Proceedings of the HPSG09 Conference

Georg-August-Universität Göttingen, Germany

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/

Contents

1 Editor's Note

The 16th International Conference on Head-Driven Phrase Structure Grammar (2009) was held in Göttingen.

The conference featured 2 invited talks and 19 papers selected by the program committee (Anne Abeille, Doug Arnold [chair], Olivier Bonami, Bob Borsley, Gosse Bouma, Rui Chaves, Berthold Crysmann, Markus Egg, Elisabet Engdahl, Dan Flickinger, Jonathan Ginzburg, Chikara Hashimoto, Jong-Bok Kim, Tibor Kiss, Anna Kupsc, Shalom Lappin, Bob Levine, Rob Malouf, Detmar Meurers, Stefan Müller, Tsuneko Nakazawa, Gerald Penn, Adam Przepiorkowski, Frank Richter, Louisa Sadler, Ivan Sag, Jesse Tseng, Stephen Wechsler).

In total there were 34 submissions to the conference. We want to thank the program committee for putting this nice program together.

Thanks go to Anke Holler, Manfred Sailer, Heike Walker, Gert Webelhuth [chair], who were in charge of local arrangements.

As in the past years the contributions to the conference proceedings are based on the five page abstract that was reviewed by the respective program committees, but there is no additional reviewing of the longer contribution to the proceedings. To ensure easy access and fast publication we have chosen an electronic format.

The proceedings include all the papers except those by Danièle Godard and Anne Abeillé and Polly Jacobson.

A Construction-based Analysis of Verbless Relative Adjuncts in French and Romanian

Gabriela Bîlbîie and Frédéric Laurens

LLF, Université Paris Diderot Paris 7

Proceedings of the HPSG09 Conference

Georg-August Universität Göttingen, Germany

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

French and Romanian verbless relative adjuncts are incidental adjuncts which have been described as elliptical relative clauses. We show that this analysis is not empirically adequate and propose an alternative non-elliptical analysis. We analyze verbless relative adjuncts as sentential fragments whose head can be a cluster of phrases. They are marked by a functor phrase which displays selection properties with respect to the head phrase and makes an essential contribution to the semantics of the adjunct. The analysis relies on the interaction of grammatical constraints introduced by various linguistic objects, as well as on a constructional analysis of verbless relative adjuncts distinguishing several subtypes.

1 Introduction

Both French and Romanian have verbless adjuncts whose form is reminiscent of relative clauses. These verbless adjuncts are exemplified in (1) for French and in (2) for Romanian. They are characterized by the presence of a fronted constituent (shown in italics) which can either be a prepositional phrase containing a WH form (1a, 2a, 2b) or the form *dont* in French (1b). We label those constructions Verbless Relative Adjuncts (henceforth VRAs).

(1) a. Trois personnes, [*parmi lesquelles* Jean], sont venues.
three people(FEM), [among which.FEM John], AUX come
'Three people, among which John, have come.'

b. Trois personnes sont venues, [*dont* une hier].
three people(FEM) AUX come, [DONT one.FEM yesterday]
'Three people have come, one of them yesterday.'

(2) a. Au venit trei persoane, [{ *printre* | *între* } *care* (şi) Ion].
AUX come three people, [{among | among} which.ACC (also) John]
'Three people have come, among which (also) John.'

b. Au venit trei persoane, [*dintre care* una ieri].
AUX come three people(FEM), [among which.ACC one.FEM yesterday]
'Three people have come, one of them yesterday.'

French and Romanian VRAs have been described as elliptical relative clauses (Grevisse 1993, Gheorghe 2004 and Gheorghe 2005) on the basis of semantic and syntactic similarities with non-restrictive relative clauses (3).

[†]Part of this work has been presented in Bucharest at the 11^{th} Conference of the English Department. Many thanks to A. Abeillé, D. Arnold, O. Bonami, D. Godard, J-M. Marandin, the audience of the HPSG09 Conference and three anonymous reviewers for helpful discussions or useful suggestions.

(3) a. Il a écrit de nombreux romans, [*dont* deux ont été publiés le mois dernier].
'He has written many novels, two of which have been published last month.'

b. El a scris multe romane, [*dintre care* două au fost publicate luna trecută].
'He has written many novels, two of which have been published last month.'

Part of the data used in this work is based on a corpus study. The French data comes from the French Treebank of Paris 7 (Abeillé et al. 2003). Lacking a similar corpus for Romanian, we collected examples from press texts.

The paper is structured as follows. We first focus on the constituency of VRAs. We show that the elliptical analysis of VRAs is not empirically adequate and propose an alternative non-elliptical analysis. Then, we discuss the distributional and functional properties of VRAs before turning to their semantic properties. The analysis is then couched in a constructional version of HPSG (i.e. SBCG, Sag 2007). The analysis relies on the interaction of grammatical constraints introduced by various linguistic objects, as well as on a constructional analysis of VRAs distinguishing several subtypes.

For reasons of space, properties of VRAs which are the same in both languages are only illustrated for French. Romanian examples are only used when the property is absent or less clear in French.

2 Constituency

French and Romanian VRAs are constituted of two parts. The first part (labeled the *fronted constituent*) is reminiscent of the extracted phrase or of the complementizer of a relative clause. The second part (labeled the *body*) is composed of either a single phrase or of a sequence of phrases which are not related by grammatical functions. We show that an analysis of VRAs as elliptical relative clauses is not empirically supported and propose an alternative analysis.

2.1 The fronted constituent

The fronted constituent of VRAs contains lexical items which are found in relative clauses. Those items include forms of the WH lexeme *lequel* in French and *care* in Romanian, and the form *dont* in French. While French *lequel* and Romanian *care* are found in interrogative clauses too, French *dont* is only found in relative clauses apart from its use in VRAs.

The WH forms always appear in prepositional phrases in VRAs. The prepositions heading the fronted constituent all have an abstract partitive meaning similar to that of the preposition *among* in English. This includes *parmi* in French and *dintre*, *între*, and *printre* in Romanian.

In both languages, more complex expressions are also found, such as *au* {*premier* | *second*} *rang desquels* 'in the {first | second} row of which', *au sommet desquels* 'at the top of which', *au sein desquels* 'in the middle of which', *au centre desquels* 'in the center of which' or *au nombre desquels* 'in the count of which' in French, and *în rândul cărora* 'in the row of which', *în mijlocul cărora* 'in the middle of which' or *în fruntea cărora* 'at the top of which' in Romanian. Although these expressions almost always compositionally denote spatial relations, they are only found with an abstract partitive meaning in VRAs. Note that some of these expressions additionally suggest a ranking among the subparts.

The WH form appearing in the fronted phrase is coreferential with a noun phrase preceding the VRA in the host clause, which we label the *licenser* (see section 3.1). This is signalled by morphological agreement on the WH form in French: *lesquels* (MASC) vs. *lesquelles* (FEM). Agreement cannot be observed in Romanian because both the accusative form *care* and the genitive plural form *cărora* do not vary in gender.

In French, the fronted phrase can also contain the form *dont*. Apart from its use in VRAs, the form *dont* appears only in relative clauses where it has been analyzed as a complementizer (Godard 1988).

Like the complementizers *que* and *qui* which are also found in relative clauses, the complementizer *dont* only combines with finite clauses (4a). WH relative clauses on the other hand are not always finite. Unlike prepositional WH forms like *duquel* 'of which', the complementizer *dont* cannot be embedded within a filler phrase (4b), and does not show morphological agreement with an antecedent. Finally, complementizer *dont* marks relative clauses containing a missing constituent which would be marked by the form *de* (4c) or any proform coreferential with the antecedent of the relative clause, as long as the proform is embedded under a propositional attitude predicate, such as *penser* 'to think' or *être vrai* 'to be true' (4d).

(4) a. * Un livre [*dont* parler]
  a book [DONT talk.INF]
  'A book which we should talk about'

b. Un homme, [*le frère* { * *dont* | *duquel* } est venu hier]
  a man, [the brother { DONT | of.which } AUX come yesterday]
  'A man, the brother of which has come yesterday'

c. Un livre [*dont* on parle beaucoup]
  'A book which one talks about'

d. Un livre [*dont* il est vrai qu'il coûte cher]
  'A book of which it is true that it is expensive'

VRAs' *dont* shares some of its properties with the complementizer *dont*. For instance, it cannot be embedded within the fronted constituent as well. However, it is unclear whether other selection properties of the complementizer are shared by VRAs' *dont*. Since an elliptical analysis of VRAs is not empirically adequate (see section 2.3), we argue that it is not the case. Moreover, while complementizer *dont* is assumed to have no semantic contribution, such a statement is hard to make for VRAs' *dont* since it forces a partitive semantics (5), although the partitive semantics may also be assumed to be constructionally introduced in VRAs.

(5) a. Au total, dix livres ont été commandés, [(**dont*) tous pour toi].
'In total, ten books have been ordered, all of them for you.'

b. Au total, dix livres ont été commandés, [(*dont*) deux pour toi].
'In total, ten books have been ordered, two of them for you.'

The left edge of the fronted phrase must coincide with the left edge of the VRA. It cannot be linearized in or after the body (6a) or be preceded by adverbials (6b).

(6) a. *Plusieurs personnes, [Jean *parmi lesquelles*], sont venues.
'Several people, among which John, have come.'

b. *Plusieurs personnes, [notamment *parmi lesquelles* Jean], sont venues.
'Several people, among which notably John, have come.'

2.2 The body

The body of VRAs is constituted of either a single phrase or a sequence of phrases. When the body is a single phrase, it can be either a noun phrase in the citation form (i.e. no prepositional marking in French and Romanian and nominative case in Romanian) or a phrase of any category whose form parallels the corresponding phrase in the host clause.

Not every VRA type allows for the two options. In French, WH VRAs do not allow for marked single phrases and *dont* VRAs disprefer it. *Dont* VRAs with a single marked constituent are not as bad as their WH VRA counterparts and can be improved with adverbs, such as *notamment* 'notably' (7).

(7) Un jeune homme annonce à diverses personnes sa mort prochaine, [{**parmi lesquelles* | % *dont*} notamment à un psychiatre qui décide de l'aider].
'A young man announces his imminent death to several people, {among which | DONT} *notably to a psychiatrist who decides to help him.'*

In Romanian, *dintre* is incompatible with preposition or case marking (8a)[1], while no such restriction is found with *printre* and *între* as long as the adverb *şi* 'also' precedes them (8b).

(8) a. Ion lucrează cu şapte doctori, [*dintre care* (*cu) doi ruşi].
John works with seven doctors, [among which with two Russian]
'John works with seven doctors, two of them are Russian.'

[1] Adverbs such as *mai ales* 'especially' can improve the acceptability of (8a).

b. Ion a oferit flori mai multor fete, [*printre care* şi
John AUX offered flowers ADV many.DAT girls, [among which also
{Maria | Mariei}].
{Mary.NOM | Mary.DAT}]
'John has offered flowers to many girls, Mary among them.'

The body of a VRA can also be constituted of a sequence of phrases (i.e. a cluster). Three types need to be distinguished. Clusters of type I mimick the syntax of the host clause. In those clusters, each of the phrases has to be marked like the correponding phrase in the host clause (9a). Clusters of type II contain a noun phrase in the citation form and a predicative phrase expressing a property of that noun phrase (9b). Clusters of type III are only found in VRAs whose licenser is a past participle used to express a functional property assumed by some individuals within a event. For example, *un blessé* 'an injured person' is the patient of an event in which someone gets hurt. In clusters of that type, the event relation of the cluster is contributed by the past participle. As a result, adverbs, rather than adjectives, are used to modify that relation (9c).

(9) a. Certains ont parlé à mes amis, [*dont* Marie *(à) Marc].
'Some have spoken with my friends, Mary with Mark.'

b. Je vends dix jeux, [*dont* la plupart encore dans leur boîte].
'I sell ten games, the majority of them still in their original box.'

c. L'accident a fait douze blessés, [*dont* cinq grièvement].
'The accident left twelve injured, five of them critically.'

Fronted phrases show selection properties regarding the type of the cluster they combine with. For instance, in French VRAs with *parmi*, at least one phrase in the cluster must be a noun phrase. These properties of fronted phrases combined with properties of clusters of type I can result in ungrammaticality (10).

(10) J'ai parlé à plusieurs personnes hier, [{*dont* | **parmi lesquelles*} à Marie de linguistique].
'I spoke with several people yesterday, of which with Mary about linguistics.'

2.3 VRAs are not elliptical relative clauses

VRAs are usually referred to as elliptical relative clauses. However, an elliptical analysis of VRAs faces two kinds of problems. Under an elliptical approach, VRAs are assumed to be relative clauses which have the additional property of having some of their syntactic or phonological material removed. An elliptical analysis is of interest if, and only if, (I) one can reconstruct a relative clause from any VRA in a regular fashion and (II) the semantic properties of VRAs are the same as that of relative clauses. We argue that none of these conditions are verified.

2.3.1 Arguments against syntactic reconstruction

There are at least three options for the reconstruction of a verbal form in VRAs. For an example like (11), the first option is to reconstruct a verbal form of the same lexeme as the verb of the host clause of the VRA (11b). This is the obvious option when the VRA contains a cluster of type I. The second option is the reconstruction of an existential verb (11c). The third option is the reconstruction of a quotation verb (11d). In some cases, none of these options will work.

(11) a. Plusieurs ont eu un cadeau, [*dont* Marie un livre].
'Several have had a present, of which Mary a book.'

b. *Plusieurs ont eu un cadeau, [*dont* Marie a eu un livre].
'Several have had a present, of which Mary has had a book.'

c. *Plusieurs ont eu un cadeau, [*dont* est Marie un livre].
'Several have had a present, of which is Mary a book.'

d. *Plusieurs ont eu un cadeau, [*dont* on cite Marie un livre].
'Several have had a present, of which one mentions Mary a book.'

Within a syntactic reconstruction approach, the choice of a verbal form is dependent on lexical constraints, such as subcategorization properties, which are not correlated with semantic properties. For instance, it is possible to reconstruct the verb *figurer* 'to appear' within a *parmi* VRA but not in a *dont* VRA, because *figurer* can subcategorize a PP marked by *parmi* but not a PP marked by *de*. The reverse is true for an expression such as *faire partie de* 'to belong to'.

2.3.2 VRAs do not have the same semantic properties as relative clauses

Non-restrictive relative clauses behave semantically like independent clauses that contain a proform (Arnold 2004). As a result, their semantic contribution is largely independent from that of their host clause. This is not the case for VRAs, as shown by the contrast in (12). While the sequence of utterances in (12a), which contains a relative clause, is coherent, the sequence in (12b) is contradictory because whales are said both to have and not to have apparent ears. This is so because VRAs are sentential fragments (see section 4.3). Thus, only a syntactic reconstruction which reproduces the content of the host clause is compatible with the semantics of VRAs. However, this kind of reconstruction is most often impossible on syntactic grounds.

