\, \text{BOUND} + \\ \text{INDEX} \ \boxed{1} \end{array} \right\rangle \right]
$$

If we assume that bound occurrences of pronominals are marked with the specification 'BOUND +', then (12) will subsume their resolution as well.

For the constellation where an S combines with a finite complementizer, an effect similar to that in (11) obtains, now in addition requiring that REFL-I also be 'minus'. Thereby, no reflexive can be bound across a finite S-boundary.

The above is the gist of a procedure, and the next steps will be to show how it accommodates more of the various properties of Norwegian reflexives mentioned above. To begin with, let us assume the structure has only one reflexive. The first step is to 'mount' this reflexive in the BND-RESP list of that part of the structure which will be propagated up through the combinatorial steps. Let us assume - as is also typically the case - that this part of the structure is a head of which the reflexive acts as a complement. The 'mounting' step can then be portrayed as in (13):

(13)

$$
XP \\
\left[\text{BND} - \text{RESP} \left\langle \begin{array}{l} \text{INDEX} \ \boxed{1} \\ \text{BND} - \text{PROPTS} \ \boxed{2} \end{array} \right\rangle \right]
$$

X

NP
$$
\left[\begin{array}{l} \text{INDEX} \ \boxed{1} \\ \text{BNDG} \ \boxed{2} \left[\begin{array}{l} \text{BOUND} + \\ \text{REFL} - \text{I} \ ... \\ \text{REFL} - \text{II} \ ... \end{array} \right] \end{array} \right]
$$

Every NP has a specification with regard to the features introduced inside the BNDG feature, indicating its binding potentials. Every NP also has an INDEX (actually introduced much deeper inside a feature path than we need to expose at this point - see (21) below). The dominating node encapsulates these pieces of information inside its feature BND-RESP.
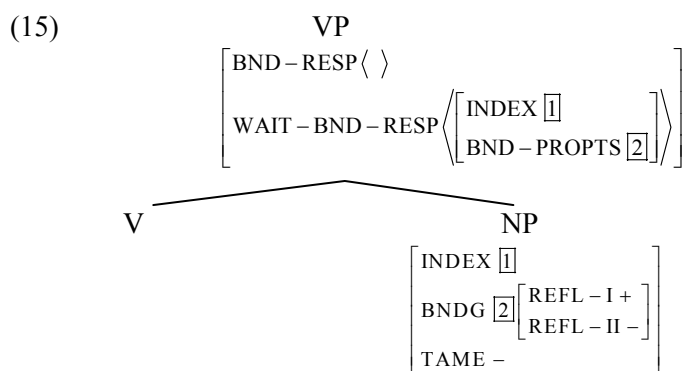
The sequence (13)-(11)-(12) gives a first rough sketch of how an anaphor can be technically resolved. Relative to this picture, we in the next two sections consider how to implement 'anti-binding' effects for bare *seg*-reflexives and how to impose the Predication Condition and its 'anti'-counterpart. After that, we address how the procedure deals with the presence of more than one reflexive in a given structure, and possibly bound by different antecedents.

3.2   Implementing 'anti-binding' effects

Suppose that the anaphor is the bare *seg*; if the NP is 'non-tamed', the rules should license a structure like (14a), but at the same time (14b) should be blocked, both repeated from earlier:

(14)   a.   ***Jon*** ba meg snakke om ***seg***
                Jon asked me to talk about him
        b.   ****Jon*** omtaler ***seg***
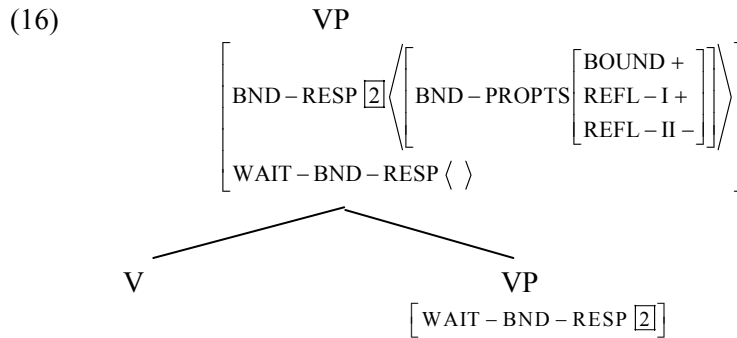   Jon talks-about himself

