Complement Extraction Lexical Rules and Argument Attraction

Stefan Müller

October, 1996

Complement Extraction Lexical Rules and Argument Attraction

Stefan Müller

A slightly different version of this paper appeared in proceedings of Konvens 96. The case principle in the current version of this paper has been simplified.

Complement Extraction Lexical Rules and Argument Attraction

Stefan Müller[*]

Stefan.Mueller@dfki.de
http://www.dfki.de/~stefan/

DFKI GmbH
Stulsatzenhausweg 3
66123 Saarbrücken

October, 1996

Abstract

In (Müller, 1994), several problems with the processing of complement extraction lexical rules (CELRs) (Pollard and Sag, 1994, Ch. 9) are discussed. CELRs destroy the ordering information implicitly contained in the COMPS list, making it impossible to assign structural case without default mechanisms (which would be an extension of the basic HPSG formalism). Furthermore, it can be shown that the ARG-S list—a static list suggested by Pollard and Sag to take over the function of the COMPS list in Binding Theory—cannot be used for case assignment.

This paper aims to demonstrate how these problems can be solved by use of a dynamically constructed ARG-S list. Together with a modified schema for the construction of German verb clusters, this dynamic list allows for an application of CELRs to both auxiliaries and main verbs, which explains the frontability of subjects and *von*-PPs in passive constructions, as well as the frontability of subjects in perfect constructions.

The proposed account also solves some implementational problems caused by a variable COMPS list.

In (Müller, 1994) habe ich viele Probleme, die sich aus der Verwendung von Komplementextraktionsregeln ergeben, diskutiert. CELR zerstören die Information, die implizit in der Reihenfolge der Elemente der COMPS-Liste enthalten ist. Dadurch wird die Zuweisung strukturellen Kasus unter Bezugnahme auf die COMPS-Liste unmöglich. Die ARG-S-Liste, die von Pollard und Sag für die Übernahme der Funktionen der COMPS-Liste in der Bindungstheorie vorgeschlagen wurde, kann für die Kasungszuweisung nicht benutzt werden.

Ich zeige, wie die erwähnten Probleme mit CELR durch die Verwendung einer dynamischen ARG-S-Liste gelöst werden können. Zusammen mit einem modifizierten Schema für den Verbalkomplex gestattet eine solche dynamische Liste die Anwendung der CELR auf Hilfs- und Vollverben. Damit ist sowohl die Voranstellbarkeit von Subjekten und *von*-PPs in Passivkonstruktionen als auch die Voranstellbarkeit von Subjekten in Perfektkonstruktionen erklärbar.

Die vorgeschlagene Analyse löst auch einige Implementationsprobleme, die durch variable COMPS-Listen entstehen.

1 Introduction

As Pollard and Sag (1994, Ch. 9.5) point out CELRs are a psycholinguistically motivated means for describing extraction phenomena. The CELR proposed there can be successfully applied in English grammars. However, the rule stated by Pollard and Sag is not equivalent to the trace solution described in Chapter 4 of their book. Rather, the following rule should be used:

$$
\begin{bmatrix}
\text{SYNSEM} & \begin{bmatrix}
\text{LOC} & \begin{bmatrix} \text{CAT}|\text{COMPS}\boxed{1} \oplus \left\langle \boxed{2} \right\rangle \oplus \boxed{3} \end{bmatrix} \\[2em]
\text{NONLOC} & \begin{bmatrix} \text{INHER}|\text{SLASH}\boxed{4} \end{bmatrix}
\end{bmatrix} \\[2em]
\textit{lexical-sign}
\end{bmatrix} \rightarrow
$$

$$
\begin{bmatrix}
\text{SYNSEM} & \begin{bmatrix}
\text{LOC} & \begin{bmatrix} \text{CAT}|\text{COMPS}\boxed{1} \oplus \boxed{3} \end{bmatrix} \\[2em]
\text{NONLOC} & \begin{bmatrix} \text{INHER}|\text{SLASH}\boxed{4} \oplus \left\langle \boxed{5} \right\rangle \end{bmatrix}
\end{bmatrix} \\[2em]
\textit{lexical-sign}
\end{bmatrix} \tag{1}
$$

Where $\boxed{2}$ corresponds to the structure in (2).

$$
\begin{bmatrix}
\text{LOC} & \boxed{5} \\[1em]
\text{NONLOC} & \begin{bmatrix} \text{INHER} & \begin{bmatrix} \text{QUE} & \langle\rangle \\ \text{REL} & \langle\rangle \\ \text{SLASH} & \left\langle \boxed{5} \right\rangle \end{bmatrix} \end{bmatrix} \\[2em]
\textit{synsem}
\end{bmatrix} \tag{2}
$$

The specification of the REL and SLASH values in $\boxed{2}$ blocks the extraction of elements with a specified REL value, as for instance the first complement of the empty relativizer[1] and the extraction of complements with either an empty set as SLASH value or with an element in the SLASH set that differs from the LOCAL value of the complement ($\boxed{5}$). *Tough*-adjectives[2] are an instance of the latter case.