(12) a. Non, tu te trompes! Bien que beaucoup de mammifères, [*dont* les baleines sont un exemple], aient des oreilles apparentes, les baleines n'en ont pas.
'No, you're wrong! While many mammals, of which whales are a example, do have apparent ears, whales do not have any.'

b. Non, tu te trompes! Bien que beaucoup de mammifères, [*dont* les baleines], aient des oreilles apparentes, les baleines n'en ont pas.
'No, you're wrong! While many mammals, whales among others, do have apparent ears, whales do not have any.'

Another problem faced by the elliptical account is that it predicts that some VRAs should be well-formed, while they are ill-formed for semantic reasons. This is so because the elliptical account assumes that the partitive semantics of VRAs comes from the elided verbal predicate rather than from the fronted phrase. In Romanian, the preposition *dintre* cannot cooccur with a body containing a referential noun phrase such as a definite one. However, a verb form can be reconstructed without difficulty yielding a well-formed relative clause (13).

(13) Au venit mai multe persoane, [*dintre care* {*Maria | o amintim pe Maria}].
'Many people have come, among which {Mary | one mentions Mary}.'

2.4 Non-elliptical alternative

Non-elliptical analyses differ from elliptical ones in that they do not link form constraints on clusters (such as those exhibited by clusters of type I) to the presence of a syntactic head in the structure. As a result, they make no prediction on the distribution of clusters. We assume that the body of VRAs has exactly the structure it seems to have at first sight: it has a flat structure and has no syntactic head. As for VRAs as a whole, we argue in favor of an analysis in which the body is the head and the fronted phrase is a functor phrase.

The selection properties of VRAs are best attributed to the body. This is so because most of the phrases functioning as the body in VRAs can also function as incidental adjuncts alone with a similar semantics. This is especially the case of those which contain adverbials such as *notamment* 'notably' (14a) or are coordinated structures (14b).

(14) a. De nombreuses espèces, [(*dont*) notamment les oursins], ont souffert de la pollution.
'Many species, (among which) notably urchins, have suffered from pollution.'

b. Plusieurs personnes, [une hier et deux ce matin], se sont plaintes de l'organisation.
'Several people, one yesterday and two this morning, complained about the organization.'

The selection properties of VRAs are distinct from those of the fronted phrase. Apart from the French form *dont* whose category is unclear, the fronted constituent is always a prepositional phrase. A preposition like *parmi* 'among' in French introduces a semantic relation between two arguments, one of which is typically realized as a complement (e.g. *lesquelles* in (15)). The other argument (the external argument) is usually not realized within the preposition phrase itself. Rather it is selected for by the prepositional phrase. In VRAs, the external argument of the fronted phrase is realized within the body (e.g. *Jean* in (15)). Thus, it is reasonable to assume that the selection properties of the fronted phrase and those of the VRA are distinct and therefore not to analyze the fronted phrase as the head.

(15) plusieurs personnes, [[*parmi lesquelles*] Jean]
several people.FEM.PL [among which.FEM.PL John(MASC.SG)]
'several people, among which John'

Following Van Eynde 2003, we use the function functor, which replaces the functions specifier, marker and (prehead) adjunct. Functors select a head and are able to mark it, hence modifying its distribution. The fronted phrase is best analyzed as a functor because (I) it must appear before the body (unlike other adjuncts, such as *notamment* 'notably'), (II) it displays selection properties, and (III) (at least in certain cases) it modifies the distribution of the phrase it combines with.

3 Function in the host phrase

VRAs are incidental adjuncts. They are linearized within a host phrase which must contain a noun phrase introducing a sum-denoting entity. That noun phrase is labeled the *licenser*. The syntactic relation between the VRA and its licenser is submitted to both linear order and locality constraints.

3.1 Properties of the licenser

The licenser must denote a plural entity whose subparts are accessible (i.e. an entity which can be described as a sum of entities) (Lasersohn 1995). This does not entail that the licenser of a VRA will always have plural morphology (16). In most corpus examples however, the licenser is a plural indefinite.

(16) **Un important volume de gaz** s'est échappé du cratère, [*dont* environ 25% de dioxyde de soufre].
'A great volume of gaz has been released from the crater, including about 25% of sulfur dioxide.'

The licenser can be a dependent of a head of any category (17a) and assume any function including adjunct (17b).

(17) a. Des représentants de **plusieurs pays**, [*dont* le Brésil], y assistent.
'Representatives of several countries, among which Brasil, attend it.'

b. Je l'ai attendu **plusieurs heures,** [*dont* une sous la pluie].
'I have been waiting for him for several hours, one of which in the rain.'

In some cases, it might be tempting to describe a VRA as having two licensers or more (18). This analysis is only possible when the fronted phrase does not make it explicit which of the noun phrase is truly the licenser of the VRA. We will come back to the analysis of these examples later in section (4.3).

(18) J'ai offert **des cadeaux à plusieurs personnes**, [*dont* un livre à Marie].
'I have offered presents to many people, among which a book to Mary.'

3.2 Ordering and locality constraints

The constraints on the linearization of VRAs need to make reference to the relative linearization of their licenser, as well as to the syntactic structure of their host.

VRAs must follow their licenser (19a). If the licenser is a dependent of a clause's head, the VRA can be linearized anywhere in the clause after the licenser (19b). However, if the licenser is not a dependent of a clause's head, the VRA needs to follow its licenser directly (19c).

(19) a. ([**Dont* Marie]), plusieurs personnes sont venues, ([*dont* Marie]).
'Several people have come, among which Mary.'

b. J'ai demandé à plusieurs personnes hier, [*dont* Marie], de m'aider un peu.
'I asked several people yesterday, among which Mary, to help me a bit.'

c. Des représentants de plusieurs pays, ([*dont* le Brésil]) se sont réunis, ([**dont* le Brésil]).
'Representatives from several countries, among which Brasil, have met.'

The relation between a VRA and its licenser obeys the right roof constraint which generally applies on rightwards non-local relations such as extraposition or right dislocation.

(20) *Que deux personnes viennent ne m'a pas étonné, [*dont* Marie].
'The fact that two people come, among which Mary, has not surprised me.'

Note that if a VRA contains a cluster of type I, the VRA must follow every phrase of the host clause which is paralleled in the cluster.

(21) Plusieurs, ([**dont* Marie un livre]), m'ont offert des cadeaux, ([*dont* Marie un livre]).
'Several have offered me presents, of which Mary a book.'

VRAs are also at least compatible with 'comma intonation'. Thus, they are a kind of incidental adjuncts, incidentality being defined as a syntactic property which correlates with both phonological and linearization properties (Bonami and Godard 2003).

4 Semantic Properties

VRAs can be associated with two distinct semantics. Although VRAs always introduce a subpart of the sum individual denoted by their licenser, the referential properties of the introduced subpart are not always the same. VRAs are non-restrictive modifiers which, unlike most other non-restrictive modifiers, must scope under propositional attitude verbs. This is explained by the fact that VRAs are sentential fragments.

4.1 VRAs have a partitive semantics

VRAs always introduce an entity which must be interpreted as a subpart of the sum individual denoted by their licenser. Sum individuals are not always composed of atomic parts. As a result, a noun phrase containing a mass noun can function as the licenser of a VRA (16). Quantified noun phrases which do not denote sum individuals are not suitable licensers for VRAs (22).

(22) *Tout étudiant doit venir, [*dont* Marie].
'Every student must come, among which Mary.'

VRAs cannot introduce any other semantic relation between their licenser and the phrase introducing the subpart. Meronymy and possession, for instance, are ruled out (23).

(23) *Plusieurs personnes sont venues, [*dont* leur chien].
'Several people have come, among them their dog.'

VRAs can have two distinct semantics depending on the referential properties of the introduced subpart. Exemplifying VRAs introduce a subpart which is referential (i.e. It can be identified independently of the fact that it is a subpart of a sum individual) (24a). This is the case of the noun phrases *Marie* and *Jean* in (24a). Partitioning VRAs, on the other hand, introduce a subpart which is not referential but can be defined within the sum individual as having some properties which are not shared by other subparts. The property in question can be a property of the entity which is the subpart or a property of the subevent in which the subpart takes part (24b). One of the most striking differences between exemplifying and partitioning VRAs is that only the latter can introduce a list of subparts which are coextensive with the sum individual denoted by the licenser (24b).

(24) a. {*Deux | trois} personnes sont venues, [*dont* Marie et Jean].
'{*Two* | *three*} *people have come, among which Mary and John.'*

b. Trois personnes sont venues, [*dont* une lundi et deux mardi].
'Three people have come, one on Monday and two on Tuesday.'

Within a noun phrase coordination, it is sufficient that one of the noun phrases be non referential in order for the partitioning semantics to be available (25).

(25) Prends deux objets, [*dont* cette bouteille et {un | *ce} couteau].
'Take two objects, that bottle and {a | that} knife.'

The semantics of the head of the fronted phrase also plays a role in the semantics of the VRA. In French, *parmi* is always exemplifying, while *dont* can be exemplifying or partitioning. In Romanian, *dintre* is always partitioning, while *printre* and *între* can be both exemplifying or partitioning. These lexical properties can be observed in other uses of the prepositions as well (26).

(26) a. Avem {majoritatea | spionii} *printre* noi.
'We have {the majority | spies} among us.' (partitioning or exemplifying)

b. {majoritatea | *spionii} *dintre* copii
'the {majority | spies} of the children' (partitioning only)

4.2 VRAs have a non-restrictive semantics

Restrictive modifiers have an intersective interpretation and therefore introduce an implicit 'contrast set', which can be accessed by anaphors like *the others* (Arnold 2004). Non-restrictive modifiers, on the other hand, are not intersective modifiers and introduce no such 'contrast set'. VRAs are non-restrictive modifiers of their licenser as well as of their host. The VRA in (27) does not restrict the set of Mary's friends to the one including John. Therefore, there is no possible antecedent for *the other friends of Mary*. Neither does the VRA in (27) restrict the event denoted by the host clause to the event such that *Some friends of Mary came and John came* as opposed to another event such that *Some friends of Mary came but John didn't*.

(27) Certains amis de Marie, [*parmi lesquels* Jean], sont venus. #Les autres amis de Marie viendront demain.
'Some friends of Mary's have come, among which John. The others friends of Mary will come tomorrow.'

Unlike most non-restrictive modifiers, the content of VRAs is part of the asserted content of the utterance containing them. As a result, VRAs which are linearized within a clause which is itself in the scope of a propositional attitude verb must be interpreted in the scope of that verb too. Thus, (28) entails that *Peter believes that vervain can heal ulcers* but not that it is actually true or even that the speaker believes it.

(28) Pierre croit que certaines plantes, [*dont* la verveine], soignent les ulcères.
'Peter believes that some plants, vervain among them, can heal ulcers.'

4.3 VRAs are sentential fragments

VRAs are sentential fragments and describe a subevent of the event denoted by their host clause. This is especially clear in partitioning examples where the subpart is not defined by a property of the subpart itself but by a property of the subevent (24b). However, this is also true of exemplifying examples as can be seen by the ungrammaticality of (29) in which a VRA is embedded within a non-event for which no subevent can be defined.

(29) *Aucun représentant de ces quatre pays, [*dont* le Brésil], n'a participé à la conférence.
'No representative of those four countries, among which Brasil, has attended the conference'

VRAs can appear in declarative, imperative (25) or interrogative clauses. In interrogative clauses, they share their abstracted parameter with the host clause and thus can be used as a hint. In (30), if one can answer the question about *France*, then one can answer the question about *countries of the OECD* too. The hint is that knowledge about France is sufficient to answer the question.

(30) En quelle année, plusieurs pays de l'OCDE, [*dont* la France], ont-ils signé ce traité?
'In which year, several countries of the OECD, among which France, did sign this treaty?'

The fact that VRAs are sentential fragments allows one to account for cases where VRAs seem to have several licensers as cases where VRAs have only one licenser but also introduce a sum-subpart relation because the event they introduce is a subevent of the one introduced by the host. In that kind of implied sub-subpart relation, the corresponding element in the host does not need to denote a sum-individual at all (31).

(31) Paul a offert un cadeau à plusieurs personnes, [*dont* un livre *(à Marie)].
'Paul has offered a present to several people, among which a book (to Mary).'

5 HPSG Grammar Fragment

The properties of VRAs are best analyzed as resulting from the interaction of various syntactic and semantic constraints applying on distinct linguistic objects. An HPSG grammar fragment accounting for the properties of VRAs minimally requires (I) a theory of fragments, (II) a theory of clusters, and (III) a theory of locality of selection. Several aspects of the grammar introduced here are constructional in nature. The analysis is couched in a constructional version of HPSG, namely SBCG (Sag 2007).

5.1 A theory of fragments

Fragments are expressions which convey a semantic content which is not given by their form alone. Rather, the semantic content conveyed by a fragment is a function of (I) the type of the fragment, (II) the semantic content of the constituent(s) in the fragment, and (III) contextual information which can be linguistic in nature or not (Fernández et al. 2007).

A sentential fragment such as the short question *when* in (32a) is interpreted as having the same semantic content as the clause *when she will come* in (32b). This semantic content comes in part from the type of the fragment (short questions have the same content type as interrogative clauses, i.e. a propositional abstract), the semantic content of the constituent(s) in the fragment (*when* provides the paramater for the propositional abstract), and contextual information (*Mary will come* functions as the antecedent of the fragment providing the proposition used to build the propositional abstract).

(32) a. Mary will come but nobody knows [when].

b. Mary will come but nobody knows [when she will come].

More generally, fragments are reminiscent of description anaphora, as opposed to instance anaphora. In instance anaphora, what is shared between an anaphoric expression and its antecedent(s) are indices. In description anaphora, what is shared is some aspects of the description of the antecedent(s) which apply to a new entity with a different index. This is the case with *one* anaphora in English (33) (Arnold and Borsley 2008).

(33) Here is a small red mugs with flowers and here is a bigger one.

In all of these cases, computing the semantics of the fragment can be achieved by expressing constraints between four semantic representations: two complete ones (the meaning of the antecedent and the meaning of the fragment) and two partial ones (the content which is anaphorically shared between the antecedent and the fragment, and the content which is given by the constituents in the fragment).

We use MRS (Copestake et al. (2005)) to represent incomplete semantic representations as the underspecification of a complete semantic representation. In MRS, meaning is represented as bags of elementary predications. Connection between these elementary predications is achieved through index-sharing and label-sharing. MRS representations are suitable to express partial meaning.

For an example such as (34), we want to obtain the four bags of elementary predications in (35). The bags are related by two meta-constraints (bag unification and bag intersection) shown at the bottom of figure (35).

(34) Plusieurs personnes sont venues, [*dont* Marie hier].
'Several people have come, of which Mary yesterday.'