Thus, we now need to account for the anti-binding effect induced by *selv*, to rule out (14b); but we at the same time must enable a successful derivation of (14a). To achieve this, we introduce a device of 'provisional mounting', by which *seg* in (14a), as realizing an NP marked as 'non-tamed', is put on a waiting list for entrance to the dominating node's BND-RESP; (15) indicates how this may be expressed.

(15)

$$
\begin{array}{c}
\text{VP} \\
\begin{bmatrix}
\text{BND} - \text{RESP} \langle\ \rangle \\
\text{WAIT} - \text{BND} - \text{RESP} \left\langle \begin{bmatrix} \text{INDEX} \;\boxed{1} \\ \text{BND} - \text{PROPTS}\;\boxed{2} \end{bmatrix} \right\rangle
\end{bmatrix}
\end{array}
$$

V   NP
$$
\begin{bmatrix}
\text{INDEX}\;\boxed{1} \\
\text{BNDG}\;\boxed{2} \begin{bmatrix} \text{REFL} - \text{I} + \\ \text{REFL} - \text{II} - \end{bmatrix} \\
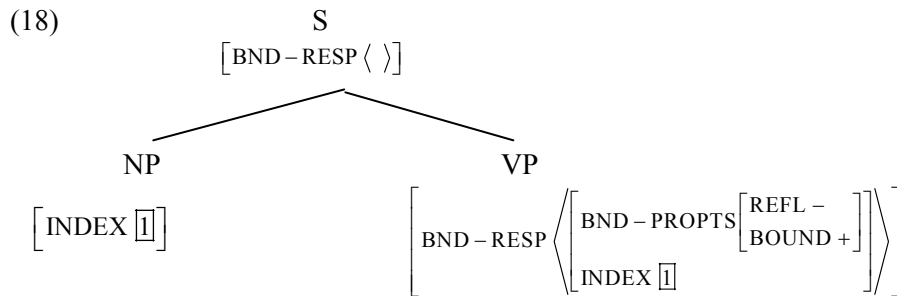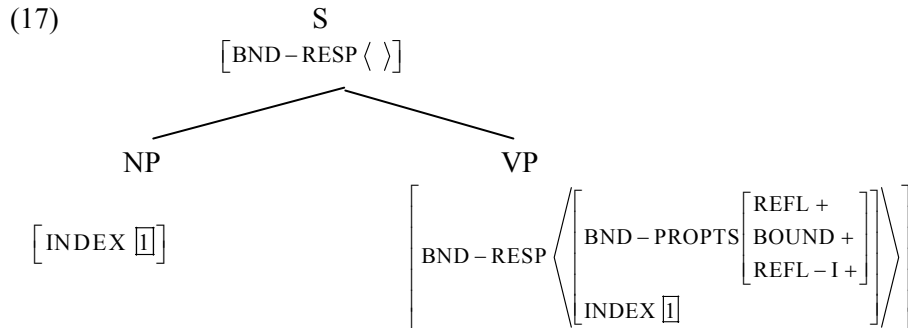\text{TAME} -
\end{bmatrix}
$$

529

If the VP in (15) is combined with a subject, that subject will not be in a position to bind the reflexive, since (12) induces binding only for items inside BND-RESP. In this way, (14b) is excluded, the anti-binding effect induced by *selv* being enforced.

For cases like *vaske seg*, in contrast, where the verb frame defines the object as 'TAME +', we will assume that the reflexive is put directly on the BND-RESP list, thereby accepting these kinds of bare *seg*.