---

[1] (Pollard and Sag, 1994, p. 216)

[2] Pollard and Sag (1994, p. 167) assume that *tough*-adjectives subcategorize for an infinitive VP with a non-empty SLASH set.

The rule (1) changes the position of complements in the COMPS list. The second element may become the first one if the first one is extracted. The position in the COMPS list was employed by several authors to account for case phenomena. I have formulated an agreement principle that refers to the first position of the COMPS list (Müller, 1997a).

While saturation by traces does not affect the position of elements in the COMPS list, extraction by a rule does. This problem will be solved in the present paper.

2   German V2

In almost all German HPSG-Analyses, verb-second (V2) sentences are analyzed by means of unbounded dependency constructions.[3] That is, the basis of a V2 sentence is a verb-initial sentence in which one constituent is topicalized. This constituent can originate from an arbitrarily deep level of embedding. Subjects, objects and more oblique complements can be topicalized.

(3) a. Der Mann   hat   den Hund   geschlagen.
the man$_{nom}$ has   the dog$_{acc}$ beaten
'The man beat the dog.'

   b. Von wem   wurde   dieses Buch   geschrieben?
by whom   was   this book$_{nom}$ written
'Who wrote this book?'

As a consequence, no distinction is made between the subject and the other complements in finite clauses. In most HPSG analyses for German, a subject insertion lexical rule (SILR) is used to generate finite forms having their subject, i.e., the value of the SUBJ list, on the COMPS list. The CELR then applies to such lexical entries.

3   Case and Passive in German

In many theories on German, there is a distinction between structural and lexical case (see Heinz and Matiasek (1994) for an introduction and detailed examples for structural and lexical case). Case assignment interacts with complement

---

Note that sentences like (ii) are not covered by the trace analysis in standard HPSG.

(i)   John is easy [to talk to $_i$].

(ii)   [To talk to $_i$]$_j$, John is easy $_j$ (but to work with, he's not).

The INHER|SLASH value of the infinitive phrase is an accusative NP and does not unify with the LOCAL value of the infinitive phrase itself. If one would ignore the NONLOCAL values in the rule (1) this would allow sentences like (ii) but sentences like (iii) would be possible as well.

(iii)   * [To talk to John]$_j$, John is easy $_j$

[3]In GPSG analyses this is assumed as well (cf. (Uszkoreit, 1987)).

saturation.[4]

In what follows, I will present an account of German passive that combines work done by several people. Firstly, there is the case principle proposed by Heinz and Matiasek (1994), secondly, the passive analysis using an ERG feature as suggested by Pollard (1994). The case principle is adapted for grammars that use a SUBJ feature to represent the subject of non-finite verbs instead of including the subject into the COMPS list.

Pollard (1994) gives a unified account for the German passive which can explain both personal and impersonal constructions. He gives lexical entries for the perfect auxiliary *haben* (have) and for the passive auxiliary *werden* (be) as follows (slightly modified):[5],[6]

haben:

$$
\begin{bmatrix}
\text{HEAD} & \begin{bmatrix} \text{SUBJ}\ \boxed{1} \\ \textit{verb} \end{bmatrix} \\
\text{ERG} & \langle\rangle \\
\text{COMPS} & \boxed{2} \\
\text{VCOMP} & V[\text{LEX+},\ \textit{ppp},\ \text{SUBJ}\ \boxed{1},\ \text{COMPS}\ \boxed{2},\ \text{VCOMP}\ \textit{none}] \\
\textit{cat} &
\end{bmatrix}
\tag{4}
$$

werden:

$$
\begin{bmatrix}
\text{HEAD} & \begin{bmatrix} \text{SUBJ}\ \boxed{1} \\ \textit{verb} \end{bmatrix} \\
\text{ERG} & \boxed{1} \\
\text{COMPS} & \boxed{2} \oplus \big\langle\ (\text{PP}[\textit{von}]{:}\boxed{3}\textit{ref})\ \big\rangle \\
\text{VCOMP} & V[\text{LEX+},\ \textit{ppp},\ \text{SUBJ}\ \big\langle\ \text{NP}[\textit{str}]{:}\boxed{3}\ \big\rangle,\ \text{ERG}\ \boxed{1},\ \text{COMPS}\ \boxed{1} \oplus \boxed{2}, \\
& \qquad\qquad \text{VCOMP}\ \textit{none}] \\
\textit{cat} &
\end{bmatrix}
\tag{5}
$$

Note that the verbal complement is not part of the COMPS list, but is listed as the value of a separate feature instead. This was proposed by Chung for Korean and by Rentier (1994) for Dutch. It is reasonable for German as well because, as was shown by Hinrichs and Nakazawa (1994a), the verb cluster has to be built up before any nonverbal complement can be saturated. Hinrichs and Nakazawa

---

[4]See (Müller, 1994) for a discussion of several approaches. As noted there, Pollard (1994) sketches a different approach employing defaults. Defaults are not part of the basic HPSG formalism.