(35)
$$\begin{bmatrix} \text{ANTECEDENT} & \boxed{A} \begin{bmatrix} \text{HOOK} & \begin{bmatrix} \text{IND} & \boxed{1} \end{bmatrix} \\ \text{RELS} & \boxed{A'} \left\langle \begin{bmatrix} \textit{message-rel} \\ \text{IND } \boxed{1} \end{bmatrix}, \begin{bmatrix} \textit{come-rel} \\ \text{IND } \boxed{1} \\ \text{ARG1 } \boxed{2} \end{bmatrix}, \begin{bmatrix} \textit{several-rel} \\ \text{IND } \boxed{2} \end{bmatrix}, \begin{bmatrix} \textit{person-rel} \\ \text{IND } \boxed{2} \end{bmatrix} \right\rangle \end{bmatrix} \\ \text{VARIABLE} & \boxed{B} \begin{bmatrix} \text{HOOK} & \begin{bmatrix} \text{IND} & \boxed{5} \end{bmatrix} \\ \text{RELS} & \boxed{B'} \left\langle \begin{bmatrix} \textit{message-rel} \\ \text{IND } \boxed{5} \end{bmatrix}, \begin{bmatrix} \textit{come-rel} \\ \text{IND } \boxed{5} \\ \text{ARG1 } \textit{ind} \end{bmatrix} \right\rangle \end{bmatrix} \\ \text{FRAGMENT} & \boxed{C} \begin{bmatrix} \text{HOOK} & \begin{bmatrix} \text{IND} & \boxed{3} \end{bmatrix} \\ \text{RELS} & \boxed{C'} \left\langle \begin{bmatrix} \textit{message-rel} \\ \text{IND } \boxed{3} \end{bmatrix}, \begin{bmatrix} \textit{name-rel} \\ \text{IND } \boxed{4} \end{bmatrix}, \begin{bmatrix} \textit{yesterday-rel} \\ \text{IND } \boxed{3} \end{bmatrix} \right\rangle \end{bmatrix} \\ \text{CONTENT} & \boxed{D} \begin{bmatrix} \text{HOOK} & \begin{bmatrix} \text{IND} & \boxed{3} \end{bmatrix} \\ \text{RELS} & \boxed{D'} \left\langle \begin{bmatrix} \textit{message-rel} \\ \text{IND } \boxed{3} \end{bmatrix}, \begin{bmatrix} \textit{come-rel} \\ \text{IND } \boxed{3} \\ \text{ARG1 } \boxed{4} \end{bmatrix}, \begin{bmatrix} \textit{name-rel} \\ \text{IND } \boxed{4} \end{bmatrix}, \begin{bmatrix} \textit{yesterday-rel} \\ \text{IND } \boxed{3} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

- $\boxed{A'} \cap \boxed{D'} = \boxed{B'}$
- $\boxed{B'} \cup \boxed{C'} = \boxed{D'}$

Bag unification: Let A, B and C be bags of elementary predications. A is the unification of B and C iff each element R in A, Q in B and S in C appears in either: (I) a pair <R,Q> where R and Q have the same description, (II) a pair <R,S> where R and S have the same description or (III) a triple <R,Q,S> where R is the unification of Q and S. No element appears in more than one tuple.

Bag intersection: Let A, B and C be bags of elementary predications. A is the intersection of B and C iff (I) B is a possible result of the unification of A and B (i.e. using only triples), (II) C is a possible result of the unification of A and C and (III) there is bag A’ such that B is a possible result of the unification of A’ and B, C is a possible result of the unification of A’ and C and A’ has more element than A.

The use of condition III in bag intersection is motivated by the fact that VRAs have a greedy interpretation (i.e. they share everything with their antecedent except what is literally introduced by the fragment). Note that in the two meta-constraints, elementary predications can only be unified if they have the exact same type.

The account is integrated within an HPSG grammar using a feature FRAGMENT. The value of the feature FRAGMENT has two features: ANTECEDENT and VARIABLE which are of type *sem-obj* (36).

(36)
$$\begin{bmatrix} \text{MTR} & \begin{bmatrix} \text{SEM} & \begin{bmatrix} \text{RELS} & \boxed{1} \end{bmatrix} \\ \text{C-SEM} & \begin{bmatrix} \text{RELS} & \boxed{2} \end{bmatrix} \\ \text{FRAGMENT} & \begin{bmatrix} \text{ANTECEDENT} & \begin{bmatrix} \text{RELS} & \boxed{3} \end{bmatrix} \\ \text{VARIABLE} & \begin{bmatrix} \text{RELS} & \boxed{4} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{DTRS} & \left\langle \begin{bmatrix} \text{SEM} & \begin{bmatrix} \text{RELS} & \boxed{5} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{SEM} & \begin{bmatrix} \text{RELS} & \boxed{6} \end{bmatrix} \end{bmatrix}, \ldots \right\rangle \end{bmatrix}$$

- $\boxed{3} \cap \boxed{1} = \boxed{4}$
- $\boxed{4} \cup \text{union}\left(\boxed{5},\boxed{6},\boxed{2}\right) = \boxed{1}$

Additional constraints on the semantics of fragments come from the syntax-semantics interface. Constraints can be imposed on by clusters or by the VRA-construction. Incomplete meaning for the fragment (i.e. bags of elementary predications that cannot be mapped into a connex graph) should be excluded as well.

5.2 A theory of clusters

Clusters are sequences of phrases which are not related by functions but nevertheless display constituent properties. Clusters are fragments but the reverse is not always true. The fragment *when* in (32a), for instance, is composed of a single phrase.

Clusters do not always have the same distribution as a headed phrase with a similar semantic content. For French, it has been noticed that some items or constructions are compatible with clusters which are sentential fragments but not with clauses. This is the case of the conjunction *ainsi que* 'as well as', which may combine with a cluster but not with a finite clause (Abeillé and Godard 1996, Mouret 2006).

(37) Paul offrira un livre à Marie, ainsi qu'(*il offrira) un CD à Paul.
'Paul will offer a book to Mary, as well as (he will offer) a CD to Paul.'

Some cluster types are submitted to form constraints which instantiate lexical knowledge about subcategorization properties of lexical items which are not realized within the cluster.

That property leads Ginzburg and Sag 2000 to analyze sentential fragments as single daughters of a phrase with full clausal properties including the syntactic category VERB. This is problematic for VRAs because sentential fragments and clauses do not have the same distribution. We use the definition given by Mouret 2006 (38). The feature CLUSTER allows one to select or subcategorize for a sign which is a cluster.

(38) $$\text{cluster-cxt} \Rightarrow \text{phrasal-cxt} \;\&\; \begin{bmatrix} \text{MTR} & \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \textit{cluster} \\ \text{CLUSTER } \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{DTRS} & \boxed{1}\ \textit{list}\left(\textit{sign}\right) \end{bmatrix}$$

Several constraints can be expressed on clusters, such as the fact that they must contain a noun phrase in the citation form or that their form instantiates knowledge about the subcategorization of a word which is not present in the structure (39).

(39) NP-cluster-cxt $\Rightarrow$ cluster-cxt &

$$\begin{bmatrix}\text{MTR} & \begin{bmatrix}\text{SYN} & \begin{bmatrix}\text{CAT} & \begin{bmatrix}\text{CLUSTER} & \textit{contains}\left(\begin{bmatrix}\text{SYN} & \begin{bmatrix}\text{CAT} & \begin{bmatrix}\textit{noun} & \\ \text{CASE} & \textit{nominative}\end{bmatrix} \\ \text{MRKG} & \textit{det}\end{bmatrix}\end{bmatrix}\right)\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

We will assume that VRAs with only one phrase in the body have a unary cluster body because it allows one to express generalizations in a simplified way.

5.3 A theory of locality of selection

The selection properties of VRAs and the selection properties of their fronted phrase are submitted to the same kind of locality constraints. When a VRA modifies a clause, the licenser of the VRA must be a direct dependent of the clause's head. When a fronted phrase modifies a cluster, the phrase introducing a subpart of the plural entity denoted by the antecedent must be a direct daughter of the cluster.

In order to state these locality constraints on selection, we introduce a set-valued feature ANCHORS which contains indices of semantic entities which are accessible to adjunct selection. This feature was originally proposed by Kiss 2005 to analyze extraposed relative clauses. The two following constraints on the propagation of anchors are introduced. Constraint (40) ensures that only direct dependents of a clause's head are accessible via the anchor set on the level of the clause. In (40) and (41), *c* stands for the predicate *contains* which is a relation expressing that a list contains some element. In both constraints, the element contained in the antecedent of the constraint in universally quantified.

(40)

$$\begin{bmatrix}\textit{word} & \\ \text{DEPS } c\left(\begin{bmatrix}\text{SEM} & \begin{bmatrix}\text{HOOK} & \begin{bmatrix}\textit{anchor} & \\ \text{LABEL} & \boxed{1} \\ \text{IND} & \boxed{2}\end{bmatrix}\end{bmatrix}\end{bmatrix}\right)\end{bmatrix} \Rightarrow \begin{bmatrix}\text{MTR} & \begin{bmatrix}\text{SEM} & \begin{bmatrix}\text{ANCHORS } c\left(\begin{bmatrix}\textit{anchor} & \\ \text{LABEL} & \boxed{1} \\ \text{IND} & \boxed{2}\end{bmatrix}\right)\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

Constraint (41) ensures that only direct daughters of a cluster are accessible via the anchor set on the level of the clause. Another virtue of the anchor constraints is that they restrict semantic selection to material which is literally introduced in the cluster. Thus, reconstructed semantic relations are not available for semantic selection.

(41)

$$\begin{bmatrix}\textit{cluster-cxt} & \\ \text{DTRS } c\left(\begin{bmatrix}\text{SEM} & \begin{bmatrix}\text{HOOK} & \begin{bmatrix}\textit{anchor} & \\ \text{LABEL} & \boxed{1} \\ \text{IND} & \boxed{2}\end{bmatrix}\end{bmatrix}\end{bmatrix}\right)\end{bmatrix} \Rightarrow \begin{bmatrix}\text{MTR} & \begin{bmatrix}\text{SEM} & \begin{bmatrix}\text{ANCHORS} & c\left(\begin{bmatrix}\textit{anchor} & \\ \text{LABEL} & \boxed{1} \\ \text{IND} & \boxed{2}\end{bmatrix}\right)\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

5.4 A theory of VRAs

We analyze VRA constructs as subtypes of the head-functor construct as defined in Van Eynde 2003 and Sag 2007. Functors have two main properties. They select the head sign and contribute a mark to the construct (42).

(42) hd-func-cxt $\Rightarrow$ hd-cxt &

$$\begin{bmatrix} \text{MTR} & \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{VAL} & \boxed{1} \\ \text{MRKG} & \boxed{2} \end{bmatrix} \end{bmatrix} \\ \text{DTRS} & \left\langle \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{SELECT} & \boxed{3} \end{bmatrix} \\ \text{MRKG} & \boxed{2} \end{bmatrix} \end{bmatrix}, \boxed{3} \right\rangle \\ \text{HD-DTR} & \boxed{3} \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{VAL} & \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The VRA construct itself is defined in (43). The head of the construct is a sentential fragment. It is selected for by the fronted phrase which contributes a sum-subpart relation which is characteristic for the construction. The sum-subpart relation is assumed to have a partitioning subtype and an exemplifying subtype. The construction itself contributes a second subpart relation which links the event denoted by the host clause to that introduced by the sentential fragment. The construction also selects a nominal licenser. Notice the use of the set ANCHORS to express the locality of selection of the fronted phrase and of the VRA construct.

(43) VRA-cxt $\Rightarrow$ phrasal-cxt & hd-fun-cxt &

$$\begin{bmatrix} \text{MTR} & \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{CAT} & \boxed{1} \begin{bmatrix} \text{SELECT} & \begin{bmatrix} \text{SEM} & \begin{bmatrix} \text{ANCHORS} & \textit{contains}(\begin{bmatrix} \text{IND} & \boxed{2} \end{bmatrix}) \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \boxed{3} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{HOOK} & \begin{bmatrix} \text{IND} & \boxed{7} \end{bmatrix} \end{bmatrix} \\ \text{C-SEM} & \begin{bmatrix} \text{RELS} & \left\langle \begin{bmatrix} \textit{sum-subpart-rel} & \\ \text{SUBPART} & \boxed{7}\ \textit{event} \\ \text{SUM} & \boxed{6}\ \textit{event} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \\ \text{HD-DTR} & \boxed{4} \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{CAT} & \boxed{1} \\ \text{VAL} & \boxed{3}\langle\rangle \\ \text{MRKG} & \textit{none} \end{bmatrix} \\ \text{SEM} & \boxed{B} \begin{bmatrix} \text{HOOK} & \begin{bmatrix} \text{IND} & \boxed{7} \end{bmatrix} \\ \text{ANCHORS} & \textit{contains}(\begin{bmatrix} \text{IND} & \boxed{8} \end{bmatrix}) \end{bmatrix} \\ \text{FRAGMENT} & \begin{bmatrix} \text{ANTECEDENT} & \boxed{A} \begin{bmatrix} \text{HOOK} & \begin{bmatrix} \text{IND} & \boxed{6} \end{bmatrix} \end{bmatrix} \\ \text{VARIABLE} & \boxed{C} \end{bmatrix} \\ \text{GAP} & \{\} \end{bmatrix} \\ \text{DTRS} & \left\langle \begin{bmatrix} \text{SEM} & \begin{bmatrix} \text{HOOK} & \begin{bmatrix} \text{LTOP} & \boxed{5} \end{bmatrix} \\ \text{RELS} & \left\langle \begin{bmatrix} \textit{sum-subpart-rel} & \\ \text{LABEL} & \boxed{5} \\ \text{SUBPART} & \boxed{8} \\ \text{SUM} & \boxed{2} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}, \boxed{4} \right\rangle \end{bmatrix}$$

WH-VRA-construct is a subtype of VRA-construct. The fronted phrase of a VRA-construct contains a WH form which is coreferential with the nominal antecedent of the VRA-construct (44).

(44) WH-VRA-cxt ⇒ VRA-cxt &

$$\begin{bmatrix} \text{MTR} & \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{SELECT} & \begin{bmatrix} \text{SEM} & \begin{bmatrix} \text{ANCHORS} & \textit{contains}\left(\begin{bmatrix} \text{IND } \boxed{1} \end{bmatrix}\right) \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{DTRS} & \left\langle \begin{bmatrix} \text{WH} & \left\{ \begin{bmatrix} \text{HOOK} & \begin{bmatrix} \text{IND } \boxed{1} \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix}, \textit{sign} \right\rangle \end{bmatrix}$$

The prepositions functioning as the head of the fronted phrase of a WH-VRA-construct have the following lexical properties (45). They have an argument structure containing two elements, one of which is an internal argument realized as the complement of the preposition. The other argument is the external argument of the preposition and is not realized as a dependent of the preposition. Rather the preposition selects for a phrase which contains an anchor coindexed with its external argument. Finally the preposition has to introduce a sum-subpart relation between its two arguments: the internal argument denoting a sum and the external a subpart of that sum. Prepositions may differ regarding the exact type of sum-subpart relation. Some introduce an exemplifying relation (Fr. *parmi*), some a partitioning one (Ro. *dintre*), some an underspecified one (Ro. *printre* and *între*).