In the configuration pictured in (11), the reflexive specification can be shifted out of the 'waiting list' and into BND-RESP list. (11) is thus instantiated as (16) (- technically, (16) can be construed as a subtype of (11)). Thereby, (14a) can now be derived, the reflexive now being in the BND-RESP list proper:

(16) VP

$$
\left[
\begin{array}{l}
\text{BND} - \text{RESP} \boxed{2} \left\langle \left[\text{BND} - \text{PROPTS} \left[\begin{array}{l} \text{BOUND} + \\ \text{REFL} - \text{I} + \\ \text{REFL} - \text{II} - \end{array} \right] \right] \right\rangle \\
\text{WAIT} - \text{BND} - \text{RESP} \left\langle \ \right\rangle
\end{array}
\right]
$$

V   VP

$$\left[\text{WAIT} - \text{BND} - \text{RESP} \boxed{2} \right]$$

How, then, next, do we impose the anti-binding effect tied to *seg/sin*, excluding anaphors with *ham* from being predication-contained relative to their antecedent? One strategy will be to specify rules like (12), whose domain is one where the Predication Condition relative to a reflexive inside the VP is met, to exclude a pron-reflexive from its scope. It should be noted that a plain pronominal, when being in a distinct finite domain from the binder, can well be bound by a subject - the exclusion effect obtains only when *ham* is part of a reflexive. The refinement of (12) thus could consist of an exclusion of the package '[REFL-I -, REFL-II +]'. However, the exclusion effect comprises also possessive items - in the relevant domain, they must be *sin*, not *hans*, and neither form is '[REFL-II +]'. Hence, a further feature has to be added, vz. 'REFL bool', and the amendment of (12) inducing the anti-binding effect will be as in (17), whereas a possible rule inducing binding of pronominals may have the form of (18):

(17)

$$
\begin{array}{c}
\text{S} \\
[\text{BND}-\text{RESP}\langle\ \rangle]
\end{array}
$$

NP
$\qquad$
VP

$$
[\text{INDEX}\ \boxed{1}]
$$

$$
\left[\text{BND}-\text{RESP}\left\langle\begin{bmatrix}\text{BND}-\text{PROPTS}\begin{bmatrix}\text{REFL}+\\ \text{BOUND}+\\ \text{REFL}-\text{I}+\end{bmatrix}\\ \text{INDEX}\ \boxed{1}\end{bmatrix}\right\rangle\right]
$$

(18)

$$
\begin{array}{c}
\text{S} \\
[\text{BND}-\text{RESP}\langle\ \rangle]
\end{array}
$$

NP
$\qquad$
VP

$$
[\text{INDEX}\ \boxed{1}]
$$

$$
\left[\text{BND}-\text{RESP}\left\langle\begin{bmatrix}\text{BND}-\text{PROPTS}\begin{bmatrix}\text{REFL}-\\ \text{BOUND}+\end{bmatrix}\\ \text{INDEX}\ \boxed{1}\end{bmatrix}\right\rangle\right]
$$

Since, other things being equal, having one unique rule for all binding-by-subject would have been preferable, we have to count this as a possible complexity. Moreover, in the formalism we are using, a rule of subject-VP combination distinct from both (17) and (18) is needed for the case where there is no item in the VP to be bound by the subject - the LKB formalism does not allow conditional rules. So, the 'count' of subject-VP rules stands potentially at three, at the moment.

Pursuing the strategy of (17), we need to identify all constellations where a potential binder combines with a VP or other XP predicated of it. This will include those exemplified (4a) and (6a), repeated in (19a,b), along with further instances in (19), all cases where the binder is syntactically a direct object (and a binding interpretation is understood/probed):

(19) a. Vi gjorde **ham** glad i **seg selv/ *ham selv**
    we made him fond of himself
  b. Vi sparket **Jon** ut av haven **sin/ *hans**
    we kicked Jon out of his garden
  c. Vi hørte **Jon s**nakke om pengene **sine/ *hans**
    we heard Jon talk about his money
  d. Vi ba **Jon** snakke om pengene **sine/ *hans**
    we asked Jon to talk about his money

