A lexicalist account of the Dutch verbal complex

Gosse Bouma and Gertjan van Noord
Alfa-informatica and
Behavioral and Cognitive Neurosciences
Rijksuniversiteit Groningen
{gosse,vannoord}@let.rug.nl

Abstract

This paper presents a categorial analysis of several types of verbs selecting a verbal complement in Dutch. In particular, we demonstrate that the various word orders found in the verbal complex can be accounted for by mechanisms operating in the lexicon only. Our theoretical starting point is the proposal of Hoeksema (1991), who accounts for 'verb raising' by stipulating a polymorphic category for verbs inducing this type of word order. The effect of polymorphism is equivalent to using a categorial rule of division in the lexicon. We implement this rule as a recursive constraint. The coverage of the analysis is extended by showing that various other word order possibilities within the verb cluster imply that harmonic and disharmonic versions of division are needed. Finally, we argue that overgeneration can be avoided by requiring that the argument of a 'verb raising' verb must be a 'verbal complex'.

1 Introduction

In Dutch, verbs selecting a VP-complement come in at least three varieties. First, a verb, such as *verbieden* (*to disallow*), may subcategorize for a full VP to its right:

(1) ... dat Jan Marie verbiedt [het boek te lezen]
... that John Mary disallows the book to read
... that John disallows Mary to read the book

These verbs are called *extraposition* verbs, as transformational accounts assume that the VP-complement has been extraposed from a position in the midfield. Second, a verb may form a 'verbal complex' or 'verb cluster' with the head of its VP-complement. In this case, the non-verbal elements of the VP-complement occur to the left of the governing verb, whereas the verbal head of the VP-complement normally occurs to the right (2a). This is the so-called *verb-raising* (VR) construction, as transformational accounts (following Evers (1975)) assume that in this case the head of the embedded VP-complement is 'raised' from its original position left of the governing verb to a position on the right. Modals, such as *willen* (*to want*) or *kunnen* (*can*), are typical VR verbs, as well as perception verbs such as *zien* (*to see*) or *horen* (*to hear*), and causatives such as *laten* (*to let*). The VR construction gives rise to so-called *cross-serial dependencies*. This is illustrated in (2b), where verbs have been coindexed with their non-verbal arguments.

(2) a. ... dat Jan het boek wil lezen
... that John the book wants read
... that John wants to read the book

b. ... dat Jan_1 Marie_2 het boek_3 wil_1 laten_2 lezen_3
... that John Mary the book wants let read
... that John wants to let Mary read the book

Finally, there is a class of verbs (containing *proberen* (*to try*) and *verzuimen* (*to fail*) inducing word orders that are a mixture of extraposition and verb-raising. Instances of such mixed word orders (3) are known as 'the third construction' (den Besten and Rutten, 1989), or as cases of *partial* extraposition (PE) (Hoeksema, 1991).

(3) a. ... dat Jan het boek aan Marie verzuimt te geven
... that John the book to Mary fails to give
... that John fails to give the book to Mary

b. ... dat Jan het boek verzuimt aan Marie te geven

c. ... dat Jan verzuimt het boek aan Marie te geven

This paper presents a categorial analysis of all the verb classes just introduced, although we will primarily be concerned with VR verbs. In particular, we demonstrate that cross-serial word order can be accounted for by a mechanism, i.e. recursive lexical constraints, whose domain of application is restricted to the lexicon and which can be implemented using feature-based formalisms. Furthermore, we show that various other word orders in the verb cluster can be accounted for by generalizing our initial formulation of the constraint. Finally, we argue that overgeneration can be avoided by requiring that the argument of a VR verb must be a 'verbal complex'.

2 Hoeksema's Categorial Analysis

Hoeksema (1991) proposes a lexicalist, categorial, analysis of Dutch verb clusters in which VR verbs are assigned polymorphic categories of the following type:

(4) willen: $(\$\backslash \text{VP})/(\$\backslash \text{VP})$
laten: $(\$\backslash (\text{NP}\backslash \text{VP}))/(\$\backslash \text{VP})$

The \$-symbol in these category-schemata represents a variable which can be instantiated by an arbitrary number of arguments, all of which must be dominated by a $\backslash$. Thus, the category of *willen* subsumes VP/VP , (NP$\backslash$VP)/(NP$\backslash$VP), and (NP$\backslash$(NP$\backslash$VP))/ (NP$\backslash$(NP$\backslash$VP)). Example (2b), for instance, is derived as follows (VP = NP$\backslash$S):

$$
(5)\quad \ldots\quad
\begin{array}{cccccc}
\textit{Jan} & \textit{Marie} & \textit{het boek} & \textit{wil} & \textit{laten} & \textit{lezen} \\
\text{NP} & \text{NP} & \text{NP} & (\$\backslash\text{VP})/(\$\backslash\text{VP}) & \$\backslash(\text{NP}\backslash\text{VP})/(\$\backslash\text{VP}) & \text{NP}\backslash\text{VP} \\
\cline{5-6}
 & & & & \multicolumn{2}{c}{\text{NP}\backslash\text{NP}\backslash\text{VP}} \\
\cline{4-6}
 & & & \multicolumn{3}{c}{\text{NP}\backslash\text{NP}\backslash\text{VP}} \\
\cline{3-6}
 & & \multicolumn{4}{c}{\text{NP}\backslash\text{VP}} \\
\cline{2-6}
 & \multicolumn{5}{c}{\text{VP}} \\
\cline{1-6}
\multicolumn{6}{c}{\text{S}}
\end{array}
$$