(45) PARMI-word ⇒ word &

$$\begin{bmatrix} \text{ARG-ST} & \left\langle \boxed{1} \begin{bmatrix} \text{syn} & \begin{bmatrix} \text{cat} & \textit{noun} \\ \text{VAL} & \langle\rangle \\ \text{MRKG} & \textit{det} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{HOOK} & \begin{bmatrix} \text{IND} & \boxed{3} \end{bmatrix} \end{bmatrix} \end{bmatrix}, \boxed{2} \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{CAT} & \textit{noun} \\ \text{VAL} & \langle\rangle \\ \text{MRKG} & \textit{det} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{HOOK} & \begin{bmatrix} \text{IND} & \boxed{4} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \\ \text{SYN} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \textit{preposition} & \\ \text{XARG} & \boxed{1} \\ \text{SELECT} & \begin{bmatrix} \text{SEM} & \begin{bmatrix} \text{ANCHORS} & \textit{contains}(\begin{bmatrix} \text{IND } \boxed{3} \end{bmatrix}) \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \langle \boxed{2} \rangle \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{HOOK} & \begin{bmatrix} \text{LTOP} & \boxed{5} \end{bmatrix} \\ \text{RELS} & \begin{bmatrix} \textit{exemplifying-sum-subpart-rel} & \\ \text{LABEL} & \boxed{5} \\ \text{SUBPART} & \boxed{3} \\ \text{SUM} & \boxed{4} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

We assume that *dont* is a marker which has no argument structure but introduces an underspecified sum-subpart relation, although it could also be introduced constructionally. *Dont* selects for a phrase which (I) contains an anchor for the subpart argument of its sum-subpart relation and (II) selects for a phrase containing an anchor for the sum argument of its sum-subpart relation.

(46) DONT-word ⇒ word &

$$\begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{SELECT} & \begin{bmatrix} \text{SYN} \mid \text{CAT} \mid \text{SELECT} \mid \text{SEM} \mid \text{ANCHORS } \textit{contains}\left\{\begin{bmatrix}\text{IND} & \boxed{4}\end{bmatrix}\right\} \\ \text{SEM} \mid \text{ANCHORS } \textit{contains}\left\{\begin{bmatrix}\text{IND} & \boxed{3}\end{bmatrix}\right\} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \langle\rangle \\ \text{MRKG} & \textit{dont} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{HOOK} & \begin{bmatrix}\text{LTOP} & \boxed{5}\end{bmatrix} \\ \text{RELS} & \left\{\begin{bmatrix} \textit{sum-subpart-rel} & \\ \text{LABEL} & \boxed{5} \\ \text{SUBPART} & \boxed{3} \\ \text{SUM} & \boxed{4} \end{bmatrix}\right\} \end{bmatrix} \end{bmatrix}$$

With the entry in (46) for *dont*, the only thing that must be stated in the subtype DONT-VRA-cxt is that the mother of the construct has a feature MRKG whose value is *dont* (47).

(47) DONT-VRA-cxt ⇒ VRA-cxt & $\begin{bmatrix}\text{MTR} & \begin{bmatrix}\text{SYN} & \begin{bmatrix}\text{MRKG} & \textit{dont}\end{bmatrix}\end{bmatrix}\end{bmatrix}$

Subtypes of VRAs must include constraints on the syntax of the head. For example, French WH VRAs must constrain one of the phrases in the cluster to be a noun phrase and link the index of that noun phrase to the subpart feature of the sum-subpart relation expressed by the fronted phrase.

6 Conclusion

VRAs are incidental adjuncts. Like other incidental adjuncts, they are licensed within a phrase, as long as their selection properties are satisfied. Adjacency between VRAs and their sum-denoting licenser is not always required, but locality constraints can be formulated nonetheless. VRAs are not elliptical relative clauses. They are sentential fragments which function as adjuncts with two different kinds of partitive semantics (exemplifying vs. partitioning). The partitive semantics is enforced by the presence of a fronted phrase which displays selection properties regarding the phrase it combines with to form the VRA. The body of a VRA can contain a cluster of phrases. Clusters are submitted to internal form constraints which cannot be explained by reconstruction because their distributional properties are distinct from those of their alleged source. Instead, the instanciation of subcategorization knowledge without the realization of a head must be allowed under description anaphora.

References

Abeillé, A., Clément, L. and Toussenel, F. 2003. Building a treebank for French. In A. Abeillé (ed.), *Treebanks*, Dordrecht: Kluwer.

Abeillé, A. and Godard, D. 1996. La complémentation des auxiliaires en français. *Langages* 122, 32–61.

Arnold, D. 2004. Non-Restrictive Relative Clauses in Construction Based HPSG. In St. Müller (ed.), *Proceedings of the HPSG04 Conference*, pages 27–47, Stanford: CSLI Publications.

Arnold, D. and Borsley, R. D. 2008. Non restrictive relative clauses, ellipsis and anaphora. In St. Müller (ed.), *Proceedings of the HPSG08 Conference*, pages 5–25, Stanford: CSLI Publications.

Bonami, O. and Godard, D. 2003. Incidental adjuncts: an overlooked type of adjunction. Presentation given at *the 10th International Conference on HPSG*. East Lansing: Michigan State University.

Copestake, A., Flickinger, D., Pollard, C. and Sag, I. A. 2005. Minimal Recursion Semantics: An Introduction. *Research on Language and Computation* 3(4), 281–332.

Fernández, R., Ginzburg, J. and Lappin, Sh. 2007. Classifying Non-Sentential Utterances in Dialogue: A Machine Learning Approach. *Computational Linguistics* 33(3), 397–427.

Gheorghe, M. 2004. *Propoziţia relativă*. Bucharest: Editura Paralela 45.

Gheorghe, M. 2005. Construcţii cu propoziţii relative. In *Gramatica limbii române*, volume II, Bucharest: Editura Academiei Române.

Ginzburg, J. and Sag, I. A. 2000. *Interrogative Investigations: the form, meaning and use of English interrogatives*. Stanford: CSLI Publications.

Godard, D. 1988. *Syntaxe des relatives en français*. Paris: CNRS Editions.

Grevisse, M. 1993. *Le Bon Usage*. Boeck Duculot, 13. edition.

Kiss, T. 2005. Semantic constraints on relative clause extraposition. *Natural Language and Linguistic Theory* 23(2), 281–334.

Lasersohn, P. 1995. *Plurality, Conjunction and Events*. Dordrecht: Kluwer.

Mouret, F. 2006. A Phrase Structure Approach to Argument Cluster Coordination. In St. Müller (ed.), *Proceedings of the HPSG06 Conference*, pages 247–267, Stanford: CSLI Publications.

Sag, I. A. 2007. Sign-Based Construction Grammar: An informal synopsis. Manuscript. Stanford University.

Van Eynde, F. 2003. On the Notion 'Determiner'. In St. Müller (ed.), *Proceedings of the HPSG03 Conference*, pages 391–396, Stanford: CSLI Publications.

Inflectional periphrasis in Persian

Olivier Bonami

Université Paris-Sorbonne &
Laboratoire de Linguistique Formelle

Pollet Samvelian

Université Paris 3 Sorbonne nouvelle &
Mondes iranien et indien

Proceedings of the HPSG09 Conference

Georg-August-Universität Göttingen, Germany

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

Modern Persian conjugation makes use of five periphrastic constructions. We contrast the properties of these five constructions and argue that they call for different analyses. We propose contrasting analyses relying on the combination of an HPSG approach to feature geometry and syntactic combination, and an approach to paradigm organization and morphological exponence based on Paradigm Function Morphology. This combination of analytic tools allows us to treat the whole array of periphrastic constructions as lexical in origin—no phrasal construction or multi-word lexical entry of any kind is required.

Grammars of Persian (e.g. Lazard et al., 2006) distinguish five conjugational periphrastic construction types. The passive construction is based on an inflected form of *šodan* 'become' preceded by a perfect participle (1). So-called 'perfect' forms are based on an inflected form of *budan* 'be' preceded by a perfect participle (2). The auxiliary is a full word (2a) or a clitic, (2b) depending on tense and mood, and giving rise to different syntactic and semantic properties. The future is formed with a special present tense form of *xâstan* 'want' followed by a bare stem (3). Finally, the progressive is based on an inflected form of *dâštan* 'have' followed by a finite form (4).[1]

(1) In tâblo foruxte mi-šav-ad.
this painting sold UNBD-become.S1-3SG
'This painting is sold.'

(2) a. Maryam in tâblo=râ foruxte bud.
Maryam this painting=DDO sold be.S2.3SG
'Maryam had sold this painting.'

b. Maryam in tâblo=râ foruxe=ast.
Maryam this painting=DDO sold=be.PRS.3SG
'Maryam has sold this painting.'

(3) Maryam in tâblo=râ xâh-ad foruxt.
Maryam this painting=DDO want.S1-3SG sell.S2
'Maryam will sell the painting'

[†]Aspects of this work have been presented at the *Décembrettes 6 International Morphology Conference* (December, 2008), at the HPSG Seminar at U. Paris Diderot (March, 2009), at a Morphology Meeting in Surrey (April, 2009), and at the HPSG 2009 Conference (Göttingen, July 2009). We thank for their comments and suggestions the audiences at these events, and in particular Anne Abeillé, Gilles Boyé, Dunstan Brown, Patricia Cabredo Hofherr, Greville Corbett, Berthold Crysmann, Gerald Gazdar, Stefan Müller, Ivan A. Sag, Gregory Stump, Jesse Tseng, and Gert Webelhuth. This work was supported by a grant from Agence Nationale de la Recherche and Deutsche Forschungsgemeinschaft to the Franco-German project 'PER-GRAM: Theory and Implementation of a Head-driven Phrase Structure Grammar for Persian'.

[1]The glosses use the following abbreviations. BD: bounded aspect; DDO: definite direct object; EZ: Ezafe; NEG: negation; PAF: pronominal affix; PRF: perfect; PRS: present; PST: past; S1: first stem (a.k.a. the present stem); S2: second stem (a.k.a. the past stem); SBJV: subjunctive; UNBD: unbounded aspect.

(4) Maryam dâr-ad in tâblo=râ mi-foruš-ad.
Maryma have.PRS-3SG this painting=DDO UNBD-sell.S1-3SG
'Maryam is selling the painting.'

The differing properties of these five types of periphrasis stem from different origins as finite, infinitival or participial complements, and different degrees of grammaticalization, going from the quasi-analytic passive to the recently morphologized present perfect, through truly periphrastic forms that need to be integrated into inflectional paradigms despite being multi-word expressions. We assume that the different properties call for different analyses. We propose five contrasting analyses relying on the combination of an HPSG approach to feature geometry and syntactic combination, and an approach to paradigm organization and morphological exponence based on Paradigm Function Morphology (PFM; Stump, 2001). Interestingly, this combination of analytic tools allows us to treat the whole array of periphrastic constructions as lexical in origin—no phrasal construction or multi-word lexical entry of any kind is required.

1 Synthetic conjugation in HPSG/PFM

Before we address the analysis of periphrastic forms, we start with an account of synthetic conjugation. (5) lists the synthetic subparadigms of the lexeme *xaridan* 'buy', using the positive 2SG form as an illustration.

(5) a. Finite forms:
   i. Simple present: *mi-**xar**-i*
   ii. Simple bounded past: ***xarid**-i*
   iii. Simple unbounded past: *mi-**xarid**-i*
   iv. Simple subjunctive: *be-**xar**-i*
   v. Imperative: *be-**xar***
b. Nonfinite forms:
   i. Infinitive: ***xarid**-an*
   ii. Present participle: ***xar**-ande*
   iii. Perfect participle: ***xarid**-e*
   iv. Gerund: ***xar**-ân*

Persian verbs exhibit a morphomic stem alternation (here *xar* vs. *xarid*). Neither stem is predictable from the other in general, and both stems are used in a combination of contexts which do not form a natural class. Affixal exponents realize unbounded aspect in the indicative (*mi-*), irrealis mood (*be-*), negation (*na-* or *ne-*, not illustrated here), type of nonfinite form (*-e* vs. *-ande* vs. *-an* vs. *-ân*), and subject agreement for finite forms. Within Paradigm Function Morphology, this rather simple position class system can be accounted for using the series of rule

III	II	I	IV	V
na-	*mi-*	stem-selection	*-e*	*-am*
ne-			*-ande*	*-i*/$\emptyset$
be-			*an*	*-ad*/$\emptyset$
				-im
				-id
				-and

Table 1: Rule blocks for Persian synthetic conjugation

blocks outlined in table 1. Remember that in PFM, realization rules are organized in successive blocks. When attempting to realize a given set of morphosyntactic feature, the most specific applicable rule within the block is chosen. (6) are sample rules from block V, written in an attribute-value matrix format.[2]: while (6a) asks that finite verbs with a 2SG subject take the suffix *-i*, the more specific (6b) indicates that the suffix is dropped in the imperative.

(6) a. $$\begin{bmatrix}\textsc{phon} & X \\ \textsc{lid} & Y\end{bmatrix}, \sigma : \begin{bmatrix}\textsc{per} & \mathit{2} \\ \textsc{nb} & \mathit{sg}\end{bmatrix} \longrightarrow \begin{bmatrix}\textsc{phon} & X \oplus \mathrm{i} \\ \textsc{lid} & Y\end{bmatrix} \qquad \text{(block V)}$$

b. $$\begin{bmatrix}\textsc{phon} & X \\ \textsc{lid} & Y\end{bmatrix}, \sigma : \begin{bmatrix}\textsc{per} & \mathit{2} \\ \textsc{nb} & \mathit{sg} \\ \textsc{mood} & \mathit{imper}\end{bmatrix} \longrightarrow \begin{bmatrix}\textsc{phon} & X \\ \textsc{lid} & Y\end{bmatrix} \qquad \text{(block V)}$$

Since the integration of HPSG and PFM will be essential to our account of periphrastic conjugation, it is important that we specify how we intend to do it. The task is not trivial, because of PFM's reliance on comparisons of feature structure descriptions, which can not easily be formulated in existing description languages for HPSG grammars. Rather than attempting a direct integration, we propose to use a PFM grammar to further constrain the class of signs satisfying an HPSG theory. Specifically, we rely on a slight reorganization of the feature geometry for head values as in (7), where MORSYN groups features that get realized in inflection and LID assigns a specific index to each lexeme (Spencer, 2005; Sag, 2007). We then define a version of PFM that is exactly like that of (Stump, 2001) except for the fact that typed feature structures are used to model morphosyntactic feature bundles instead of category structures *à la* (Gazdar et al., 1985). The meta-constraint in (8) then links the two grammars.

[2]Two different conventions are currently used to write PFM rules, defined respectively in (Stump, 2001) and (Ackerman and Stump, 2004). The AVM format we use here is meant to ease the integration with HPSG, although the change is little more than syntactic sugar.