[5]I assume that SUBJ is a head feature since the SUBJ value must be projected to account for control phenomena.

Pollard does not use the VCOMP feature.

[6]Note that such an analysis presupposes that the elements on the COMPS list are arranged in order of obliqueness (cf. (Pollard and Sag, 1987, p. 120)), i.e:

SUBJECT => PRIMARY OBJECT => SECOND OBJECT => OTHER COMPLEMENTS

use a feature NPCOMP to prevent the saturation of nonverbal complements before the verbal cluster is built. However, the use of VCOMP turned out to be the more elegant way.[7]

The complements of the perfect and the passive auxiliaries have the same morphological form. As can be seen by comparing (4) and (5), the difference between the perfect and passive construction manifests itself in the set of complements taken over from the embedded verb. The form of the verbal complements is *ppp* in both cases.

(6) a. Der Mann   hat den Hund geschlagen.
       the man$_{nom}$ has the dog$_{acc}$ beaten
       'The man has beaten the dog.'

   b. Der Hund   wurde (von dem Mann) geschlagen.
       the dog$_{nom}$ was   by the man   beaten
       'The dog was beaten (by the man).'

As the examples in (6) show, the complement of *geschlagen* can surface in two distinct cases depending on its syntactic environment. The case of *der Mann* and *den Hund/der Hund* is said to be structural because it depends on the syntactic structure of the phrase. The value of CAS is therefore a feature structure with the features CASE-TYPE and SYN-CASE. *snom* is a shorthand for structural nominative and *ldat* for lexical dative, respectively.

The entry for *geschlagen* then has the following LOCAL value.[8]

geschlagen:

$$
\begin{bmatrix}
\text{CAT} & \begin{bmatrix}
\text{HEAD} & \begin{bmatrix} \text{SUBJ} \left\langle \text{ NP}[str]_{\boxed{1}} \right\rangle \\ verb \end{bmatrix} \\
\text{ERG} & \left\langle \boxed{2} \right\rangle \\
\text{COMPS} & \left\langle \boxed{2}\ \text{NP}[str]_{\boxed{3}} \right\rangle
\end{bmatrix} \\
\text{CONT} & \begin{bmatrix}
\text{AGENT} & \boxed{1} \\
\text{PATIENT} & \boxed{3} \\
schlagen
\end{bmatrix} \\
loc
\end{bmatrix}
\tag{7}
$$

---

[7]Compare the discussion of the PVP account of Hinrichs and Nakazawa in (Müller, 1997a; Müller, 1997b) for instance.

[8]The ERG value contains a subject or complement of the verb that has 'accusative' properties. This feature is needed for a unified account of the German personal and impersonal passive (see (Pollard, 1994) for details).

Heinz and Matiasek use a case principle which employs the order information of the COMPS list.[9] The least oblique element receives nominative case if its CASE value is *structural*. The other elements receive accusative case if their CASE value is *structural*. If a SUBJ feature is used, i.e. if the subject of non-finite verbs is not part of the COMPS list, the case principle must include a special implication for non-finite verbs.[10]

Principle 1 (Case Principle—Case Assignment on COMPS)

$$
a. \quad \begin{bmatrix} \text{SYNSEM} \begin{bmatrix} \text{LOC|CAT|HEAD} \begin{bmatrix} \text{VFORM} \textit{fin} \\ \textit{verb} \end{bmatrix} \end{bmatrix} \\ \text{DTRS} \begin{bmatrix} \text{H-DTR|SYNSEM|LOC|CAT|COMPS} \langle \ \text{NP}[\textit{str}] \ \rangle \oplus \boxed{1} \\ \textit{head-comp-structure} \end{bmatrix} \\ \text{DTRS|H-DTR|SYNSEM|LOC|CAT|COMPS} \langle \ \text{NP}[\textit{snom}] \ \rangle \oplus \boxed{1} \end{bmatrix} \Rightarrow \\ \land
$$

---

[9]In fact, they are using a SUBCAT list as was common in earlier versions of HPSG. In their account, both subjects and complements are elements of the SUBCAT list.

Heinz and Matiasek are using a type hierarchy for the value of CASE. In their account *snom* is a subtype of *structural* and *nominative*. With such a hierarchy, (i) is not explainable as the generalization about (i) is that the second and the third argument of *nennen* bear the same case.