Extraposition verbs are simply assigned a category which selects for a VP complement on its right, thus eliminating the need for a special *extraposition*-operation:

(6) verbieden: (NP$\backslash$VP)/VP

PE verbs, finally, are treated as VR verbs, that is, they also receive a polymorphic category (7). The difference between the two verb types is that for VR verbs, cross-serial word order is obligatory, whereas for PE verbs this constraint does not hold. We will explain below how this difference is accounted for.

(7) verzuimen: (\$$\backslash$VP)/(\$$\backslash$VP)
proberen: (\$$\backslash$VP)/(\$$\backslash$VP)

The analysis of Hoeksema (1991) differs from previous categorial analyses in that it derives the possibility of cross-serial word order as a consequence of the category of certain lexical items. Other categorial analyses have accounted for cross-serial word order either by combining a non-directional flexible categorial grammar with word order constraints (Steedman, 1984; Houtman, 1984; Steedman, 1985) or by introducing disharmonic rules of composition (Moortgat, 1988; Cremers, 1993). This has the disadvantage that quite detailed and intricate constraints are needed to ensure that cross-serial word order arises only in the context of VR verbs. By using lexically assigned category schemata the latter is achieved naturally.

3 Implementing Polymorphism

Most of the machinery employed in categorial grammars can be translated into unification-based grammar formalisms without problem. Complex categories, for instance, can be represented by feature structures in which the attributes VAL, DIR, and ARG are assigned values representing the value, directionality, and argument of the complex category, respectively. Categorial rules, such as application, can be encoded as a rule rewriting a *value* as a *functor* and *argument*, with feature constraints expressing the categorial interpretation of this rule.

The variable notation used by Hoeksema, however, cannot be translated so easily, as in general there is no feature structure subsuming all possible instantiations of a \$-category. Therefore, we have chosen for a different approach. Instead of assigning polymorphic types to VR and PE verbs directly, we stipulate that these verbs are assigned all categories that can be derived from some initial category by means of applying *rightward disharmonic division* zero or more times:

(8) $X/Y \rightarrow (Z\backslash X)/(Z\backslash Y)$ **Rightward Disharmonic Division**

The initial category of *willen* is VP/VP, for instance, and thus it is assigned the categories VP/VP, (NP\VP)/(NP\VP), etc. Note that rightward disharmonic division associates *willen* with a set of categories that is identical to all possible instantiations of the polymorphic categories proposed by Hoeksema.

The reason for pointing out this equivalence is that it suggests a method for incorporating polymorphism into unification-based frameworks. The various categories of a VR verb can be defined using a recursive, relational, constraint *division*. An example of a lexical entry, as well as a first approximation of the definition of *division* is presented in (9). Note that we use definite clauses to implement lexical entries as well as constraints. Symbols starting with a capital represent variables, matrices represent feature structures, and VP represents the feature structure encoding of the corresponding linguistic category.

(9) $$\textit{lex}(\text{willen}, \text{Sign}) \text{ :- } \textit{division}(\begin{bmatrix} \textit{val}\ \text{VP} \\ \textit{dir}\ \backslash \\ \textit{arg}\ \text{VP} \end{bmatrix}, \text{Sign}).$$

$$\textit{division}(\text{In}, \text{In}).$$

$$\textit{division}(\text{In}, \begin{bmatrix} \textit{val} & \begin{bmatrix} \textit{val}\ \text{X} \\ \textit{dir}\ \backslash \\ \textit{arg}\ \text{Z} \end{bmatrix} \\ \textit{dir}\ / & \\ \textit{arg} & \begin{bmatrix} \textit{val}\ \text{Y} \\ \textit{dir}\ \backslash \\ \textit{arg}\ \text{Z} \end{bmatrix} \end{bmatrix}) \text{ :- } \textit{division}(\text{In}, \begin{bmatrix} \textit{val}\ \text{X} \\ \textit{dir}\ / \\ \textit{arg}\ \text{Y} \end{bmatrix}).$$

The *division*-relation holds if its second argument can be derived by applying the categorial rule (rightward disharmonic) division to the first argument an arbitrary number of times.

A consequence of defining lexical entries using recursive constraints is that certain entries may be infinitely ambiguous. Carpenter (1991) has shown that such grammars in general are not decidable. In Bouma and van Noord (1994) we argue that processing with the particular kind of grammar used here is possible if one interleaves the evaluation of recursive lexical constraints with the derivation of syntactic structure.

4 Extending Coverage

The definition of division presented in the previous section only allows for disharmonic division of right-directional functors. To account for inverted word orders within the verb cluster, other instances of division are also needed.