Whether or not all of these constellations can be subsumed under one rule, it or they will include a specification like that found for the VP in (17). And the proliferation necessitated for cases where a pronominal is to be bound, and where no item is to be bound, is repeated at this level. Crucially not covered by the rule(s) in question is the constellation in (20a,b), where an indirect object binds into a direct object; here the relevant specification will be as indicated in (20c), where the dotted area will include the binder, although in a way we will not try to make more precise at this point, as it involves technicalities arising from a consistently binary branching view of phrasal structure:

(20)   a.   Vi fratok ham pengene **hans/ *sine**
           we took his money from him
       b.   Vi fratok Jon **ham selv / *seg selv**
           we from-took Jon himself
           'we deprived Jon of himself'

       c.   VP
              $\left[\text{BND} - \text{RESP}\left\langle\ \right\rangle\right]$

       ...   NP
              $\left[\text{BND} - \text{RESP}\left\langle\left[\text{BND} - \text{PROPTS}\begin{bmatrix}\text{REFL} + \\ \text{BOUND} + \\ \text{REFL} - \text{I} -\end{bmatrix}\right]\right\rangle\right]$

       Summarizing, we have suggested implementing anti-binding effects in essentially two ways: those relating to *selv* are enforced through the mounting rule (15) and the phrasal combination rule (16) (and possible counterparts for configurations we have not looked at here), whereas those relating to the predication condition are enforced through the statement of the binding rules themselves, such as (17), its counterpart for cases like (19), and (20c).


3.3   A unified implementation of the Predication Condition?

Ideally, one would like the Predication Condition as stated in (3) to be implementable in one single specification in the grammar. In general, LKB grammars allow generalizations to come out either through type inheritance - a supertype encodes what is common to the subtypes - or through unary rules - a certain specification sits in the input to one or more unary rules each producing

different structures/types/rules, but all preserving the input specification in question. In the present context, this is to say that either, rules like those portrayed in (17) and (20c) could be construed as subtypes of a given type, or they could be constructed off from some sort of 'basic' binding rule covering the Predication constellation. Either way, one would need to abstract away from the specific configurational or functional status of the binders - such as the status of subject/sister of VP in (17) - and identify a status corresponding to 'predicate' and 'predication subject'. Let us consider how this could be done.

Arguably, adjectives like *glad* in (4a) (*Vi gjorde **ham** glad i **seg** selv* 'we made him fond of himself') are not valence defined for a subject, but they do have a semantic representation of a 'logical subject', something which in the Matrix grammars is entered as an ARG1 of the predicate in its semantic specification. Also verbs have an ARG1, often corresponding to their subject; however, in passive constructions, the ARG1 systematically is not what is expressed as subject, still, also subjects in passive sentences can bind a reflexive, as in (5) (*Jon ble skutt av naboen **sin*** 'Jon was shot by his neighbor'). In the framework in question, a common denominator for these options is XARG: for a verbal lexeme, this is the participant expressed by the (surface) subject, and for non-verbal lexemes, it is identical to its ARG1. Using these terms, the Predication Condition will require, for any reflexive *R* composed with *seg/sin*, that its binder have a status as XARG. To illustrate, in the 'generalized type' of binding rule imposing the Predication Condition (of which the case instantiated in (17) would be a subtype, or a 'unary rule' derivative), one thus would envisage the partial specifications in (21) (to stay faithful to the actual formalism, we show the full feature paths introducing XARG and INDEX):