(i) a. Sie nannte ihn einen Lügner.

 b. Er wurde ein Lügner genannt.

However, the case type must be different because if the case type of *ihn* and *einen Lügner* would be structural the case of *ein Lügner* in the second sentence would have to be accusative. To express this with structure sharing two different features are needed.

[10]Note that the implication c. applies to adjectives as well. The case in (i) is therefore not assigned lexically.

(i) die den Mann liebende Frau

Contrary to the claim of Heinz and Matiasek (1994, p. 211), the case assignment for adjectives in the lexicon is impossible because there are sentences like (ii) where verbs occur in a prenominal verbal complex.

(ii) a. das neue, zischende, immer explodieren wollende Kraftwerk Fortuna Nord. (Günter Grass, Die Blechtrommel, Deutscher Taschenbuchverlag, 1993, p. 543)

 b. sagt sie in schneidend ironisch sein wollendem Ton (Wiglaf Droste, taz, 09.12.97, p. 20)

 c. Ich sehe heute: Sie, die Partei, braucht das sich selbst bestimmen wollende Individuum nicht. (Die Welt, 04.07.97, „Männer hinter Masken")

 d. ? die den Mann lieben wollende Frau

In a sentence where a verb that assigns structural case is governed by *wollend-* case cannot be assigned in the lexicon.

$$
b. \begin{bmatrix} \text{SYNSEM} \begin{bmatrix} \text{LOC}|\text{CAT}|\text{HEAD} \begin{bmatrix} \text{VFORM}\,fin \\ verb \end{bmatrix} \end{bmatrix} \\ \text{DTRS} \begin{bmatrix} \text{H-DTR}|\text{SYNSEM}|\text{LOC}|\text{CAT}|\text{COMPS} \Big\langle\, \big[{}_{synsem} \big] \,\Big\rangle \oplus \boxed{1} \oplus \\ \qquad\qquad\qquad\qquad\qquad \Big\langle\, \text{NP}[str]\, \Big\rangle \oplus \boxed{2} \\ head\text{-}comp\text{-}structure \end{bmatrix} \end{bmatrix} \Rightarrow
$$

$$
\begin{bmatrix} \text{DTRS}|\text{H-DTR}|\text{SYNSEM}|\text{LOC}|\text{CAT}|\text{COMPS} \Big\langle\, \big[{}_{synsem} \big] \,\Big\rangle \oplus \boxed{1} \oplus \\ \qquad\qquad\qquad\qquad\qquad\qquad\quad \Big\langle\, \text{NP}[sacc]\, \Big\rangle \oplus \boxed{2} \end{bmatrix} \wedge
$$

$$
c. \begin{bmatrix} \text{SYNSEM} \begin{bmatrix} \text{LOC}|\text{CAT}|\text{HEAD} \begin{bmatrix} \text{SUBJ} \Big\langle\, synsem \,\Big\rangle \\ verb \vee adj \end{bmatrix} \end{bmatrix} \\ \text{DTRS} \begin{bmatrix} \text{H-DTR}|\text{SYNSEM}|\text{LOC}|\text{CAT}|\text{COMPS}\,\boxed{1} \oplus \Big\langle\, \text{NP}[str]\, \Big\rangle \oplus \boxed{2} \\ head\text{-}comp\text{-}structure \end{bmatrix} \end{bmatrix} \Rightarrow
$$

$$
\begin{bmatrix} \text{DTRS}|\text{H-DTR}|\text{SYNSEM}|\text{LOC}|\text{CAT}|\text{COMPS}\,\boxed{1} \oplus \Big\langle\, \text{NP}[sacc]\, \Big\rangle \oplus \boxed{2} \end{bmatrix}
$$

4 The CELR and Structural Case

The problem with this formulation of the case principle is that if the CELR is applied to *geschlagen* the relevant element of the COMPS list will be moved into the SLASH set. Once it is contained in the SLASH set case assignment is not possible anymore. As was shown in (Müller, 1994), it is not possible to use the static ARG-S list proposed in Chapter 9 of (Pollard and Sag, 1994) for case assignment. The static ARG-S list would contain SUBJ+COMPS in order of obliqueness and would not vary according to the syntactic environment in which the lexical entry is used. But such a variation is exactly what is needed to properly assign case using an ARG-S list.

5 The Dynamic ARG-S list

The ARG-S list can be constructed dynamically by mirroring the process that takes place during the construction of COMPS lists in systems without CELRs.