Finite modals as well as auxiliaries allow for inversion in the verb cluster, that is, the governed verb may occur to their left, rather than to the right:

(10) a.

…dat	Jan	een brief	schrijven	wil	
…that	John	a letter	write	wants	

… *that John wants to write a letter*

b.

…dat	Jan	Marie	gekust	heeft
…that	John	Mary	kissed	has

… *that John has kissed Mary*

Verb clusters with inverted word order may seperate a head from its argument, if that head selects for an argument to the right:

(11)

…dat	Jan	geleerd	heeft	zich	aan	te	passen
…that	John	learned	has	himself	PRT	to	adapt

… *that John has learned to adapt himself*

To account for inversion we assume that an auxiliary such as *hebben* is assigned an initial category VP\VP and that it is subject to *division*. Note, however, that since governors inducing inverted word order are left-directional, a leftward version of disharmonic division is needed to derive examples such as (11) (in (13) and below we use $A \Rightarrow B$ to represent an instance of *division*):

(12) $Y\backslash X \rightarrow (Y/Z)\backslash(X/Z)$ **Leftward Disharmonic Division**

(13) …

$$
\begin{array}{cccc}
\textit{Jan} & \textit{geleerd} & \textit{heeft} & \textit{zich aan te passen}\\
\text{NP} & \text{VP/VP} & \text{VP}\backslash\text{VP} & \text{VP}\\
 & & \Downarrow & \\
 & & (\text{VP/VP})\backslash(\text{VP/VP}) & \\ \cline{2-3}
 & \multicolumn{2}{c}{\text{VP/VP}} & \\ \cline{2-4}
 & & \multicolumn{2}{c}{\text{VP}}
\end{array}
$$

Next, consider cases in which the governed verb is an extraposition verb selecting an NP-object:

(14)

…dat	Jan	Marie	beloofd	heeft	de diskette	mee	te	brengen
…that	John	Mary	promised	has	the diskette	PRT	to	bring

… *that John has promised Mary to bring along the diskette*

The category of *beloven* is $(\textsc{np}\backslash\textsc{vp})/\textsc{vp}$. This implies that the auxiliary can only combine with the preceding participle if it can undergo leftward harmonic (15) as well as disharmonic division:

(15) $Y\backslash X \rightarrow (Z\backslash Y)\backslash(Z\backslash X)$ **Leftward Harmonic Division**

(16) … *beloofd* *heeft* *de diskette mee te brengen*

$$
\dfrac{\dfrac{(\textsc{np}\backslash\textsc{vp})/\textsc{vp} \qquad \dfrac{\dfrac{\textsc{vp}\backslash\textsc{vp}}{(\textsc{np}\backslash\textsc{vp})\backslash(\textsc{np}\backslash\textsc{vp})}\Downarrow}{((\textsc{np}\backslash\textsc{vp})/\textsc{vp})\backslash((\textsc{np}\backslash\textsc{vp})/\textsc{vp})}\Downarrow}{(\textsc{np}\backslash\textsc{vp})/\textsc{vp}} \qquad \textsc{vp}}{\textsc{np}\backslash\textsc{vp}}
$$

For modal verbs, inversion is possible only if the modal is finite. For auxiliaries, however, inversion is also possible if the auxiliary is non-finite:

(17) … dat Jan geleerd moet hebben zich aan te passen
… that John learned must have himself PRT to adapt
… *that John must have learned to adapt himself*

Note that in these cases the participle does not occur left-adjacent to the auxiliary, but at the left-periphery of the verb-cluster. Assuming that *hebben* is assigned the category $(\textsc{vp}/\textsc{vp})\backslash(\textsc{vp}/\textsc{vp})$ (as in (13)), a derivation of (17) is possible only if the VR verb *moet* can undergo rightward harmonic division:

(18) $X/Y \rightarrow (X/Z)/(Y/Z)$ **Rightward Harmonic Division**

(19) … *geleerd* *moet* *hebben* …

$$
\dfrac{\textsc{vp}/\textsc{vp} \qquad \dfrac{\dfrac{\dfrac{\textsc{vp}/\textsc{vp}}{(\textsc{vp}/\textsc{vp})/(\textsc{vp}/\textsc{vp})}\Downarrow}{((\textsc{vp}/\textsc{vp})\backslash(\textsc{vp}/\textsc{vp}))/((\textsc{vp}/\textsc{vp})\backslash(\textsc{vp}/\textsc{vp}))}\Downarrow \qquad \dfrac{\textsc{vp}\backslash\textsc{vp}}{(\textsc{vp}/\textsc{vp})\backslash(\textsc{vp}/\textsc{vp})}\Downarrow}{(\textsc{vp}/\textsc{vp})\backslash(\textsc{vp}/\textsc{vp})}}{\textsc{vp}/\textsc{vp}}
$$

The upshot of the discussion in this section is that leftward and rightward, disharmonic as well as harmonic, division is needed to account for the word order posibilities within the verb cluster. Note that all four instances of division can be captured in a unification-based setting by generalizing the previous definition in (9) as shown in (20). From now on, we assume that VR and PE verbs are subject to this general constraint.