(7) $head \rightarrow \begin{bmatrix} head & \\ \textsc{lid} & lexemic\text{-}index \\ \textsc{morsyn} & morsyn \end{bmatrix}$

(8) **Morphology-syntax interface (preliminary version)**

A sign of type *word* meeting the description $\begin{bmatrix} \textsc{phon} & \boxed{1} \\ \textsc{head} & \begin{bmatrix} \textsc{lid} & \boxed{3} \\ \textsc{morsyn} & \boxed{4} \end{bmatrix} \end{bmatrix}$ is well-formed only if the PFM grammar licenses phonology $\boxed{1}$ as a realization of the features $\boxed{4}$ for the lexeme $\boxed{3}$.

2 The passive

The passive in Persian is a typical complex predicate construction, whose properties are parallel to those of copula-predicative complement constructions. The auxiliary *šodan* is clearly the head: all inflectional information, e.g. negation (9), is realized on the auxiliary. The participle-auxiliary sequence is syntactically flexible: adverbs may intervene (10), the auxiliary may be scrambled over the participle (11), and long-distance fronting of the participle is possible (12).

(9) In tâblo foruxte ne-mi-šav-ad.
this painting sold NEG-UNBD-become.S1-3.SG
'This painting is not sold.'

(10) In tâblo foruxte hatman šod.
this painting sold certainly become.S2
'This painting was certainly sold.'

(11) In tâblo šod robude va foruxte.
this painting become.S2 stolen and sold
'It is this painting which was stolen and sold.'

(12) Foruxte fekr mi-kon-am [agar in tâblo __
sold thought UNBD-do.S1-1SG if this painting
be-šav-ad, mi-tavân-im bâ pul-aš yek
SBJV-become.S1-3SG UNBD-can.S1-1PL with money-PAF.3SG a
mâšin be-xar-im].
car SBJV-buy.S1-1PL
'I think that if this painting is sold, we can buy a car with the money.'

To account for this we rely on an argument composition analysis in the spirit of (Hinrichs and Nakazawa, 1994) and subsequent work. Specifically we propose the lexical entry in (13) for the auxiliary lexeme *šodan*, giving rise to analyses such as that in Figure 1. Under our analysis there is no passive participle, and subject demotion is effected directly in the auxiliary's entry. This is appropriate

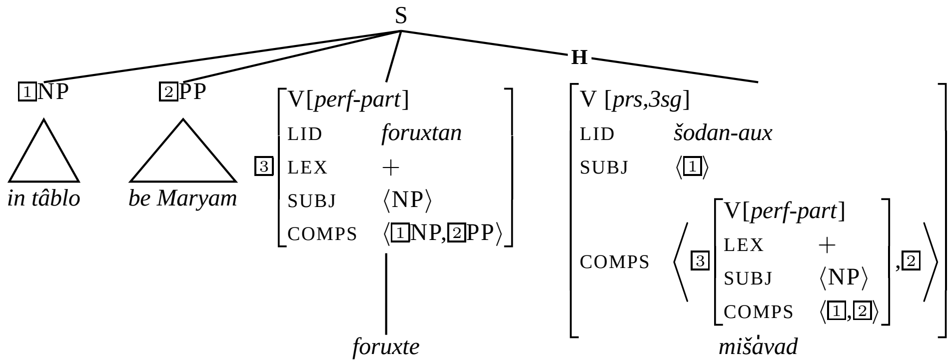


Figure 1: Analysis of a passive sentence

because (i) perfect participles are always active except in the periphrastic passive constructions—participial clauses with transitive head verbs take direct objects (14), and (ii) for semantic reasons there is no hope of using the same lexical entry for the auxiliary *šodan* and the full verb *šodan* (contrary to what happens in languages where the passive auxiliary coincides with the copula). Moreover, we assume a flat structure, wich allows for an easy account of the free reordering of the participle, auxiliary and valents. The specification [VC −] on the participle inhibits the formation of a verbal complex—see below for a contrasting analysis of perfect periphrases.

(13)
$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix}\text{LID} & \textit{šodan-aux}\end{bmatrix} \\ \text{CONT} & \boxed{2} \\ \text{ARG-ST} & \left\langle \boxed{1}, \begin{bmatrix} \text{FORM} & \textit{part} \\ \text{PERFECT} & + \\ \text{POL} & + \\ \text{CONT} & \boxed{2} \\ \text{ARG-ST} & \langle \text{NP}, \boxed{1} \rangle \oplus \boxed{L} \\ \text{LEX} & + \\ \text{VC} & - \end{bmatrix} \right\rangle \oplus \boxed{L} \end{bmatrix}$$

(14) Maryam tâblo=râ xarid-e va be Omid dâd.
Maryam painting=DDO buy.S2-PRP and to Omid give.S2
'Having bought the painting, Maryam gave it to Omid.'

Notice that under our analysis voice is not an inflectional category in Persian: the active-passive opposition is dealt with entirely within syntax.

3 Two sets of forms based on *budan*

There are five different subparadigms based on *budan*, illustrated here in (15). These contrast in two independent ways.

(15) a. Complex present: *xaride=i*
b. Complex bounded past: *xaride bud-i*
c. Complex unbounded past: *mixaride=i*
d. Complex subjunctive: *xaride bâš-i*
e. Compex perfect: *xaride bude=i*

3.1 Morphologized vs. truly periphrastic forms

In the complex present and the complex unbounded past, the perfect participle combines with the present clitic form of the auxiliary, which is homophonous with the exponent of subject agreement except for 3SG (there is also a nonclitic form of present *budan*, but it may not be used in this construction). In the complex bounded past and complex subjunctive, the perfect participle combines respectively with the bounded past and subjunctive forms of the auxiliary. Finally the complex perfect cumulates two forms of the auxiliary: the participle *bude* and the present form clitic (here =*i*).

There is strong evidence that the forms historically based on the clitic auxiliary have undergone morphologization in contemporary Persian. First, the sequence cannot be interrupted in any way; in particular, adverbs are excluded (16), as is participle fronting (17). Second, the distribution of the unbounded aspect marker *mi-* is otherwise unexplainable: it is the full construction, not the participle, that is unbounded. Finally, colloquial Persian allows a form of vowel reduction in the 3SG that is peculiar to these forms (18a): comparable contructions where the clitic auxiliary combines with an adjective do not give rise to the same pattern (18b).[3]

(16) *Rafte hatman=ast.
left certainly=be.S1.3SG
'(S)he has certainly left.'

(17) *Ne-mi-rafte sâlhâ Maryam be madrase=ast.
NEG-UNBD-gone years Maryam to school=be.S1.3SG
'For years, Maryam didn't go to school'

(18) a. mord'e=ast → mord'e:
died=be.S1.3SG
'(S)he has died.'

[3]The only piece of evidence pointing in the other direction is the possibility for the auxiliary to have wide scope over a coordination of participles. However the existence of sublexical coordination in numerous languages calls into question whether this is a strong argument against a morphological analysis. We leave this issue for future research.

b. mordˈe=ast → mordˈast
corpse=be.S1.3SG
'It is a corpse.'

Compare now the situation of forms that are based on a nonclitic auxiliary. The participle-auxiliary combination is more constrained than it is in the passive; in particular, neither adverbs (19) nor pronominal affixes (20) can occur between the two verb forms, and negation must be realized on the participle (21). In addition, scrambling is excluded (22). However, the combination is not lexical, since the participle can be extracted (23).

(19) * Maryam dide hatman bud-aš
Maryam seen certainly be.S2-PAF.3SG

(20) a. Maryam dide budaš.
Maryam seen be.S2-PAF.3SG
'Maryam had seen him.'

b. * Maryam dide-aš bud.
Maryam seen-PAF.3SG be.S2

(21) Maryam Omid-râ na-dide bud.
Maryam Omid-DDO NEG-seen be.S2
'Maryam hadn't seen Omid.'

(22) * Maryam Omid-râ bud dide.
Maryam Omid-DDO be.S2 seen

(23) Foruxte fekr ne-mi-kon-am [_ bâš-ad in
sold thought NEG-UNBD-do.S1-1SG be.SBJV-3SG this
tâblo=râ].
painting=DDO
'I don't think that s/he has sold this painting.'

3.2 Morphosyntactic import

The use of a form based on *budan* may realize two distinct morphosyntactic features. The complex bounded past (24) and complex subjunctive (25) express respectively the past perfect and the subjunctive perfect. The complex unbounded past however does not express perfectivity at all. Rather, it has an evidential value (Windfuhr, 1982; Lazard, 1985; Jahani, 2000). Whereas the simple bounded past is used when the speaker has direct evidence for what she is asserting, the complex bounded past is used in contexts where the evidence is only indirect, as in (26).

(24) Qabl az inke Omid be-res-ad, Maryam birun rafte bud.
before from that Omid SBJV-arrive.S1-3SG Maryam out gone be.S2
'Maryam had left (before Omid arrived).'

(25) Fekr mi-kon-am Maryam mariz bude baš-ad.
thought UNBD-do.S1-1SG Maryam sick been be.SBJV-3SG
'I think Maryam has been sick.'

(26) (Banâ bar gofte-ye Omid) Maryam dar sâl-e 1950 in xâne-râ
According to-EZ Omid Maryam in year-EZ 1950 this house-DDO
mi-sâxte=ast.
UNBD-built=be.S1.3SG
'According to Omid, Maryam was building this house in 1950.'

The complex present is ambiguous between a perfect and an evidential value: it can be interpreted either as a present perfect (27a) or as a bounded past with indirect evidentiality (27b). Finally, the complex perfect expresses both perfectivity and indirect evidentiality: it is the indirect evidential equivalent of the complex bounded past (28). Note that this corresponds transparently to the fact the the complex perfect includes two realizations of the copula.

(27) a. Maryam tâze reside=ast.
Maryam new arrived=be.S1.3SG
'Maryam has just arrived.'

b. (Banâ bar gofte-ye Omid) Maryam in xâne-râ dar sâl-e 1950
According to-EZ Omid) Maryam this house-DDO in year-EZ 1950
xaride=ast.
bought=be.S1.3SG
'According to Omid, Maryam bought this house in 1950.'

(28) (Az qarâr), qabl az inke Omid be-res-ad, Maryam birun
apparently before from that Omid SBJV-arrive.S1-3SG, Maryam out
rafte bude=ast.
gone been=be.S1.3SG
'Apparently, Maryam had left before Omid arrived.'

As can be seen in Table 3.2, if the present perfect is ignored, morphosyntactic properties align nicely with morphologized vs. syntactic combinations: the morphologized forms are used for indirect evidentiality, as stated by rules (29); while the truly periphrastic forms are used to express the perfect. The fact that the present perfect is unexpectedly synthetic calls for an paradigmatic analysis: this seems to be a standard case of syncretism, where the exponents used to realize a certain feature set (here indirect bounded past) are reused in some unrelated part of the paradigm. Specifically one should assume a rule of referral along the lines of (30). The rule states that any present perfect form of a lexeme Y will be identical to the indirect bounded past form of Y with the same specifications for all features except tense, aspect and evidentiality (here, the relevant remaining features are person, number and polarity).[4]

[4]This is a portmanteau rule of referral covering blocks I to V, thus bypassing completely synthetic exponence. $\sigma \setminus \tau$ is the description that is identical to σ except where the features mentioned in

	PRESENT	PAST		SBJV
		DIR. EV.	IND. EV.	
BD	***	bounded past	complex present	simple sbjv
UNBD	simple present	unbd past	cpl. unbd. past	
PRF	complex present	complex bnd. past	complex perfect	complex sbjv

Table 2: Morphosyntactic features expressed by Persian subparadigms

(29) a. $\begin{bmatrix} \text{PHON} & X \\ \text{LID} & Y \end{bmatrix}, \sigma : \begin{bmatrix} \text{EVID} & \textit{indir} \end{bmatrix} \longrightarrow \begin{bmatrix} \text{PHON} & X \oplus \mathsf{e} \\ \text{LID} & Y \end{bmatrix}$ (block IV)

b. $\begin{bmatrix} \text{PHON} & X \\ \text{LID} & Y \end{bmatrix}, \sigma : \begin{bmatrix} \text{EVID} & \textit{indir} \\ \text{PER} & \textit{3} \\ \text{NB} & \textit{sg} \end{bmatrix} \longrightarrow \begin{bmatrix} \text{PHON} & X \oplus \mathsf{ast} \\ \text{LID} & Y \end{bmatrix}$ (block V)

(30) $\begin{bmatrix} \text{PHON} & X \\ \text{LID} & Y \end{bmatrix}, \sigma : \begin{bmatrix} \text{TNS} & \textit{prst} \\ \text{PRF} & + \end{bmatrix} \longrightarrow$

$\begin{bmatrix} \text{PHON} & \mathsf{refer}\left(\begin{bmatrix} \text{PHON} & X \\ \text{LID} & Y \end{bmatrix}, \sigma \backslash \begin{bmatrix} \text{TNS} & \textit{pst} \\ \text{ASP} & \textit{bnd} \\ \text{PRF} & - \\ \text{EVID} & \textit{ind} \end{bmatrix}, \text{I-V} \right) \\ \text{LID} & Y \end{bmatrix}$ (blocks I-V)

4 Analyzing the perfect periphrases

We construct the analysis of perfect periphrases in two steps. First we present a syntactic analysis of perfect forms based on argument composition, and show what is unsatisfactory with such an approach. Next we present a way of arriving at the same syntactic analysis by inflectional means. Finally we discuss alternatives and potential problems.

τ differ from those in σ. The function refer takes as arguments an indexed phonological form, a morphosyntactic specification and a rule block sequence, and outputs the result of applying to this indexed phonological form and this morphosyntactic specification the restriction of the PFM grammar to these rule blocks. The motivation for deriving the present perfect from the indirect bounded past rather than the other way around is the economy of paradigms: this allows us to state the rules of exponence realizing suffixes *-e* and *-ast* in a natural way, as applying to all and only evidential forms. Notice that the orientation of the rule of referral might not correspond to the directionality of the diachronic morphologization process.

4.1 A failed analysis based on argument composition

As a first step, we present an analysis that is a variation of the analysis presented above for the passive. (31) is a candidate entry for the present form of the auxiliary *bud*. This states that the auxiliary is a past perfect form which takes a perfect participle complement and inherits the participle's arguments. Because the past participle is marked as [VC +], the auxiliary and participle form a verbal complex, as indicated in figure 2 and thus can not be seperated by elements that are not allowed to occur inside a verbal complex. Rigid word order is a consequence of the LP rule in (32). In addition, since the participle is an argument of the auxiliary, this analysis will allow for the extraction of the participle within any HPSG approach to extraction.