The appropriate entries for *haben* and *werden* are then as follows:

haben (version with dynamically constructed ARG-S list):

$$
\begin{bmatrix}
\text{HEAD} & \begin{bmatrix} \text{SUBJ}\,\boxed{1} \\ verb \end{bmatrix} \\
\text{ARG-S} & \boxed{2} \\
\text{ERG} & \langle\,\rangle \\
\text{COMPS} & \boxed{3} \\
\text{VCOMP} & \text{V}[\text{LEX}+,\ ppp,\ \text{SUBJ}\ \boxed{1},\ \text{ARG-S}\ \boxed{2},\text{COMPS}\ \boxed{3},\ \text{VCOMP}\ none] \\
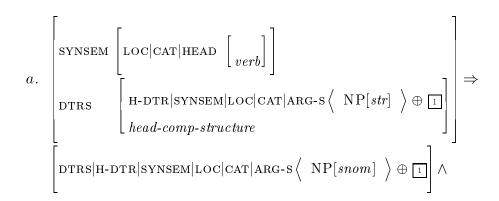cat &
\end{bmatrix}
\tag{8}
$$

werden (version with dynamically constructed ARG-S list):

$$
\begin{bmatrix}
\text{HEAD} & \begin{bmatrix} \text{SUBJ}\,\boxed{1} \\ verb \end{bmatrix} \\
\text{ARG-S} & \boxed{1} \oplus \boxed{2} \oplus \big\langle\ (\text{PP}[von]{:}\boxed{3}ref)\ \big\rangle \\
\text{ERG} & \boxed{1} \\
\text{COMPS} & \boxed{4} \oplus \big\langle\ (\text{PP}[von]{:}\boxed{3}ref)\ \big\rangle \\
\text{VCOMP} & \text{V}[\text{LEX}+,\ ppp,\ \text{SUBJ}\ \boxed{5}\ \big\langle\ \text{NP}[str]\ {:}\boxed{3}\ \big\rangle,\ \text{ERG}\ \boxed{1}, \\
& \quad\quad \text{ARG-S}\ \boxed{5} \oplus \boxed{1} \oplus \boxed{2},\ \text{COMPS}\ \boxed{1} \oplus \boxed{4},\ \text{VCOMP}\ none] \\
cat &
\end{bmatrix}
\tag{9}
$$

The following two implications are part of the case principle:[11]

Principle 2 (Case Principle—Case Assignment on ARG-S)

$$
a.\quad
\begin{bmatrix}
\text{SYNSEM} & \begin{bmatrix} \text{LOC}|\text{CAT}|\text{HEAD} & \begin{bmatrix} verb \end{bmatrix} \end{bmatrix} \\
\text{DTRS} & \begin{bmatrix} \text{H-DTR}|\text{SYNSEM}|\text{LOC}|\text{CAT}|\text{ARG-S}\big\langle\ \text{NP}[str]\ \big\rangle \oplus \boxed{1} \\ \text{head-comp-structure} \end{bmatrix}
\end{bmatrix}
\Rightarrow
$$

$$
\begin{bmatrix}
\text{DTRS}|\text{H-DTR}|\text{SYNSEM}|\text{LOC}|\text{CAT}|\text{ARG-S}\big\langle\ \text{NP}[snom]\ \big\rangle \oplus \boxed{1}
\end{bmatrix} \wedge
$$

---

[11]Genitive complements of nouns bear structural case as well. The implication for this case is omitted because it is irrelevant here.

$$
b. \quad
\begin{bmatrix}
\text{SYNSEM} & \begin{bmatrix} \text{LOC|CAT|HEAD} & \begin{bmatrix} verb \lor adj \end{bmatrix} \end{bmatrix} \\[2em]
\text{DTRS} & \begin{bmatrix} \text{H-DTR|SYNSEM|LOC|CAT|ARG-S} \left\langle \begin{bmatrix} synsem \end{bmatrix} \right\rangle \oplus \boxed{1} \oplus \\[1em] \qquad\qquad \left\langle \; \text{NP}[str] \; \right\rangle \oplus \boxed{2} \\[1em] head\text{-}comp\text{-}structure \end{bmatrix}
\end{bmatrix} \Rightarrow
$$

$$
\begin{bmatrix}
\text{DTRS|H-DTR|SYNSEM|LOC|CAT|ARG-S} \left\langle \begin{bmatrix} synsem \end{bmatrix} \right\rangle \oplus \boxed{1} \oplus \\[1em]
\qquad\qquad \left\langle \; \text{NP}[sacc] \; \right\rangle \oplus \boxed{2}
\end{bmatrix}
$$

Note that by assining case at ARG-S, it is sufficient to have two implications for case assignment in verbal environments as the subject of finite and non-finite verbs is contained in ARG-S, which is not the case for COMPS. Instead of having the implications b and c of principle 1, it is sufficient to state the implication b of principle 2.