(20) $division(\mathrm{In}, \mathrm{In}).$

$$division(\mathrm{In}, \begin{bmatrix} val & \begin{bmatrix} val\ \mathrm{X} \\ dir\ \mathrm{D}_1 \\ arg\ \mathrm{Z} \end{bmatrix} \\ dir\ \mathrm{D}_2 & \\ arg & \begin{bmatrix} val\ \mathrm{Y} \\ dir\ \mathrm{D}_1 \\ arg\ \mathrm{Z} \end{bmatrix} \end{bmatrix}) \mathrel{:\!-} division(\mathrm{In}, \begin{bmatrix} val\ \mathrm{X} \\ dir\ \mathrm{D}_2 \\ arg\ \mathrm{Y} \end{bmatrix}).$$

5 Avoiding Overgeneration

The analysis of Dutch verb-clusters presented so far has demonstrated that cross-serial word order can be accounted for by means of a lexical constraint equivalent to a generalized version of division. The account still overgenerates, however, as nothing *forces* cross-serial word order for VR verbs. In this section we argue that overgeneration can be avoided by requiring that VR verbs take a *verbal complex* as argument and by introducing a feature that marks certain constituents as not being a *verbal complex*.

A VR or PE verb may combine with the verbal head of its complement before combining with the (non-verbal) arguments of that head, but nothing in the analysis presented so far enforces that this is obligatory. For PE verbs this is fine, as in these cases both the cross-serial and the extraposed word order is allowed, as well as cases that are a mixture of both (the 'partially extraposed' VP appears is italic):

(21) a. ...dat Jan *de trainer van zijn gelijk* meent *te hebben overtuigd*
...that John the trainer of his right believes to have convinced
...that John believes to have convinced the trainer that he was right

b. ...dat Jan meent *de trainer van zijn gelijk te hebben overtuigd*

c. ...dat Jan *de trainer* meent *van zijn gelijk te hebben overtuigd*

For VR verbs, however, only the cross-serial word order is grammatical:

(22) a. ...dat Jan *de trainer van zijn gelijk* schijnt *te hebben overtuigd*
...that John the trainer of his right seems to have convinced
...that John seems to have convinced the trainer that he was right

b. * ...dat Jan schijnt *de trainer van zijn gelijk te hebben overtuigd*

c. * ...dat Jan *de trainer* schijnt *van zijn gelijk te hebben overtuigd*

The constraint we have failed to incorporate is that for VR verbs, cross-serial word order is obligatory. In other words, a VR verb must combine with the verbal head of its complement, before this head has been combined with a complement of, for instance, category NP or PP.

In previous work (Houtman, 1984; Moortgat, 1988; Hoeksema, 1991) it has been suggested that VR verbs take an argument that must be a lexical item. We believe that this suggestion is problematic for a number of reasons.

First of all, whereas it is usually the case that only cross-serial word order is grammatical, there are important exceptions.

(23) a. ...dat Jan Marie durft aan te spreken
...that John Mary dare PRT to speak
...that John dares to speak to Mary

b. ...dat Jan Marie aan durft te spreken

The particle *aan* in (23a) is a so-called seperable prefix of the verb *aanspreken*, which is illustrated in (23b) by the fact that a VR verb may seperate the prefix/particle from its root. One could argue that eventhough *aan* in (23b) is clearly a particle, *aan te spreken* in (23a) is a morphologically complex verb, so that this example is not immediately problematic for an account that requires that the verbal argument of VR verbs must be lexical.

Next, consider (24), where an auxiliary has been added to the verb cluster. In these cases, the particle may occur in three positions.

(24) a. ...dat Jan Marie heeft durven aan te spreken
...that John Mary has dare PRT to speak
...that John has dared to speak to Mary

b. ...dat Jan Marie heeft aan durven te spreken

c. ...dat Jan Marie aan heeft durven te spreken

Example (24b) is problematic if VR verbs take lexical arguments. A verb cluster V_1 V_2 V_3, with V_1 and V_2 VR verbs, must be left-branching (i.e. $[[V_1\ V_2]\ V_3]$), as otherwise V_1 would not take a lexical argument.[1] For (24b), however, left-branching implies that *heeft* has to combine with the particle *aan* before combining with its verbal argument. Such a derivation is not possible given the rules presented in the previous section.