(31)
$$\begin{bmatrix} \text{PHON} & \text{bud} \\ \text{HEAD} & \begin{bmatrix} \text{LID} & \textit{budan-aux} \\ \text{MORSYN} & \begin{bmatrix} \textit{tns} & \textit{pst} \\ \text{PRF} & + \\ \text{AGR} & \begin{bmatrix}\text{PER} & 3 \\ \text{NB} & \textit{sg}\end{bmatrix} \\ \text{POL} & + \end{bmatrix}\end{bmatrix} \\ \text{ARG-ST} & \left\langle \boxed{1}, \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{FORM} & \textit{part} \\ \text{PRF} & + \\ \text{POL} & + \end{bmatrix} \\ \text{LEX} & + \\ \text{VC} & + \\ \text{ARG-ST} & \langle \boxed{1} \rangle \oplus \boxed{L} \end{bmatrix} \right\rangle \oplus \boxed{L} \end{bmatrix}$$

(32)
$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{FORM} & \textit{part} \\ \text{PRF} & + \\ \text{VC} & + \end{bmatrix} \end{bmatrix} \prec [\ ]$$

While this analysis is appropriate as far as syntax is concerned, its integration with the analysis of synthetic conjugation is problematic. First, the perfect auxiliary must be stipulated to be defective for all nonperfect forms, and to have perfect forms that are homonymous to the nonperfect forms of the ordinary copula; thus the purported perfect auxiliary is inflectionally deeply abnormal. Second, we need to derive the fact that there is no present form of the perfect auxiliary (remember that the present perfect is a morphologized form). There are two ways this could be done. We could further stipulate that the perfect auxiliary is defective for

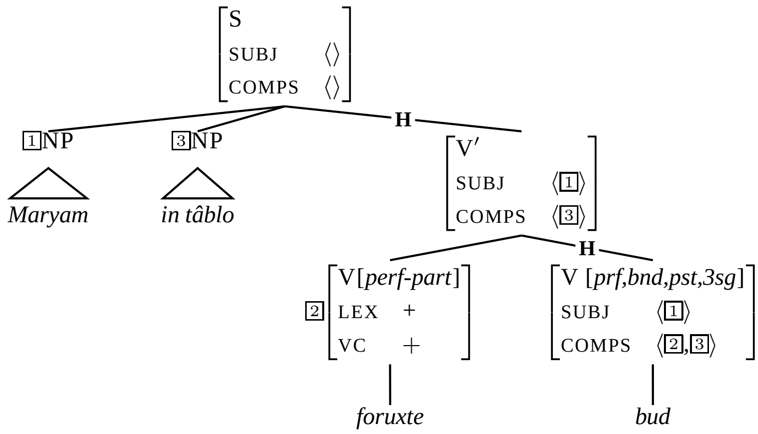


Figure 2: The syntactic structure of perfect periphrases

the present, despite the fact that the ordinary copula it derives from has perfectly good present forms (in fact, two sets of such forms: clitic and nonclitic ones). Or we could assume that some form of competition between morphology and syntax is taking place (Poser, 1992)—but the postulation of such competitions is notoriously difficult to state precisely, and quite alien to the design properties of HPSG. Finally, we need to find a way of stating that the passive auxiliary can not take the perfect auxiliary as its complement: while (33a) is well-formed, (33b) is not.

(33) a. In tâblo foruxte šode bud.
this painting sold become be.PST.3SG
'This painting had been sold.'

b. * In tâblo foruxte bude šod.
this painting sold be become.PST.3SG

While these problems can definitely be circumvented by specifying an appropriately complex inflectional paradigm for the perfect auxiliary, it is striking that many conterintuitive stipulations are needed just because it is not possible to state that the periphrastic perfect is part of the inflectional paradigm of the main verb. The next subsection attempts to modify the framework in a way that allows for the formulation of such an analysis.

4.2 An alternative solution: exponence as valence

As the last subsection stressed, what we need is a way to treat perfect forms as part of the inflectional paradigm (Ackerman and Stump, 2004), while allowing for the fact that they correspond to a combination of two words, one of which may be extracted. The solution we explore here can be stated informally as follows: a perfect form of a lexeme Y is a word whose phonology is borrowed from that of a form of the lexeme *budan*, but which subcategorizes for a perfect participle of this same lexeme Y. For instance, the 3SG positive complex bounded past of *xaridan* meets the description in (34), which is exactly like (31) except for the fact that it is an instance of the lexeme *xaridan*.

(34)
$$
\begin{bmatrix}
\textsc{phon} & \mathsf{bud} \\
\textsc{head} & \begin{bmatrix}
\textsc{lid} & \textit{xaridan} \\
\textsc{morsyn} & \begin{bmatrix} \textit{tns} & \textit{pst} \\ \textsc{prf} & + \\ \textsc{agr} & \boxed{2} \\ \textsc{pol} & + \end{bmatrix}
\end{bmatrix} \\
\textsc{arg-st} & \left\langle \boxed{1}, \begin{bmatrix}
\textsc{head} & \begin{bmatrix} \textit{verb} & \\ \textsc{form} & \textit{part} \\ \textsc{prf} & + \\ \textsc{pol} & + \end{bmatrix} \\
\textsc{lex} & + \\
\textsc{vc} & + \\
\textsc{arg-st} & \langle \boxed{1} \rangle \oplus \boxed{L}
\end{bmatrix} \right\rangle \oplus \boxed{L}
\end{bmatrix}
$$

The challenge now is to derive (34) in a principled way, while integrating it within an inflectional system where perfect forms may be realized either synthetically or periphrastically. The approach we propose is based on an extension of the power of realization rules in the spirit of (Spencer, 2005). In classical PFM, realization rules relate phonology-lexemic index pairs to phonology-lexemic index pairs. We propose that valence lists be added to the picture: realization rules now relate triplets of a phonological representation, a lexemic index, and an argument structure specification. The meta-constraint in (8) is updated as in (35), so that argument structure is examined at the morphology-syntax interface.[5]

(35) **Morphology-syntax interface (preliminary version)**

A sign of type *word* meeting the description
$$
\begin{bmatrix}
\textsc{phon} & \boxed{1} \\
\textsc{arg-st} & \boxed{2} \\
\textsc{head} & \begin{bmatrix} \textsc{lid} & \boxed{3} \\ \textsc{morsyn} & \boxed{4} \end{bmatrix}
\end{bmatrix}
$$
is well-formed only if the PFM grammar licenses phonology $\boxed{1}$ and arguments $\boxed{2}$ as a realization of the features $\boxed{4}$ for the lexeme $\boxed{3}$.

The rule licensing (34) is given in (36). To realize a feature structure σ verifying [PRF $+$], one should refer the phonology to that of the corresponding bounded

[5]The formulation of this constraint presupposes that the HPSG grammar says nothing about individual lexical entries, and that most of the usual HPSG theory of the lexicon is recast as part of the morphological component.

positive nonperfect form of *budan*, and add to the argument list a requirement for a form of Y realizing the same feature set except for the fact that it is a participle.

(36)
$$\begin{bmatrix}\text{PHON} & X\\ \text{LID} & Y\\ \text{VAL} & Z\end{bmatrix}, \sigma : \begin{bmatrix}\text{PRF} & +\end{bmatrix} \longrightarrow$$
$$\begin{bmatrix}\text{PHON} & \text{refer}\left(\begin{bmatrix}\text{PHON} & X\\ \text{LID} & \textit{budan}\\ \text{VAL} & Z\end{bmatrix}, \sigma \backslash \begin{bmatrix}\text{PRF} & -\\ \text{ASP} & \textit{bnd}\\ \text{POL} & +\end{bmatrix}, \text{I-V}\right)\\ \text{LID} & Y\\ \text{VAL} & Z \oplus \left\langle \begin{bmatrix}\text{LEX} & +\\ \text{VC} & +\\ \text{HEAD} & \begin{bmatrix}\text{LID} & Y\\ \text{MORSYN} & \sigma \backslash \begin{bmatrix}\text{FORM} & \textit{part}\end{bmatrix}\end{bmatrix}\end{bmatrix} \right\rangle\end{bmatrix}$$

The proposed analysis makes the following correct predictions. First, negation is handled correctly: the phonology of the head word is constrained to be that of a positive form of *budan*, whereas the participle shares its polarity value with that of the head word. Thus the head will never carry a negation prefix, but its negative polarity value will be realized as a prefix on the participle it selects. Second, the complex perfect is predicted to exist without stipulation: because evidentiality is morphologized and available for all past forms, rule (36) will generate an indirect past perfect with the phonology of an indirect bounded past form of *budan*. Figure 3 illustrates the relevant analysis. Third, the analysis correctly predicts that (33a), and not (33b), is grammatical. This is because the passive auxiliary, as a lexeme, can be put in the perfect; whereas there is no sense in which one can put the perfect auxiliary in the passive, because there is no such thing as a perfect auxiliary lexeme. The analysis of (33a) is shown in Figure 4.

Finally we account straightforwardly for the nonexistence of a periphrastic present perfect. Since (36) is an inflectional realization rule, it interacts with the rule of referral in (30) under the logic of rule specificity: thus the existence of (30) overrides the application of (36). In this sense the current analysis of the periphrastic perfect is syntactically reductionist: periphrasis is reduced to valence; no phrasal constructions or lexical entries are needed. We assume a notion of rule competition, but this competition is segregated to the inflectional component, where it is arguably needed for independent reasons. Thus no competition between morphology and syntax (e.g. Poser, 1992; Bresnan, 2001) needs to be orcherstrated.

4.3 Discussion

The analysis of the Persian perfect outlined above attempts to capture the traditional intuition of periphrastic inflection. While there are many ways one might

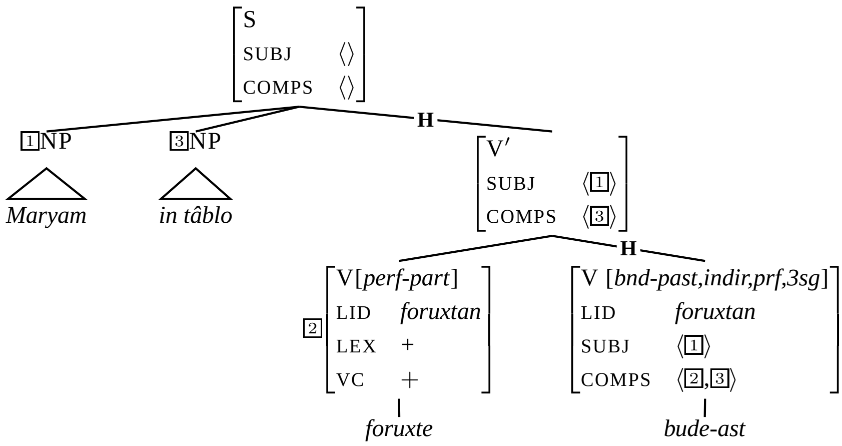


Figure 3: Analysis of a sentence in the complex pefect: 'Reportedly, Maryam had sold this painting.'

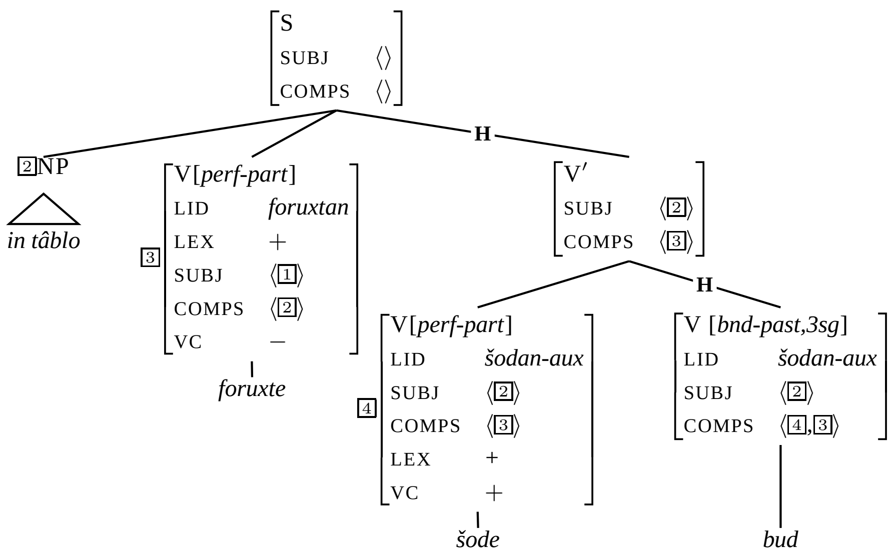


Figure 4: Perfect-passive interaction: analysis for (33a)

attempt to reach this goal in the context of HPSG (see in particular Ackerman and Webelhuth, 1998), the specific design goal here has been to devise an analysis that meets as much as possible both the analytical habits of HPSG syntax and of realizational morphology. Thus as far as clausal syntax is concerned, our analysis is undistinguishable from an argument composition analysis, and we have attempted to account for all relevant syntactic features of the construction. On the other hand, the lexical analysis is as close as possible to standard realizational morphology; in particular it relies heavily on the architecture of rule blocks and rule competition to generate the correct lexical representations.

While we fully assume this research strategy, alternatives are readily conceivable that meet different design goals but produce very similar analyses. For instance, turning the analysis into a standard HPSG analysis, with item-and-process morphology encoded via lexical rules, is easy: one just needs to recast rules such as (30) and (36) as lexical rules, and modify the morphosyntactic descriptions (using quite a bit of negation and disjunction and/or fine-tuning the type system) so as to make them mutually exclusive. The resulting system is more conservative from an HPSG perspective, although one may doubt that it is more perspicuous.

A different issue raised by the current analysis is its interaction with the analysis of coordination. Coordination of participles is possible in the perfect in Persian, just as it is in the passive (37). This can not be treated as a simple instance of constituent coordination under our analysis: because we assume that the auxiliary is really an inflected form of the main verb, there is no single lexeme of which *bud* is the realization in (37). While this is definitely a problem, it is a familiar one, reminiscent of issues pertaining to coordinations of unlikes. We see two potential solutions. First, we could assume an ellipsis-based analysis of (37) along the lines of analyses proposed by (Yatabe, 2001; Crysmann, 2003; Beavers and Sag, 2004). Second, we could assume a richer ontology of LID values where a neutralized value common to both participles is assigned to the coordinate phrase in (37), extending work in the tradition of (Daniels, 2002; Levy and Pollard, 2002; Sag, 2003). This neutralized value could then serve as an appropriate input for rule (36).[6] Whether these strategies prove fruitful will have to wait for future research, and in particular for a detailed empirical study of coordination in Persian.

(37)	Maryam	tâblo-râ	pasandide	va	xaride	bud.
	Maryam	painting-DDO	liked	and	bought	be.PST
	'Maryam had liked and bought the painting.'					

5 The future

For the periphrastic future, a number of different analytic options are available. As in the case of the periphrastic perfect, the verb sequence can not be interrupted, and

[6]Notice that the postulation of neutralized LID values is needed anyway to allow for constituent coordination under the assumptions of (Sag, 2007). Thus the issue raised by our analysis is an issue that needs to be addressed anyway.

occurs in a rigid order.