(21)
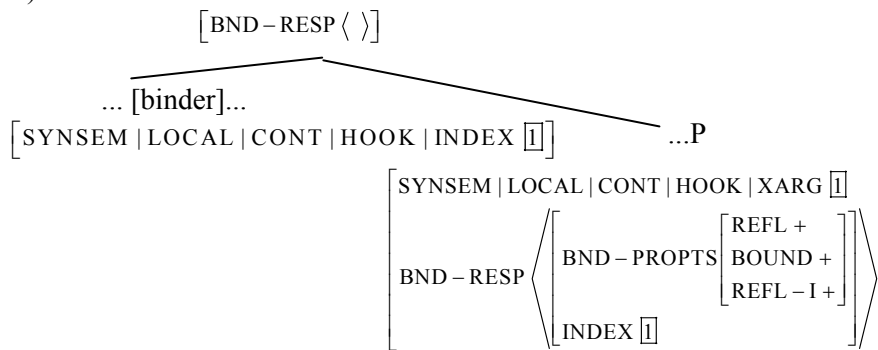
Since a direct object does not have the preceding indirect object as its XARG, the structure in (20) would not satisfy the specification in (21), and so would not inherit the 'Predication Condition' type.

Whether a supertype like (21) can realistically be constructed relative to the relevant rules in the grammar, is a question that goes beyond the scope of the present discussion; however, it is reasonably clear what we would be looking for. The same goes for a unary rule utilization of such an underspecified representation.

The reasoning around (21) warrants a slight digression. The procedure conceptualized presupposes a general one-to-one correspondence between NP tokens in a sentence and ref-indices representing the NPs in the semantic representation. In the standard procedures of assigning semantic representations to sentences, this is indeed obeyed, but an anaphor and its binder are traditionally assumed to share referential index; here, thus, there is no one-to-one-correspondence NP - referential index being made. However, somewhat subtle situations can arise where exactly this might be desirable. Consider the contrast in (22), where in (22a), the lines indicate licensing relations that are per se acceptable:

(22)

a. *Jon hørte seg selv snakke om seg
        Jon heard himself talk about himself
b. Jon hørte seg selv snakke om seg selv
        Jon heard himself talk about himself

Underlying this contrast seems to be a constraint to the effect that if two reflexives are licensed as bound by the same NP, but are arguments of different predicates, then for the second of the reflexives it must also be verified that it is licensed as bound by the first of the reflexives. In (22a), *seg* indeed cannot be bound by *seg selv*, due to the 'long distance' requirement inherent in bare *seg* (as 'non-tamed'). In checking if *seg selv* in (22a) is an eligible binder of *seg*, it is crucial that it is the potential XARG status (relative to *snakke*) of *seg selv* we are checking, and not that of *Jon*. But if these have the same referential index, it is not obvious how to formally guarantee this.

Two options present themselves for resolving this kind of situation. One is to let the binding rules introduce explicit *identity relations between indices*, thereby letting each referential index be unique to one NP token. Another might be to rely on a bottom-up application of binding rules, by which the second reflexive in (22) would be necessarily first related to the first reflexive for possible binding, if these were to be represented as coreferential with each other at all. At this point, we just state these possibilities.

3.4 Dealing with multiple reflexives

The limitation to one reflexive per sentence, as we adopted in the previous paragraphs, is not representative of how reflexives occur in Norwegian. (23) are examples where two reflexives share a binder, whereas in (24), each reflexive has a different binder (in (b), as one of two readings):

(23) a. **Jon** omtaler **seg selv** i alle **sine** bøker
          'Jon mentions himself in all his books'
      b. **Jon** så **sine** motstandere komme mot **seg**
          'Jon saw his adversaries coming against him'
(24) a. **Jon** så *meg* krysse **seg** ut av registeret *mitt*
          'Jon saw me crossing him out of my register'
      b. **Jon** ba *Marit* fortelle **seg** om *sine* opplevelser
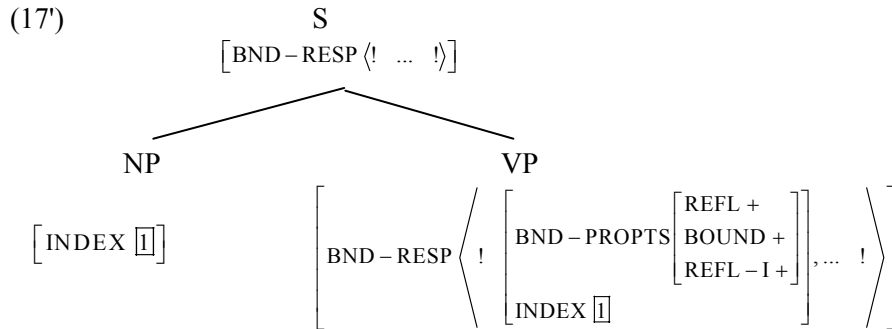   'Jon asked Marit to tell him about her experiences'