Note also, that by assigning case at ARG-S, the order of the elements of the COMPS list becomes completely not relevant. This gives the grammar writer free choice in selecting means for describing the relatively free constituent order in German. So for instance, a COMPS list which need not be ordered gives the grammar writer the choice between using reordering lexical rules or a COMPS set.

6   The Extraction of Subjects and *von*-PPs

Not all the problems discussed in (Müller, 1994) are solved yet. The remaining question is how subjects and *von*-PPs can be extracted. Because both subjects and other complements can appear sentence initially, German V2 sentences are usually described by the topicalization of the subject or of one of the other complements. For finite verbs the SILR inserts the SUBJ value into the COMPS list from which it can be extracted by the CELR.

The variable COMPS list of auxiliaries poses implementational difficulties since it precludes the application of the *append* relation (unless one uses late evaluation techniques). The reason for this is that the arguments of *append* might both contain variables at the time the SILR has to be applied. For the same reason, it is not possible to apply the CELR to auxiliaries. The restriction that the SLASH set can contain at most one element would lead to spurious ambiguities because complements could be extracted from both the auxiliary and from the embedded verb.[12] Without this restriction one would get an infinite lexicon and spurious ambiguities. Without being able to apply the CELR, sentences like those in (3)–repeated here as (10)–, where a subject or a *von*-PP is fronted, cannot be accounted for.

---

[12]Whether one gets spurious ambiguities or wrong analyses depends on the view one takes on lexical rules. Höhle (1994) shows that one gets wrong analyses under certain assumptions regarding the semantics of lexical rules.

(10) a. Der Mann   hat den Hund geschlagen.
      the man$_{nom}$ has the dog$_{acc}$ beaten
      'The man has beaten the dog.'

   b. Von wem wurde dieses Buch   geschrieben?
      by whom was   this book$_{nom}$ written
      'Who did write this book?'

The *von*-PP is not part of the COMPS list of *geschrieben* (written), but an argument of *wurde* (was).

In the following, I will give a reformulation of the verb cluster schema that operates with closed lists, i.e., lists that don't contain any variables. Therefore both subjects and *von*-PPs can be extracted from the auxiliary entry by the CELR. The new entries for *haben* and *werden* are:[13]

haben (version with closed COMPS list):

$$
\begin{bmatrix}
\text{HEAD} & \begin{bmatrix} \text{SUBJ}\,\boxed{1} \\ verb \end{bmatrix} \\
\text{ARG-S} & \boxed{2} \\
\text{ERG} & \langle\rangle \\
\text{COMPS} & \langle\rangle \\
\text{VCOMP} & \text{V}[\text{LEX}+,\ ppp,\ \text{SUBJ}\ \boxed{1},\ \text{ARG-S}\ \boxed{2}, \text{COMPS}\ \boxed{3},\ \text{VCOMP}\ none] \\
cat
\end{bmatrix}
\tag{11}
$$

werden (version with closed COMPS list):

$$
\begin{bmatrix}
\text{HEAD} & \begin{bmatrix} \text{SUBJ}\,\boxed{1} \\ verb \end{bmatrix} \\
\text{ARG-S} & \boxed{1} \oplus \boxed{2} \oplus \left\langle\ (\text{PP}[von]{:}\boxed{3}ref)\ \right\rangle \\
\text{ERG} & \boxed{1} \\
\text{COMPS} & \left\langle\ (\text{PP}[von]{:}\boxed{3}ref)\ \right\rangle \\
\text{VCOMP} & \text{V}[\text{LEX}+,\ ppp,\ \boxed{5}\ \text{SUBJ}\ \left\langle\ \text{NP}[str]\ {:}\boxed{3}\ \right\rangle,\ \text{ERG}\ \boxed{1}, \\
& \quad \text{ARG-S}\ \boxed{5} \oplus \boxed{1} \oplus \boxed{2},\ \text{COMPS}\ \boxed{1} \oplus \boxed{4},\ \text{VCOMP}\ none] \\
cat
\end{bmatrix}
\tag{12}
$$

If a finite form is built, the subject will be inserted in front of the optional *von*-PP, resulting in a COMPS list containing the subject alone or the subject and the PP. The information that the arguments of the embedded verb are attracted is contained implicitly in the VCOMP values. Verbs that do not attract the complements of their verbal complement do have the VCOMP value *none*. In this case, a verbal complement is then listed as an ordinary oblique complement in the COMPS list. The revised version of the verb cluster schema accounts for the argument attraction:

---

[13]The tags $\boxed{3}$ in (11) and $\boxed{4}$ in (12) do not express structure sharing.