(38) a. Maryam Omid=râ xâh-ad did.
Maryam Omid=DDO want.S1-3.SG see.S2
'Maryam will see Omid.'

b. *Maryam xâh-ad Omid=râ did.
Maryam want.S1-3.SG Omid=DDO see.S2

c. *Maryam Omid-râ did xâh-ad.
Maryam Omid-DDO see.S2 want.S1-3.SG

The periphrastic future does not enter into paradigmatic relations with syncretic inflection. Thus it could be accounted for entirely within syntax. On the other hand, syntactic rules do not manipulate portions of the periphrastic construction—notably, the nonauxiliary part of the future can not be fronted. Thus nothing precludes either a purely morphological analysis.

There is however one argument favouring a purely morphological analysis, although it is not a very strong one. The future auxiliary looks like a present tense form of *xâstan* 'want', except that it does not carry the unbounded auxiliary normally found in the present. If we were to treat the future construction as phrasal, we would thus need to set up the grammar so that the morphology output supplementary forms, the distribution of which we would then need to constrain drastically within syntax.[7] We thus opt for a purely morphological analysis.[8] We propose to use the rule in (39), which is a double portmanteau rule of referral. To find the

[7]The nonfinite form appears to be a bare past stem. Words homophonous to a bare past stem are used in two other contexts: in the bounded past with a 3SG subject, where the exponent of agreement is null; and in the impersonal complement of some modal verbs such as *bâyastan* 'must, be necessary' (1).

(1) a. Maryam (hatman) bây-ad be madrasa be-rav-ad.
Maryam certainly must.S1-3SG to school IRR-go.S1-3SG
'Maryam definitely has to go to school.'

b. (Hatman) bây-ad be madrase raft.
certainly must.S1-3SG to school go.S2
'It is definitely necessary to go to school.'

[8]One could argue from the fact that object clitics can be realized either on the auxiliary (i) or on the nonfinite form (ii) that they should be treated as two distinct syntactic atoms; but since we treat object clitics as affixes anyway, the question is moot. In any case, the analysis in (39) can readily be extended to account for (i), but an account of (ii) will need to rely on more extensive revisions.

(i) Maryam xâh-ad did-aš
Maryam want.S1-3.SG see.S2-PAF.3.SG
'Maryam will see her/him.'

(ii) Maryam xâh-ad-aš did
Maryam want.S1-3.SG-PAF.3.SG see.S2
'Maryam will see her/him.'

phonology of a future form, one needs to concatenate the output of block IV on the form *xâh* with a bare past stem of the lexeme being realized.

(39)
$$\begin{bmatrix}\text{PHON} & X\\ \text{LID} & Y\\ \text{ARG-ST} & Z\end{bmatrix}, \sigma : \begin{bmatrix}\text{TNS} & \textit{fut}\end{bmatrix} \longrightarrow$$
$$\begin{bmatrix}\text{PHON} & \textsf{refer}\left(\begin{bmatrix}\text{PHON} & \textsf{xâh}\\ \text{LID} & Y\\ \text{ARG-ST} & Z\end{bmatrix}, \sigma, \text{V}\right) \oplus \\ & \textsf{refer}\left(\begin{bmatrix}\text{PHON} & X\\ \text{LID} & Y\\ \text{ARG-ST} & Z\end{bmatrix}, \sigma \backslash \begin{bmatrix}\text{TNS} & \textit{pst}\end{bmatrix}, \text{I}\right)\\ \text{LID} & Y\\ \text{ARG-ST} & Z\end{bmatrix}$$

6 The progressive

All unbounded forms may give rise to a progressive interpretation, but that interpretation can also be forced by using the periphrastic construction illustrated in (4). Unlike the ones we discussed so far, this construction results from the grammaticalization of a finite complement clause construction, and all relevant evidence points to the fact that an embedded clausal structure is still present.[9] The nonauxiliary verb is unmistakably a finite form; it occurs on the right of the auxiliary, as finite complement clauses occur on the right of their head. No complementizer can be used, but complementizers are optional for finite complements (40). Complements normally occur between the two verbs; they can scramble to the left of the auxiliary, but this is also possible with clausal complements (41). Finally, object clitic pronouns must be realized on the nonauxiliary verb, and cannot climb to the auxiliary (42).

(40) a. Maryam dâr-ad (*ke) ketâb mi-xân-ad
Maryam have.S1-3SG COMP book UNBD-read.S1-3SG
'Maryam is reading a book.'

b. Maryam mi-xâh-ad (ke) bâ Omid har ruz be sinemâ
Maryam UNBD-want.S1-3SG COMP with Omid every day to theatre
be-rav-ad
SBJV-go.S1-3SG
'Maryam wants to go to theatre with Omid everyday.'

[9]Persian raising and control constructions normally rely on a finite unsaturated complement clause. Infinitival complements are available only in a very formal register.

(41) a. Maryam in ketâb=râ dâr-ad mi-xân-ad
Maryam this book=DDO have.S1-3SG UNBD-read.S1-3SG
'Maryam is reading this book.'

b. Maryam bâ Omid mi-xâh-ad (ke) har ruz be sinemâ
Maryam with Omid UNBD-want.S1-3SG COMP every day to theatre
be-rav-ad
SBJV-go.S1-3SG
'Maryam wants to go to theatre with Omid everyday.'

(42) a. Maryam dâr-ad mi-xân-ad=aš
Maryam have.S1-3.SG UNBD-read.S1-3SG-=3SG
'Maryam is reading it.'

b. * Maryam dâr-ad=aš mi-xân-ad
Maryam have.S1-3SG=3SG UNBD-read.S1-3SG

This data can be accounted for by assuming a slightly idiosyncratic lexemic entry for the auxiliary *dâštan*. This entry assumes that *prog* is a subtype of the ASPECT value *unbd* (unbounded). As a result of its lexeme-level specification, this auxiliary is defective for all subparadigms except the present, the unbounded past and the complex unbounded past, in accordance with the facts. The subject of the complement is constrained to be an *nc-pro*, the type of pro-dropped subjects, and coindexed with the auxiliary's subject. The analysis is illustrated in Figure 5.

(43)
$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{LID} & \textit{dâštan-aux} \\ \text{MORSYN} & \boxed{1}\begin{bmatrix} \text{ASP} & \textit{prog} \end{bmatrix} \end{bmatrix} \\ \text{CONT} & \boxed{2} \\ \text{ARG-ST} & \left\langle \begin{bmatrix} \text{IND} & \boxed{3} \end{bmatrix}, \begin{bmatrix} \text{MORSYN} & \boxed{1} \\ \text{MARKING} & \textit{none} \\ \text{CONT} & \boxed{2} \\ \text{SUBJ} & \left\langle \begin{bmatrix} \textit{nc-pro} & \\ \text{IND} & \boxed{3} \end{bmatrix} \right\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \right\rangle \end{bmatrix}$$

References

Ackerman, Farrell and Stump, Gregory T. 2004. Paradigms and Periphrastic Expression. In Louisa Sadler and Andrew Spencer (eds.), *Projecting Morphology*, pages 111–157, Stanford: CSLI Publications.

Ackerman, Farrell and Webelhuth, Gert. 1998. *A Theory of Predicates*. Stanford: CSLI Publications.

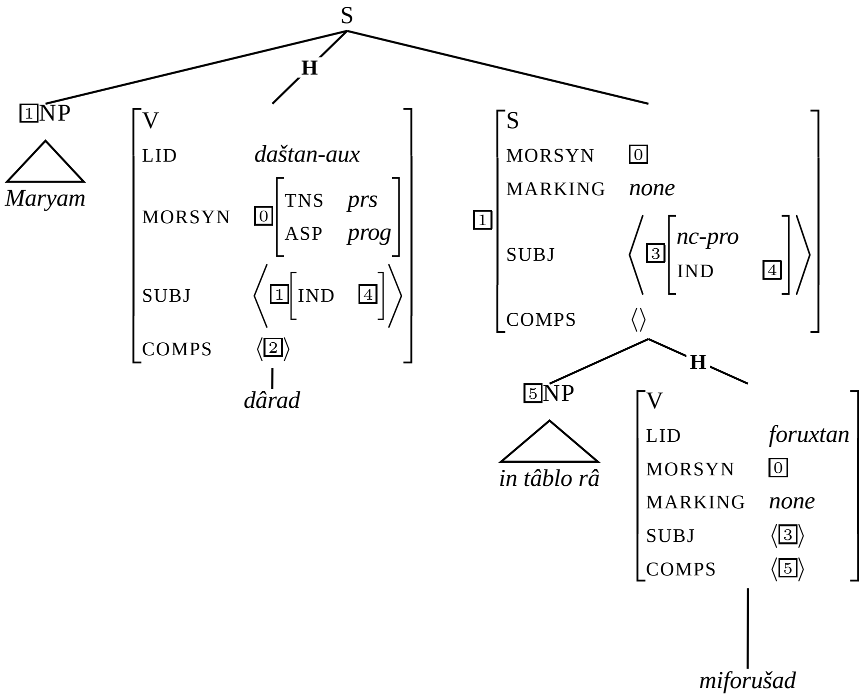


Figure 5: Analysis for (4) ‘Maryam is selling the painting.’

Beavers, John and Sag, Ivan A. 2004. Coordinate Ellipsis and Apparent Non-Constituent Coordination. In *Proceedings of the HPSG04 Conference*, Stanford: CSLI Publications.

Bresnan, Joan. 2001. *Lexical-functional syntax*. Oxford: Basil Blackwell.

Crysmann, Berthold. 2003. An Asymmetric Theory of Peripheral Sharing in HPSG: Conjunction Reduction and Coordination of Unlikes. In *Proceedings of Formal Grammar 2003*.

Daniels, Michael W. 2002. On a Type-Based Analysis of Feature Neutrality and the Coordination of Unlikes. In Frank Van Eynde, Lars Hellan and Dorothee Beermann (eds.), *Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar*, pages 137–147, Stanford: CSLI Publications.

Gazdar, Gerald, Klein, Ewan, Pullum, Geoffrey K. and Sag, Ivan A. 1985. *Generalized Phrase Structure Grammar*. Harvard: Harvard University Press.

Hinrichs, Erhard and Nakazawa, Tsuneko. 1994. Linearizing AUXs in German verbal complexes. In John Nerbonne, Klaus Netter and Carl Pollard (eds.), *German in HPSG*, pages 11–37, Stanford: CSLI Publications.

Jahani, Carina. 2000. Expressions of indirectivity in spoken Modern Persian. In Lars Johanson and Bo Utas (eds.), *Evidentials: Turkic, Iranian and Neighbouring Languages*, pages 185–207, Berlin: Mouton de Gruyter.

Lazard, Gilbert. 1985. L'inférentiel ou passé distancié en persan. *Studia Iranica* 14, 27–42.

Lazard, Gilbert, Richard, Yann, Hechmati, Rokhsareh and Samvelian, Pollet. 2006. *Grammaire du persan contemporain*. Tehran: IFRI and Farhag-e Moaser.

Levy, Roger and Pollard, Carl. 2002. Coordination and Neutralization in HPSG. In Frank Van Eynde, Lars Hellan and Dorothee Beermann (eds.), *Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar*, pages 221–230, Stanford: CSLI Publications.

Poser, William J. 1992. Blocking of phrasal constructions by lexical items. In Ivan A. Sag and Anna Szabolcsi (eds.), *Lexical matters*, pages 111–130, Stanford: CSLI Publications.

Sag, Ivan A. 2003. Coordination and underspecification. In *Proceedings of the 9th HPSG Conference*, pages 267–291, Stanford: CSLI Publications.

Sag, Ivan A. 2007. SIgn-Based Construction Grammar: An Informal Synopsis, ms., Stanford University.

Spencer, Andrew. 2005. Inflecting clitics in Generalized Paradigm Function Morphology. *Lingue e Linguaggio* 4, 179–193.

Stump, Gregory T. 2001. *Inflectional Morphology. A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

Windfuhr, Gernot. 1982. The verbal category of inference in Persian. In *Momentum Georg Morgenstierne II*, pages 263–287, Leiden: E. G. Brill.

Yatabe, Shûichi. 2001. The Syntax and Semantics of Left-Node Raising in Japanese. In Dan Flickinger and Andreas Kathol (eds.), *Proceedings of the 2001 HPSG Conference*, Stanford: CSLI Publications.

Construction-based Cumulation and Adjunct Extraction

Rui P. Chaves

University at Buffalo – SUNY

Proceedings of the HPSG09 Conference

Göttingen, Germany

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

Previous HPSG accounts of extraction blur the distinction between valents and adjuncts by allowing verbs to lexically control the modifiers that combine with their phrasal projections. However, assuming that adjuncts are valents runs into various difficulties. This paper argues that the distinction between complements and adjuncts can be maintained, and that certain semantic phenomena that challenge traceless theories of extraction can be seen as an instance of a more general process. Finally, this paper also discusses a uniform mechanism for case assignment to valents and adverbial nominals.

1 Introduction

Pollard and Sag (1994) and others have noted that certain verbal adjuncts can be extracted, as in (1). Although extractable, these phrases behave like adjuncts in many other aspects (they are not semantic arguments of the verb that they modify, are optional, can be iterated, are canonically VP-final, pass the *do-so* test, and have a freer distribution than true arguments).

(1) a. [Yesterday], it seems that [Kim arrived home very early _].
b. [(On) that day], I think that [Kim went home very late _].
c. [How often] do you think that [Fred was late this week _]?
d. It was [with a stick] that [we killed the snake _].

The distinction between adjuncts and complements also appears to be blurred cross-linguistically, in extraction pathway marking phenomena (see Clements et al. (1983); McCloskey (1979); Hukari and Levine (1995) *inter alia*), in case assignment to adverbial nominals (see Maling (1989, 1993) *inter alia*), and in adverbial scope (van Noord and Bouma, 1994). Thus, it can be argued that at least some verbal adjuncts are selected or controlled by the verbal head. This poses various puzzles, and runs counter the standard distinction between arguments and adjuncts. §2 discusses previous accounts of these phenomena, and §3 proposes a new analysis that allows for a simpler view of the adverbial argument-adjunct puzzle.

2 Previous accounts

There are two main approaches to adjunct extraction that have been proposed within HPSG. One is lexical (the lexical entry of the verb hosts adverbial gaps in SLASH), and another is phrasal (adverbial gaps are introduced syntactically). Both run into problems in the presence of conjunction, as discussed below.

[†]I thank the audience of the HPSG09 conference and reviewers for their comments and criticism. A very special dept is owed to Doug Arnold. I am also grateful to Olivier Bonami, Philip Hofmeister, EunHee Lee, Stefan Müller, and Ivan Sag. I am the sole responsible for any errors or omissions.

2.1 Lexical gap addition

Pollard and Sag (1994, 387) propose the lexical rule in (2), which adds one adjunct to the SLASH feature of verbs that subcategorize for clauses.