As the discussion has already shown, the propagation of reflexive specification can come from all sorts of constituents: VPs, PPs, NPs, and any depth of embedding of NPs within PPs, for instance. Defining BND-RESP as taking a *list* as value seems in principle the right choice, since a given constituent can host many reflexives. Composition of lists from lists through the phrasal combinatorics technically will have to deal with what is called 'difference lists', informally marked as '<!...!>' rather than '<...>'. Acceptance of a sentence will require an empty such list in the end. We now consider how this can be obtained.

By assumption, (23a), repeated as (25a), may have a difference list of the form (25b) by the time a first binding rule (such as (17)) comes into operation:
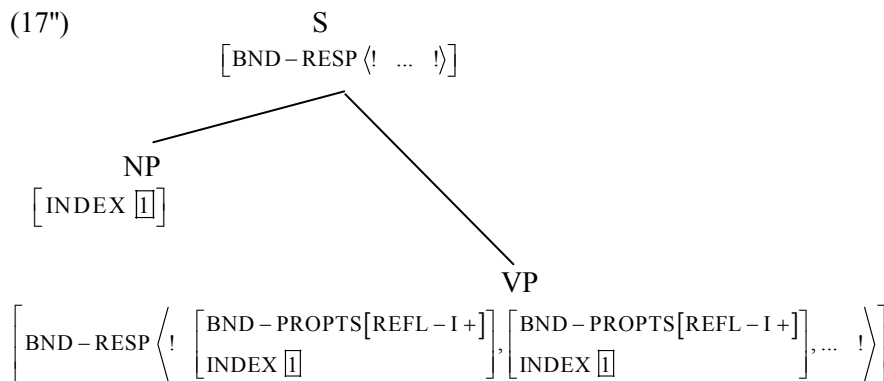
(25) a. **Jon** omtaler **seg selv** i alle **sine** bøker
          'Jon mentions himself in all his books'

      b.

$$\text{BND}-\text{RESP}\left\langle !\begin{bmatrix} \text{BND}-\text{PROPTS}\begin{bmatrix} \text{REFL}-\text{I}+ \\ \text{REFL}-\text{II}+ \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{BND}-\text{PROPTS}\begin{bmatrix} \text{REFL}-\text{I}+ \\ \text{REFL}-\text{II}- \end{bmatrix} \end{bmatrix}! \right\rangle$$

The exact order in which the items appear on the list will depend on which head projection one starts with, and on whether new items are (on lists) prefixed or suffixed to the existing list; the order in (b) results, e.g., if one starts with the verb projection, and suffixes new items. A binding rule will, technically, have to address difference lists rather than lists. Suppose that we amend (17) minimally

535

(as (17')) to meet this requirement; the operation informally indicated is that of removing the *first* item on the VP's BND-RESP list:

(17')