Schema 1 (Verb Cluster Schema)

$$
\begin{bmatrix}
\text{SYNSEM} & \begin{bmatrix} \text{LOC|CAT|COMPS} \boxed{1} \oplus \boxed{2} \end{bmatrix} \\[2em]
\text{DTRS} & \begin{bmatrix}
\text{HEAD-DTR} & \begin{bmatrix} \text{SYNSEM|LOC|CAT} & \begin{bmatrix} \text{ERG} & \boxed{3} \\ \text{COMPS} & \boxed{1} \\ \text{VCOMP} & \boxed{4} \end{bmatrix} \end{bmatrix} \\[2em]
\text{CLUSTER-DTR} & \begin{bmatrix} \text{SYNSEM} \boxed{4} & \begin{bmatrix} \text{LOC|CAT} & \begin{bmatrix} \text{COMPS} \boxed{3} \oplus \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\[1em]
\textit{head-cluster-structure}
\end{bmatrix}
\end{bmatrix}
$$

In this schema, the value of the ERG feature of the head daughter plays a crucial role in the construction of the COMPS list of the resulting sign. In the case of passive constructions the value is instantiated by the embedded verb. Here the empty list or the complement with 'accusative' properties is the value of ERG. In the case of perfect constructions, the value of ERG is the empty list. This means that the COMPS list of the embedded verb is taken over without any changes.
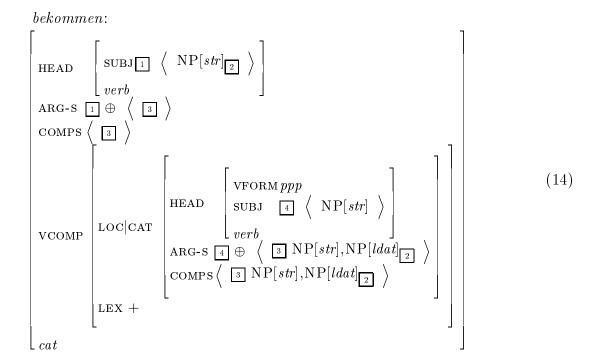
7   Problems

A problem for the account presented above is one very particular construction: the *bekommen* passive. In German, it is possible to use this construction with verbs like *bekommen*, *erhalten* and *kriegen*. These verbs also attract the COMPS list of their verbal complements. They can change the dative case of a complement of the embedded verb to the structural case.

(13) a. Der Mann   schenkt dem Jungen ein Buch.
     the man$_{nom}$ gives   the boy$_{dat}$ a book$_{acc}$
     'The man gives the boy a book.'

   b. Der Junge   bekommt (von dem Mann) ein Buch   geschenkt.
      the boy$_{nom}$ receives   by the man   a book$_{acc}$ given
      'The boy was given a book (by the man).'

The entry for *bekommen* is shown in (14).[14] Schema 1 does not account for such

---

[14]Note, that the ARG-S list and the COMPS list of the embedded verb are identical. This ensures that no lexical rule was applied to the *ppp*-verb in a *bekommen* construction. The CELR only applies to the matrix verb. Because of the identity of both lists, it is clear that a complement reordering lexical rule could not have been applied to the entry of the embedded verb. If one uses a COMPS set the INHER|SLASH value of the embedded verb has to be specified {}.

bekommen:

$$
\begin{bmatrix}
\text{HEAD} & \begin{bmatrix} \text{SUBJ}\boxed{1} \ \left\langle \ \text{NP}[str]_{\boxed{2}} \ \right\rangle \\ verb \end{bmatrix} \\
\text{ARG-S} \ \boxed{1} \oplus \left\langle \ \boxed{3} \ \right\rangle \\
\text{COMPS} \left\langle \ \boxed{3} \ \right\rangle \\
\text{VCOMP} \begin{bmatrix} \text{LOC}|\text{CAT} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{VFORM}\,ppp \\ \text{SUBJ} \quad \boxed{4} \ \left\langle \ \text{NP}[str] \ \right\rangle \\ verb \end{bmatrix} \\ \text{ARG-S} \ \boxed{4} \oplus \left\langle \ \boxed{3} \ \text{NP}[str], \text{NP}[ldat]_{\boxed{2}} \ \right\rangle \\ \text{COMPS} \left\langle \ \boxed{3} \ \text{NP}[str], \text{NP}[ldat]_{\boxed{2}} \ \right\rangle \end{bmatrix} \\ \text{LEX} \ + \end{bmatrix} \\
cat
\end{bmatrix}
\tag{14}
$$

an auxiliary. One has to use an exception feature. If the value of this feature—let us call it ATTR—is +, the COMPS list of the embedded verb is ignored and only the COMPS list of the matrix verb is taken into account.

Another problem is that the value of SUBJ might be a variable. For instance, verbs like *scheinen* (*seem*) may embed verbs with or without a subject.