A related problem for analyses requiring that VR verbs take lexical arguments is the fact that there is a - somewhat heterogeneous - class of complements that can occur to the right of a VR verb. Some examples are listed below (see also ANS (Geerts et al., 1984), p. 1012 ff.):

(25) a. ...dat deze stichting veel goeds heeft tot stand gebracht
...that this foundation much good has [IDIOM] brought
...that this foundation has done a lot of good things

[1]The alternative is to give verb clusters lexical status as well. We discuss this option below.

b. ... dat de man zich aan zwendel heeft schuldig gemaakt
... that the man REFL. on fraud has guilty made
... that this man is guilty of fraud

c. ... dat Deng het de Russen heeft duidelijk gemaakt
... that Deng it the Russians has clear made
... that Deng has made it clear to the Russians

While there are many idiomatic expressions in this class, it seems unlikely that phrases such as *tot stand gebracht, schuldig gemaakt, duidelijk gemaakt* are in fact lexical items.[2] Note also that, as with particles, cross-serial word orders are possible as well (and, for some speakers, are preferred):

(26) a. ... dat deze stichting veel goeds tot stand heeft gebracht

b. ... dat deze man zich aan zwendel schuldig heeft gemaakt

c. ... dat Deng het de Russen duidelijk heeft gemaakt

These can only be derived if the idiomatic expressions involved are not treated as lexical units.

Second, an account which assumes that the arguments of VR verbs are lexical, must either assume that complex verb clusters are left-branching or else that verb clusters themselves are derived lexically. We already pointed out that a left-branching analysis of verb-clusters cannot easily account for the distribution of seperable verb-prefixes or other complements within the verb-cluster.[3] An analysis which assumes that the verb cluster itself is derived lexically (as has been proposed in Moortgat (1988)) is not very plausible either. The examples in (25) can only be explained by assuming that quite complex phrases can have lexical status. Furthermore, if verb clusters are complex lexical items, it remains to be explained why parts of it can be fronted (27) and why coordination of parts of the verb cluster is possible (28).[4] Note that in example (28c) (from Steedman (1985)), the conjuncts consist of two NP's and a verb. In an analysis where verb clusters are derived lexically, this would require that coordination may compose phrasal and (sub-)lexical elements into one conjunct.

(27) werken zou Piet nooit willen
work should Pete never want
Pete should never want to work

[2]In southern dialects of Dutch, the possibility of NP or PP arguments violating cross-serial word order is clearly not restricted to idiomatic expressions only (Haegeman and van Riemsdijk, 1986).

[3]Most analyses outside the categorial tradition assume a right-branching structure for the verb cluster. Evers (1975), Bresnan et al. (1983), and den Besten and Edmondson (1983) argue for right brancing structures, but some of their arguments are questioned in Kroch and Santorini (1987), who present a left-branching analysis instead.

[4]Example (28b) is taken from Steedman (1985), who presents it as an argument against Bresnan et al. (1983), who consider a structurally similar example ungrammatical.

(28) a. ... dat Jan Marie heeft zien zwemmen en horen zingen
... that John Mary has seen swim and hear sing
... that John has seen Mary swimming and has heard Mary singing

b. ... dat ik Henk de kinderen hoorde leren en zag helpen zwemmen
... that I Hank the children heard learn and saw help swim
... that I heard Hank learn the children to swim and saw Hank help the children to swim

c. ... dat ik Cecilia de nijlpaarden zag en Henk de olifanten hoorde wassen
... that I Cecilia the hippo's saw and Hank the elephants heard wash
... that I saw Cecilia wash the hippo's and heard Hank wash the elephants

To avoid the problems that arise form the assumption that the arguments of VR verbs must be lexical, we impose the constraint that the verbal argument of a VR verb must be a *verbal complex*. Assuming a feature VC to distinguish between verbal complexes and other (verbal) constituents, we can implement the relevant restriction as follows (where the subscribt *Sign* indicates a reentrancy with the second argument of *division*):

(29) $$lex(\text{willen}, \begin{bmatrix} arg \begin{bmatrix} vc\ + \end{bmatrix} \end{bmatrix}_{Sign}) \texttt{ :- } division(\begin{bmatrix} val\ \text{VP} \\ dir\ \backslash \\ arg\ \text{VP} \end{bmatrix}, \text{Sign}).$$

The next, and more difficult, matter that needs to be settled is when exactly a constituent is marked as $-$VC. Consider a simple example.

(30) a. ... *een boek* *wil* *lezen*

$$\dfrac{\text{NP} \quad \dfrac{(\text{NP}\backslash\text{VP})/(\text{NP}\backslash\text{VP})_{[+\text{VC}]} \quad \text{NP}\backslash\text{VP}_{[\text{-VC}]}}{\text{NP}\backslash\text{VP}}}{\text{VP}}$$

b. *... *wil* *een boek* *lezen*

$$\dfrac{\text{VP}/\text{VP}_{[+\text{VC}]} \quad \dfrac{\text{NP} \quad \text{NP}\backslash\text{VP}_{[\text{-VC}]}}{\text{VP}_{[\text{-VC}]}}}{***}$$

Example (30a) is acceptable, whereas (30b) needs to be ruled out. A simple way to achieve this would be to assign the category NP\VP[-VC] to *lezen*. This blocks the derivation of (30b), since *een boek lezen* is now of category VP[-VC], which is not unifiable with the argument of *wil*. (30a), on the other hand is still derivable, as NP\VP[-VC] and (NP\VP)[+VC] are unifiable (note that the category of *lezen* only specifies that its *value* is -VC).