$$S$$
$$\left[\text{BND}-\text{RESP}\ \langle!\ \ldots\ !\rangle\right]$$

NP
$$\left[\text{INDEX}\ \boxed{1}\right]$$

VP
$$\left[\text{BND}-\text{RESP}\ \left\langle!\ \begin{bmatrix}\text{BND}-\text{PROPTS}\begin{bmatrix}\text{REFL}\ +\\ \text{BOUND}\ +\\ \text{REFL}-\text{I}\ +\end{bmatrix}\\ \text{INDEX}\ \boxed{1}\end{bmatrix},\cdots\ !\right\rangle\right]$$
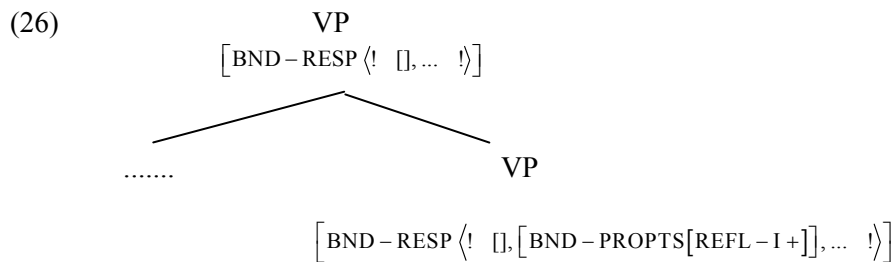
By its underspecification regarding 'REFL-II', it might seem that (17') should be able to apply to either item on the list in (25b) - a not unnatural way of conceiving the operation of this rule could indeed be that it applies to *all* items on the list satisfying its description. The INDEX of the subject NP being constant, this would mean inducing the same binder for both of the reflexives, which in this case would be correct as far as the reading is concerned. The obvious problem is that in the LKB architecture, such quantification over 'all items' is illicit - a rule can apply only to specific parts of a structure, for which it is explicitly declared. Thus, as it stands, (17') can apply only to the first item on the list. Since (17') defines a subject-VP combination, moreover, and there is only one subject, the rule cannot be reiterated. Thus, to get both items in the list in (25b) bound by the same subject, an alternative binding rule will have to be applied, explicitly binding the *first two* items on the list - cf. (17"):

(17")

$$S$$
$$\left[\text{BND}-\text{RESP}\ \langle!\ \ldots\ !\rangle\right]$$

NP
$$\left[\text{INDEX}\ \boxed{1}\right]$$

VP
$$\left[\text{BND}-\text{RESP}\ \left\langle!\ \begin{bmatrix}\text{BND}-\text{PROPTS}\left[\text{REFL}-\text{I}\ +\right]\\ \text{INDEX}\ \boxed{1}\end{bmatrix},\begin{bmatrix}\text{BND}-\text{PROPTS}\left[\text{REFL}-\text{I}\ +\right]\\ \text{INDEX}\ \boxed{1}\end{bmatrix},\cdots\ !\right\rangle\right]$$

In (24b) (repeated), in turn, on the reading indicated,

(24b) **Jon** ba *Marit* fortelle **seg** om *sine* opplevelser
'Jon asked Marit to tell him about her experiences'

the VP "fortelle **seg** om *sine* opplevelser" will have a 2-membered list where, to represent this reading, we want the binding by *Marit* to apply only to the *second* item on the list. We thus need a VP specification in the binding rule as in (26), removing only the second item from the list:

(26)

$$\text{VP}$$
$$\left[\text{BND} - \text{RESP}\left\langle!\ \ [],\ldots\ \ !\right\rangle\right]$$

....　　　　　　　　　　　　　　VP

$$\left[\text{BND} - \text{RESP}\left\langle!\ \ [],\left[\text{BND} - \text{PROPTS}[\text{REFL} - \text{I} +]\right],\ldots\ \ !\right\rangle\right]$$

There is in principle no end to how many reflexives a binder may have to resolve - an indication is given in (27a), where the VP combining with *Jon* has four items - or how far out in a BND-RESP sequence a rule may have to go in order to select a reflexive licit on a given reading - in (27b), this is the third in the list under the VP starting with *beskrive*, and on another interpretation it could be number two and three, for instance - it is especially the possessive reflexives which can bring the number up in these ways:

(27)
a.
Jon satte **sin** bok om **sine** tvister med naboen **sin** aller høyest i **sitt** forfatterskap
'Jon valued his book about his controversies with his neighbor highest of his writings'
b.
Marit ba Jon beskrive **sin** bok om **sine** tvister med naboen *sin* som den aller beste i **sitt** forfatterskap
'Marit asked Jon to describe his book about his controversies with her neighbor as the very best of his writings'

For the strategy we are pursuing, the numbers now alluded to would be enlarged if we also treat pronominal binding by this same mechanism, and for every relevant configuration, there is of course also a rule for the case where BND-RESP is empty. We have been vague about how many binding configurations there actually are - for instance, the NP internal domains have not been mentioned - and the possibilities of generalizing over constellations are not

settled, as indicated in the discussion in 3.3. (And even for such a straightforward constellation as subject-VP, it is not to be taken for granted that main and subordinate clauses can be treated by a uniform rule, or inverted and non-inverted structures - but they all unfold the same possibilities of reflexive binding.)