(15) a. Karl scheint zu schlafen.
      Karl seems   to sleep
      'Karl seems to be asleep.'

   b. Hier scheint getanzt zu werden.
      here seems   danced   to be
      'There appears to be dancing going on here.'

But such verbs pose a problem for the analysis presented above only if one does not have the constraint that in German the CELR must not be applied recursively to its own output. This constraint is reasonable for German because only one constituent can take part in a nonlocal dependency construction at a time. Hinrichs and Nakazawa (1994b) developed an analysis for partial verb phrase fronting which is based on the assumption that the SLASH set may contain an arbitrary subset of the COMPS list of a verb. As was shown in (Müller, 1997a; Müller, 1997b) such an assumption is not necessary to describe PVP fronting. But if one allows for more than one element in the SLASH set, one is either forced to have two different entries for verbs like *scheinen*—one without subject and one with a subject—or to state the constraint that the SUBJ-list can have at most one element.

8   Conclusion

In this paper, I have demonstrated how a dynamically constructed ARG-S list can be used for case assignment. This ARG-S list can also be employed for Binding Theory and for the agreement principle.[15] The proposed approach makes it unnecessary to specify an order of increasing or decreasing obliqueness for the COMPS list. Furthermore, I demonstrated how a modified verb cluster schema can interact with a closed COMPS list, which makes the application of the CELR to auxiliaries possible.

References

Heinz, Wolfgang and Johannes Matiasek. 1994. Argument structure and case assignment in German. In John Nerbonne, Klaus Netter, and Carl J. Pollard, editors, *German in Head-Driven Phrase Structure Grammar*, CSLI Lecture Notes, Number 46. Center for the Study of Language and Information, Stanford, chapter 6, pages 199–236.

Hinrichs, Erhard W. and Tsuneko Nakazawa. 1994a. Linearizing AUXs in German verbal complexes. In John Nerbonne, Klaus Netter, and Carl J. Pollard, editors, *German in Head-Driven Phrase Structure Grammar*, CSLI Lecture Notes, Number 46. Center for the Study of Language and Information, Stanford, chapter 1, pages 11–38.

Hinrichs, Erhard W. and Tsuneko Nakazawa. 1994b. Partial-VP and split-NP topicalization in German—an HPSG analysis. In Erhard W. Hinrichs, Walt Detmar Meurers, and Tsuneko Nakazawa, editors, *Partial-VP and Split-NP Topicalization in German—An HPSG Analysis and its Implementation*, Arbeitspapiere des Sonderforschungsbereiches 340, Bericht Nr. 58. Universität Tübingen. ftp://ling. ohio-state.edu/pub/HPSG/Workshop.LLI.94/hinrichs-nakazawa.ps.Z. 09.07.97.

Höhle, Tilman. 1994. Spurenlose Extraktion. ms, Seminar für Sprachwissenschaften, Universität Tübingen.

Müller, Stefan. 1994. Problems with complement extraction lexical rules. http:// www.dfki.de/~stefan/Pub/e_celr.html. 09.23.1997.

Müller, Stefan. 1997a. *Head-Driven Phrase Structure Grammar für das Deutsche*. Lecture Notes Computational Linguistics. Humboldt Universität zu Berlin. http://www.dfki.de/~stefan/Pub/e_hpsg.html. 09.23.1997.

Müller, Stefan. 1997b. Yet another paper about partial verb phrase fronting in German. Research Report RR-97-07, Deutsches Forschungszentrum für Künstliche Intelligenz, Saarbrücken. A shorter version appeared in *Proceedings of COLING 96*. http://www.dfki.de/~stefan/Pub/e_pvp.html. 09.23.1997.

---

[15]See (Müller, 1997a).

Pollard, Carl J. 1994. Toward a unified account of passive in German. In John Nerbonne, Klaus Netter, and Carl J. Pollard, editors, *German in Head-Driven Phrase Structure Grammar*, CSLI Lecture Notes, Number 46. Center for the Study of Language and Information, Stanford, chapter 8, pages 273–296.

Pollard, Carl J. and Ivan A. Sag. 1987. *Information-Based Syntax and Semantics Volume 1 Fundamentals*. CSLI Lecture Notes, Number 13. Stanford: Center for the Study of Language and Information.

Pollard, Carl J. and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. University of Chicago Press.

Rentier, Gerrit. 1994. Dutch cross serial dependencies in HPSG. In *Proceedings of COLING 94*, pages 818–822, Kyoto, Japan. http://xxx.lanl.gov/abs/cmp-lg/9410016. 07.24.97.

Uszkoreit, Hans. 1987. *Word Order and Constituent Structure in German*. CSLI Lecture Notes, Number 8. Stanford: Center for the Study of Language and Information.

Complement Extraction Lexical Rules and Argument Attraction

Stefan Müller