This solution does not work for more complex examples, however. Consider for instance the following example:

(31) *...

$$\begin{array}{cccc}
\textit{heeft} & \textit{een boek} & \textit{willen} & \textit{lezen}\\
\mathrm{VP/VP}_{[+\mathrm{VC}]} & \mathrm{NP} & (\mathrm{NP}\backslash\mathrm{VP})/(\mathrm{NP}\backslash\mathrm{VP})_{[+\mathrm{VC}]} & \mathrm{NP}\backslash\mathrm{VP}_{[-\mathrm{VC}]}\\ \cline{3-4}
 & & \multicolumn{2}{c}{\mathrm{NP}\backslash\mathrm{VP}}\\ \cline{2-4}
 & \multicolumn{3}{c}{\mathrm{VP}}
\end{array}$$

If we add the specification -VC to the value of *lezen* only, the complex phrase *een boek willen lezen* will remain unspecified for the feature VC. One might consider solving this problem by adding the specification -VC to the value of VR verbs. But this would block the derivation of right-branching verb clusters, an option we argued against above. For instance, this solution would block the derivation of verb clusters containing particles:

(32) ...

$$\begin{array}{cccc}
\textit{heeft} & \textit{op} & \textit{moeten} & \textit{schieten}\\
\textit{has} & \mathrm{PRT} & \textit{must} & \textit{hurry}\\
\mathrm{VP/VP}_{[+\mathrm{vc}]} & \mathrm{PRT} & (\mathrm{PRT}\backslash\mathrm{VP}_{[-\mathrm{VC}]})/(\mathrm{PRT}\backslash\mathrm{VP})_{[+\mathrm{VC}]} & \mathrm{PRT}\backslash\mathrm{VP}\\ \cline{3-4}
 & & \multicolumn{2}{c}{\mathrm{PRT}\backslash\mathrm{VP}_{[-\mathrm{VC}]}}\\ \cline{2-4}
 & \multicolumn{3}{c}{\mathrm{VP}_{[-\mathrm{VC}]}}
\end{array}$$

In (32), the specification -VC is added to the value category of *moeten*. Since particles can occur as part of the verb cluster, no specification -VC is added to the value of *schieten*. The derivation is blocked nevertheless, as *op moeten schieten* is -VC, irrespective of the fact that the value of *schieten* is unmarked for VC.

The problem with the solutions just considered is that they fail to take into account the fact that it is the argument of a verb which usually determines whether the result of combining that verb with its argument is -VC or not. This information should be preserved if a verb does not combine with its argument directly, but instead is combined with a VR verb that 'inherits' the arguments of the verbs it governs. One way to achieve this is to extend the rule of *division* with a constraint that makes the values of VC on the value categories of functor and argument reentrant:

(33) $\textit{division}(\mathrm{In},\mathrm{In}).$

$$\textit{division}\left(\mathrm{In},\ \begin{bmatrix} \textit{val} & \begin{bmatrix} \textit{val} & \begin{bmatrix} \textit{vc}\ \mathrm{V} \end{bmatrix}_X \\ \textit{dir} & \mathrm{D}_1 \\ \textit{arg} & \mathrm{Z} \end{bmatrix} \\ \textit{dir} & \mathrm{D}_2 \\ \textit{arg} & \begin{bmatrix} \textit{val} & \begin{bmatrix} \textit{vc}\ \mathrm{V} \end{bmatrix}_Y \\ \textit{dir} & \mathrm{D}_1 \\ \textit{arg} & \mathrm{Z} \end{bmatrix} \end{bmatrix}\right) :\!-\ \textit{division}\left(\mathrm{In},\ \begin{bmatrix} \textit{val}\ \mathrm{X} \\ \textit{dir}\ \mathrm{D}_2 \\ \textit{arg}\ \mathrm{Y} \end{bmatrix}\right).$$

The derivation of simple as well as more complex verb clusters now proceeds as required. Example (30a) is still derivable, whereas (30b) is not:

(34) a. … *een boek* *wil* *lezen*

$$\dfrac{\mathrm{NP} \qquad \dfrac{(\mathrm{NP}\backslash\mathrm{VP}_{[\alpha\mathrm{VC}]})/(\mathrm{NP}\backslash\mathrm{VP}_{[\alpha\mathrm{VC}]})_{[+\mathrm{VC}]} \qquad \mathrm{NP}\backslash\mathrm{VP}_{[-\mathrm{VC}]}}{\mathrm{NP}\backslash\mathrm{VP}_{[-\mathrm{VC}]}}}{\mathrm{VP}_{[-\mathrm{VC}]}}$$

b. *… *wil* *een boek* *lezen*

$$\dfrac{\mathrm{VP}/\mathrm{VP}_{[+\mathrm{VC}]} \qquad \dfrac{\mathrm{NP} \qquad \mathrm{NP}\backslash\mathrm{VP}_{[-\mathrm{VC}]}}{\mathrm{VP}_{[-\mathrm{VC}]}}}{***}$$

Note that the phrase *een boek wil lezen* is marked -VC. This implies that the ungrammatical example (31) is no longer derivable:

(35) *… *heeft* *een boek willen lezen*

$$\dfrac{\mathrm{VP}/\mathrm{VP}_{[+\mathrm{VC}]} \qquad \dfrac{\textit{een boek willen lezen}}{\mathrm{VP}_{[-\mathrm{VC}]}}}{***}$$

For complex verb clusters containing a particle, however, no problem arises. In particular, the word order in which the particle appears in the middle of the verb cluster (i.e neither at the left-periphery nor adjacent to its governing verb) remains derivable:

(36) … *heeft* *op* *moeten* *schieten*
has PRT *must* *hurry*

$$\dfrac{\mathrm{VP}/\mathrm{VP}_{[+\mathrm{VC}]} \qquad \dfrac{\mathrm{PRT} \qquad \dfrac{(\mathrm{PRT}\backslash\mathrm{VP}_{[\alpha\mathrm{VC}]})/(\mathrm{PRT}\backslash\mathrm{VP}_{[\alpha\mathrm{VC}]})_{[+\mathrm{VC}]} \qquad \mathrm{PRT}\backslash\mathrm{VP}}{\mathrm{PRT}\backslash\mathrm{VP}}}{\mathrm{VP}}}{\mathrm{VP}}$$

The analysis can be extended to verb clusters containing a modal or auxiliary inducing inverted word order. In section 4 we observed that if such a verb selects an extraposition verb, which in its turn selects for a full VP to its right, only cross-serial word order is allowed:

(37) a. … dat Jan leren moet zijn mond te houden
… that John learn must to keep his mouth shut
… *that John must learn to keep his mouth shut*

b. * … dat Jan leren zijn mond te houden moet

This suggests that inverted modals and auxiliaries are like other VR verbs in that they select a +VC argument[5] and that extraposition verbs give rise to -VC constituents:

[5]Note that a 'side-effect' of this assumption is that it eliminates a spurious ambiguity that might

(38) ... *leren* *moet* *zijn mond te houden*

$$\dfrac{\dfrac{\mathrm{VP}_{[\text{-VC}]}/\mathrm{VP} \qquad (\mathrm{VP}_{[\alpha\mathrm{VC}]}/\mathrm{VP})_{[+\mathrm{VC}]}\backslash(\mathrm{VP}_{[\alpha\mathrm{VC}]}/\mathrm{VP})}{\mathrm{VP}_{[\text{-VC}]}/\mathrm{VP}} \qquad \mathrm{VP}}{\mathrm{VP}_{[\text{-VC}]}}$$

(39) *... *leren* *zijn mond te houden* *moet*

$$\dfrac{\dfrac{\mathrm{VP}_{[\text{-VC}]}/\mathrm{VP} \qquad \mathrm{VP}}{\mathrm{VP}_{[\text{-VC}]}} \qquad \mathrm{VP}_{[+\mathrm{VC}]}\backslash\mathrm{VP}}{***}$$

A slightly more complex situation arises if the verb inducing inverted word order is itself governed by a VR verb. This is possible for auxiliaries, but not for modals, as with the latter the possibility of inversion is restricted to finite forms. The examples below show that in such cases, the participle selected by the auxiliary must occur at the left periphery of the verb cluster:

(40) a. ...dat Jan het boek gelezen moet hebben
...that John the book read must have
...*that John must have read the book*

b. * ...dat Jan het boek moet gelezen hebben

To account for the ungrammaticality of (40b), we must assume that *hebben* not only selects a +VC argument but also that it marks its value as -VC:

$$(41)\ \ lex(\text{hebben}, \begin{bmatrix} arg & \begin{bmatrix} vc\ + \end{bmatrix} \\ val & \begin{bmatrix} vc\ \text{-} \end{bmatrix} \end{bmatrix}_{Sign}) \text{ :- } division(\begin{bmatrix} val\ \mathrm{VP} \\ dir\ \backslash \\ arg\ \mathrm{VP} \end{bmatrix}, \text{Sign}).$$

Under this assumption, the derivation of the verb cluster in (40b) is blocked (TV abbreviates NP\VP):

---

otherwise arise if we derive a phrase consisting of a transitive verb governed by an inverted modal or auxiliary:

i. ... *een boek* *lezen* *wil*

$$\dfrac{\mathrm{NP} \qquad \dfrac{\mathrm{NP}\backslash\mathrm{VP}_{[\text{-VC}]} \qquad (\mathrm{NP}\backslash\mathrm{VP})\backslash(\mathrm{NP}\backslash\mathrm{VP})}{\mathrm{NP}\backslash\mathrm{VP}}}{\mathrm{VP}}$$

ii. ... *een boek* *lezen* *wil*

$$\dfrac{\dfrac{\mathrm{NP} \qquad \mathrm{NP}\backslash\mathrm{VP}_{[\text{-VC}]}}{\mathrm{VP}_{[\text{-VC}]}} \qquad \mathrm{VP}\backslash\mathrm{VP}}{\mathrm{VP}}$$

The right-branching derivation in (ii) is ruled out if *willen* requires a verbal complex as argument.