For an implemented grammar using the design in question, it seems clear that an attempted coverage of all possibilities would constitute too much of a load on the grammar. For a grammar of Norwegian, it is equally clear that failing to cope with reflexives altogether makes the grammar inadequate. For what we have called 'tamed' reflexives, the device of augmented lexical frames will work, but these constructions cover only a small part of the domain, so some steps need to be taken to accommodate patterns with 'non-tamed' reflexives. Rules of the types in (17), (17') and (17'') then have to be adopted, and it will be a calibration question at what number of reflexives per sentence (such as, for instance, two) and on how many positions in a list, to set the limit.[3]

4 Conclusion

The article has laid out the rather complex patterns of Norwegian reflexives for accommodation in the restrictive typed feature formalism of the LKB/ Matrix systems. On the one hand, we have shown that constraints tied to both types of reflexive elements, as well as their anti-binding effects, can be concisely formalized - this holds both of domain-size constraints (co-argumenthood) and containment constraints (predication). On the other hand, in dealing with patterns of multiple reflexives, we have seen that the strategy chosen may face challenges in the form of rule proliferation.

References

Beermann, Dorothee, and Lars Hellan. 2004. A treatment of Directionals in two implemented HPSG grammars. In Stefan Müller (ed) *Proceedings of the HPSG04 Conference*. Center for Computational Linguistics, Katholieke Universiteit Leuven. CSLI Publications /http://csli-publications.stanford.edu/

Bender, Emily, Daniel Flickinger and Stephan Oepen. 2002. The Grammar Matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In: *Proceedings of the Workshop on Grammar Engineering and Evaluation at*

---

[3] Aside from presenting challenges within the LKB design, these phenomena may also invite to comparison with other platforms or architectures that could provide a more expressive formalism, such as TRALE (cf. Meurers et al. (2002)).

the 19th International Conference on Computational Linguistics, 8-14, Taipeh, Taiwan.

Branco, António. 2001. Without an Index: a lexicalist account of binding theory. In Frank van Eynde, Dorothee Beermann and Lars Hellan (eds) *Proceedings of the HPSG Conference 2001*, NTNU, Trondheim. CSLI Publications /http://csli-publications.stanford.edu/

Chomsky, Noam. 1981. *Lectures on Government and Binding.* Foris Publications, Dordrecht.

Copestake, Ann. 2002. *Implementing Typed Feature Structure Grammars*. CSLI Publications.

Flickinger, Daniel. 2000. On building a more efficient grammar by exploiting types. *Natural Language Engineering* 6(1): 15-28, Cambridge University Press.

Hellan, Lars. 1988. *Anaphora in Norwegian and the Theory of Grammar.* Foris Publications, Dordrecht.

Jackendoff, Ray. 1990. *Semantic Structures* MIT Press.

Kracht, Marcus. 2002. On the Semantics f Locatives. *Linguistics and Philosophy* 25: 157-232.

Manning, Christopher, and Ivan Sag. 1998. Argument Structure, Valence and Binding. In *Nordic Journal of Linguistics*, 21, 107-144.

Meurers, Detmar, Gerald Penn and Frank Richter. 2002. A Web-Based Instructional Platform for Constraint-Based Grammar Formalisms and Parsing. In: Dragomir Radev and Chris Brew (eds) *Effective Tools and Methodologies for Teaching NLP and CL*, 18-25. The Association for Computational Linguistics. New Brunswick, NJ.

Pollard, Carl and Ivan Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago Univ. Press.