(42) *... *moet* *gelezen* *hebben*

$$
\begin{array}{llll}
{}^{*}\ldots & \textit{moet} & \textit{gelezen} & \textit{hebben} \\
 & \mathrm{TV}/\mathrm{TV}_{[+\mathrm{VC}]} & \mathrm{TV} & \mathrm{TV}_{[+\mathrm{VC}]}\backslash\mathrm{TV}_{[-\mathrm{VC}]} \\
\cline{3-4}
 & & \multicolumn{2}{c}{\mathrm{TV}_{[-\mathrm{VC}]}} \\
\cline{2-3}
 & \multicolumn{2}{c}{***} & \\
\end{array}
$$

whereas the verb cluster in (40a) can be derived as follows:

(43) ... *gelezen* *moet* *hebben*

$$
\begin{array}{llll}
\ldots & \textit{gelezen} & \textit{moet} & \textit{hebben} \\
\ldots & \mathrm{TV} & (\mathrm{TV}\backslash\mathrm{TV}_{[\alpha\mathrm{VC}]})/(\mathrm{TV}\backslash\mathrm{TV}_{[\alpha\mathrm{VC}]})_{[+\mathrm{VC}]} & \mathrm{TV}_{[+\mathrm{VC}]}\backslash\mathrm{TV}_{[-\mathrm{VC}]} \\
\cline{3-4}
 & & \multicolumn{2}{c}{\mathrm{TV}_{[+\mathrm{VC}]}\backslash\mathrm{TV}_{[-\mathrm{VC}]}} \\
\cline{2-3}
 & \multicolumn{2}{c}{\mathrm{TV}} & \\
\end{array}
$$

6 Conclusions

In this paper we have demonstrated that a lexicalist account of the Dutch verb cluster is able to deal with a variety of word order patterns typical for this construction, while at the same time overgeneration can be avoided.

There have been quite a number of proposals for a categorial analysis of Dutch cross-serial depedendencies. We believe that the proposal outlined above, which uses either polymorphic lexical category assignments (Hoeksema, 1991) or recursive lexical constraints, is promising as it directly relates cross-serial word order to certain lexical entries. This is clearly desirable as the possibility of 'verb raising' is definitely a lexical property. Furthermore, since the mechanims used to derive cross-serial word order are located in the lexicon, the interaction with general syntactic rules is restricted and thus there is no need for complex word order constraints to ensure that crossing word orders arise only in the context of 'verb raising' verbs.

References

Bouma, Gosse and Gertjan van Noord. 1994. Constraint-based categorial grammar. In *32th Annual Meeting of the Association for Computational Linguistics*, Las Cruces.

Bresnan, Joan W., Ronald M. Kaplan, Stanley Peters, and Annie Zaenen. 1983. Cross-serial dependencies in dutch. *Linguistic Inquiry*, 13.

Carpenter, Bob. 1991. The generative power of categorial grammars and head-driven phrase structure grammars with lexical rules. *Computational Linguistics*, 17(3):301–313.

Cremers, Crit. 1993. *On Parsing Coordination Categorially.* Ph.D. thesis, Rijksuniversiteit Leiden. HIL dissertations 5.

den Besten, Hans and J.A. Edmondson. 1983. The verbal complex in continental west germanic. In Werner Abraham, editor, *On the Formal Syntax of West Germanic*. Benjamins Amsterdam.

den Besten, Hans and Jean Rutten. 1989. On verb raising, extraposition, and free word order in Dutch. In D. Jaspars, W. Klooster, Y.Putseys, and P. Seuren, editors, *Sentential Complementation and the Lexicon*. Foris, Dordrecht.

Evers, Arnold. 1975. *The Transformational Cycle in Dutch and German*. Ph.D. thesis, Rijksuniversiteit Utrecht.

Geerts, G., W. Haeseryn, J. de Rooij, and M.C. van den Toorn. 1984. *Algemene Nederlandse Spraakkunst*. Groningen: Wolters-Noordhoff.

Haegeman, Liliane and Henk van Riemsdijk. 1986. Verb projection raising, scope, and the typology of verb movement rules. *Linguistic Inquiry*, 17:417–66.

Hoeksema, Jack. 1991. A categorial theory of reanalysis phenomena. ms. RUG Groningen.

Houtman, Joop. 1984. Een kategoriale beschrijving van het Nederlands. *TABU*, 14(1):1–27.

Kroch, Anthony S. and Beatrice Santorini. 1987. The derived constituent structure of the west germanic verb raising construction. In *Proceedings of the Princeton Workshop on Comparative Grammar*. MIT Cambridge Mass.

Moortgat, Michael. 1988. *Categorial Investigations*. Ph.D. thesis, University of Amsterdam.

Steedman, Mark. 1984. A categorial theory of intersecting dependencies in Dutch infinitival complements. In W. de Geest and Y. Putseys, editors, *Sentential Complementation*. Foris, Dordrecht, pages 215–226.

Steedman, Mark. 1985. Dependency and coordination in the grammar of Dutch and English. *Language*, 61:523–68.