(2)

$$\mathrm{V}\begin{bmatrix}\textsc{comps}\ \langle \dots\ \mathrm{S}\dots \rangle\\ \textsc{slash}\ \{\}\end{bmatrix} \rightarrow \mathrm{V}\begin{bmatrix} \textsc{comps}\ \langle\dots \boxed{2}\mathrm{S}\dots\ \rangle \\ \textsc{slash}\ \left\{ \mathrm{XP}\begin{bmatrix}\textsc{mod}\ \boxed{2}\\ \textsc{index}\ \boxed{3}\end{bmatrix}\right\} \\ \textsc{cont}\mid\textsc{soa-arg}\ \boxed{3}\end{bmatrix}$$

But as Hukari and Levine (1996) note, this account is problematic on various fronts. Not only adverbs can be extracted when verbs take VP complements, but there are a number of languages for which valent and modifier extraction triggers exactly the same morphophonological processes. This is the case of Kikuyu downstep suppression (Clements et al., 1983) and Irish complementizers (McCloskey, 1979), among many others. Since extraction pathway marking for valents and (some) modifiers triggers the same phenomena in these languages, van Noord and Bouma (1994) and Bouma et al. (2001) propose that such modifiers are in fact lexically selected by the verbal head, as dependents. Thus, adjunct extraction can be handled in the same way as valent extraction. The account in Bouma *et al.* is given in (3).

(3) ARGUMENT STRUCTURE EXTENSION

$$\mathit{verb} \Rightarrow \begin{bmatrix}\textsc{synsem}\begin{bmatrix}\textsc{head}\ \boxed{1}\\ \textsc{arg-st}\ \boxed{2}\\ \textsc{deps}\ \boxed{2} \oplus \mathit{list}\left(\begin{bmatrix}\textsc{mod}\begin{bmatrix}\textsc{head}\ \boxed{1}\\ \textsc{cont}\mid\textsc{key}\ \boxed{3}\end{bmatrix}\end{bmatrix}\right)\\ \textsc{cont}\mid\textsc{key}\ \boxed{3}\end{bmatrix}\end{bmatrix}$$

However, giving up adjunction is problematic, as Levine (2003) notes. First, further assumptions are needed in order to account for cases that would otherwise be taken care as standard VP adjunction, as in (4).

(4) Nobody can [[drink four beers and eat two hotdogs] [under fifteen seconds]].

If the PP is a complement, then something else must be assumed in order to capture this sentence. For example, one would have to assume that the PP is extraposed ATB or Right-Node Raised. But, as Levine and Hukari (2006) argue, these hypotheses are at odds with the semantic interpretation that the PP obtains, which ranges over the total time interval denoted by the two conjuncts. In contrast, this reading is trivially obtained if the PP simply adjoins to the conjoined VP.

It should be clear that plurality-forming conjunction operates beyond NPs, and can form *event pluralities* (Bach, 1986; Lasersohn, 1995; Link, 1998). For example, the sentence in (5a), adapted from Oehrle (1987), can describe the frequency of two joint event-types rather than independent frequencies of occurrence.

(5) a. Often, [[I go to the beach]$_{e_1}$ and [you go to the city]$_{e_2}$]$_{e_1+e_2}$.

b. Sue [[[got dressed]$_{e_1}$ and [dried her hair]$_{e_2}$]$_{e_1+e_2}$, [in exactly twenty seconds]].

c. You can't simultaneously [[drink]$_{e_1}$ and [drive]$_{e_2}$]$_{e_1+e_2}$.

Levine (2003) also points out that the cumulative reading can occur even when the adjunct is extracted, as in (6). Here, *In how many seconds flat* predicates the total of three events denoted by the embedded coordinate VP, not each conjunct.

(6) In how many seconds flat do you think that [Robin found a chair, sat down and took off her logging boots]?

This utterance is a query about the total time occupied by the occurrence of three (possibly overlapping) events. Such a reading suggests that the extracted constituent is not a complement of anything in the sentence. If it were, then the adjunct should be predicating over each of the conjuncts separately, not the higher VP coordination node. To address this problem, Sag (2005) proposes that the extracted phrase *in how many seconds flat* is a complement that can semantically outscope the verb structure that it modifies. In a coordinate structure, the PP extracted ATB is naturally required to simultaneously outscope each of the verbs heading the conjuncts, thus obtaining wide scope over the entire coordination.

However, there is no evidence that the modifier *in X seconds* is semantically scope-bearing. Compare the unambiguous examples in (7) with the examples in (8). Only the latter contain scope-bearing modifiers and trigger an ambiguity with respect to the wide or narrow scope interpretation of the indefinite NP.

(7) a. Someone died in the arena yesterday / under twelve seconds flat.

b. Kim sang a song yesterday / in twelve seconds flat.

(8) a. Someone probably / usually died in the arena.

b. Kim probably / usually / often sang a song.

Scope cannot in general solve the cumulation problem, but in §3 I will argue that the challenging phenomenon in (6) is the consequence of other semantic aspects.

2.2 Syntactic gap addition

Assuming that adjuncts are modifiers, Levine (2003) proposes that extraction paths are terminated by traces. Thus, a modifier can instantiate the trace in (9) and adjoin to VP as usual. When it does, it creates an unbounded dependency that is percolated and linked to a filler, like any other unbounded dependency. Cumulative readings arise as a consequence of adjuncts being able to adjoin to VPs, coordinate or not.

(9) $\begin{bmatrix} \text{PHON}\ \langle\rangle \\ \text{SYNSEM} \begin{bmatrix} \text{LOC}\ \boxed{1} \\ \text{SLASH} \left\{\boxed{1}\right\} \end{bmatrix} \end{bmatrix}$

In order to deal with adverbial case assignment and extraction pathway marking, Levine and Hukari (2006) introduce a new feature ADJS, which allows the lexical entry of a verb to list adjuncts realized in arbitrarily high positions. This list is lexically underspecified, and becomes instantiated at the phrasal level, when modifiers adjoin to a VP projected by that verb. The verb can thus lexically access any gaps that may reside in an adjunct located in a higher position in the syntactic tree (see also Sato and Tam (2008)). With regard to scope ambiguities in Dutch that have been argued to follow from an adjuncts-as-complements analysis, Levine (2003) proposes that these can be captured with direct access to the relevant parts of semantic representation.[1]

There are however some concerns with the adjunct extraction account in Levine (2003). First, nothing prevents the adverb trace in (9) from being adjoined to each of the VP conjuncts, instead of the coordinate mother VP. Given the standard assumption that SLASH values are structure-shared between daughters and mother in coordination, then one would obtain an impossible interpretation where each conjunct event is the same. This is shown schematically in Figure 1; notice that both VPs become structure-shared in $\boxed{1}$). The problem is related to the one discussed in Levine and Hukari (2006, 159), where structure-sharing slashed event-modifying adjuncts originating in each conjunct yield a description that no sign can satisfy.

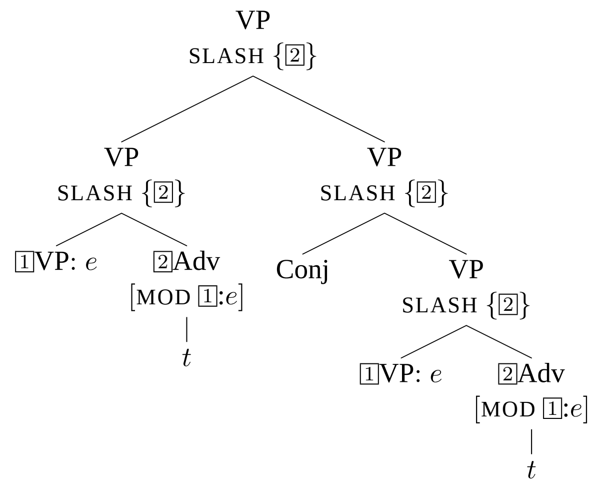


Figure 1: ATB adverbial extraction and impossible descriptions

[1]This can be achieved in a number of different ways, for example, if KEY corresponds to the predication $\phi(v_1, ..., v_n)$ semantically heading a phrase, then other heads can access the predicate's argument slots. Thus, an adverb like *often* can either predicate the verb heading the phrase that it adjoins to, or the scopal argument of that verb. See also Crysmann (2004) and Sato and Tam (2008).

This is not the same as a distributive interpretation of the adverb. For example, *yesterday* is distributive in the sense that when applying to a sum of events such as e_1+e_2, it does not yield a collective predication *yesterday*(e_1+e_2), but rather, a distributive one: $\forall e(e \leq e_1+e_2 \rightarrow$ *yesterday*$(e))$. In the latter, the adverb predicates over each mereological part e of the event sum e_1+e_2. A modifier like *for how long* on the other hand, can apply collectively to the entire sum. Which adverbs are distributive, which are collective, and which are ambiguous is a matter of lexical specification, similar to how verbs like *smile*, *meet*, and *hire* can interpret their plural NP arguments in different ways. This is what seems to be happening in the ambiguous example from Levine and Hukari (2006,186), shown below. Although the PP attaches to the higher VP coordination, it can either apply distributively to each event in the sum e_1+e_2, or apply collectively to the entire sum.

(10) Robin [stands on his head and falls off his chair]$_{e_1+e_2}$ in order to attract people's attention.

A second concern pertains to the assumption that extraction is terminated by traces. I will side with Sag and Fodor (1994) and Sag (2000) in assuming that extraction can be modelled without resorting to traces. The question is, of course, how can this be achieved parcimoniously. One possible alternative is explored in Müller (1999, 108–109,447) and Chaves (2007,Ch.7), who show that it is possible to formulate unary-branching traceless extraction rules. Consider (11), based on Chaves (2007), which allows an adjunct to become a member of the head's SLASH.

(11) ADJUNCT EXTRACTION RULE:

$$\textit{adj-extr-phr} \Rightarrow \begin{bmatrix} \text{SYNSEM} \begin{bmatrix} \text{LOC } \boxed{1} \\ \text{SLASH } \boxed{2} \cup \left\{ \text{XP}\begin{bmatrix}\text{MOD } \boxed{3}\end{bmatrix} \right\} \end{bmatrix} \\ \text{HD-DTR } \boxed{4} \\ \text{DTRS } \left\langle \boxed{4} \begin{bmatrix} \text{SYNSEM } \boxed{3} \begin{bmatrix} \text{LOC } \boxed{1} \\ \text{SLASH } \boxed{2} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

Assuming a Ginzburg and Sag (2000) framework, the GENERALIZED HEAD FEATURE PRINCIPLE would ensure that valence and head features are percolated in the tree structure. The value of SLASH is percolated as dictated by the rule in (11), overriding the default percolation of the GENERALIZED HEAD FEATURE PRINCIPLE. Since (11) is independent from coordination, and adverbs are lexically select VPs, gap insertion can also arise in non-coordinate VPs. In other languages adverbs may select S nodes or even to V nodes instead.

This analysis can also be augmented with the ADJS feature, so that whenever the rule in (11) applies, the gap adjunct is identified with a member of the head's ADJS list. This way, a verb can access adjuncts higher up in the tree, for gap threading purposes. The account is illustrated in Figure 2. The VP daughter is modified by an adverbial phrase located in SLASH. Thus, *yesterday* modifies the event plurality that the VP conjunction yields.

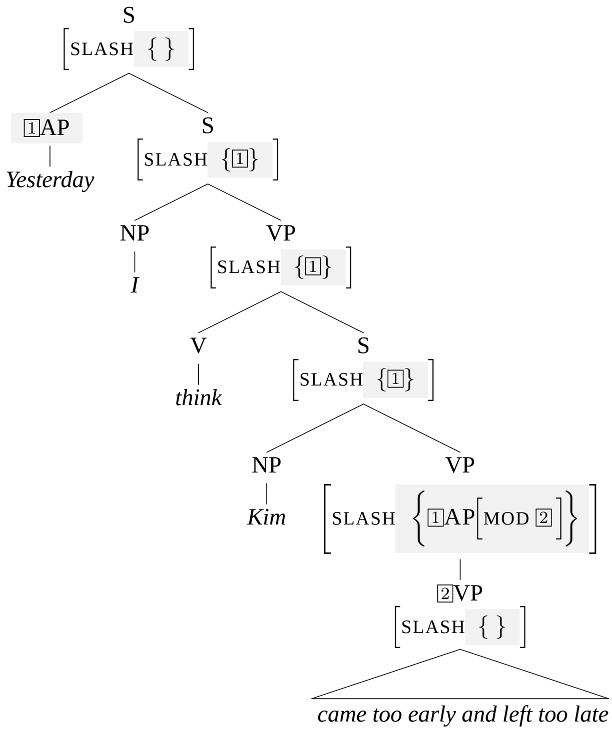


Figure 2: A modifier adjoining into SLASH

The ATB adjunct extraction in (12) can also be easily addressed. In Copestake et al. (2006), existentially quantification of events is often omitted for simplification purposes, but if one makes such quantification explicit – as in $\exists e$ ***left***(e, kim) – then the adjunct extraction rule cannot apply to each VP conjunct because the resulting semantic structure is ill-formed. Consider the parse in (12).

(12) [Under how many seconds flat] did Kim [pack _ and escape _]?

Each adjunct is located in the SLASH value of each VP conjunct, and predicates over the respective event. The modifiers must be one and the same at the coordination level, because the coordination rule imposes identity of SYN values (as for example, in Beavers and Sag (2004)). Thus, the adverbial phrase filler has to predicate the very same event across conjuncts, and has to be simultaneously located under the scope of each existential quantifier. This yields an ill-formed MRS structure because the underspecified representation cannot describe a tree structure, as depicted in Figure 3. Arrows denote underspecified semantic subordination constraints. Here, $\boxed{k}$ is the shared subject index of *Kim*, and $\boxed{e'}$ = $\boxed{e1}$ = $\boxed{e2}$. This solution follows from the coordination rule, and is valid for extraction accounts with or without traces.

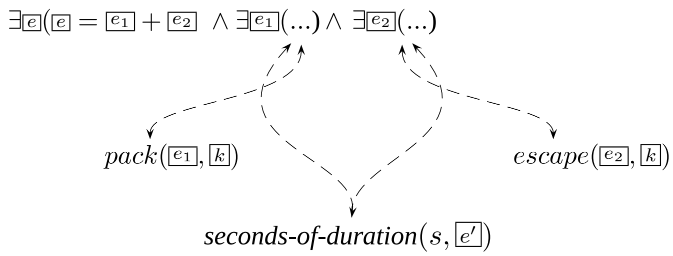


Figure 3: ATB adjunct extraction yielding an illegal MRS representation

2.3 Taking stock

All of the extraction accounts discussed so far – the adjuncts as complements analysis and the phrasal adjunction analyses – allow verbs to select for *in situ* adjuncts, blurring the distinction between adjuncts and valents. In one case this is done lexically, and in the other, via a special valence feature ADJS, so that gap threading phenomena can be dealt with.

However, there is in my view no semantic evidence that such adjuncts are complements. In fact, the semantic evidence observed in coordination indicates that adjuncts semantically combine with phrase structures. On the other hand, the feature ADJS seems to lack independent motivation because it is only relevant for a particular class of post-verbal adjuncts, and not, for example, adnominal modifiers.

In what follows I propose a simpler and more general analysis that maintains a strong distinction between adjuncts and valents, and dispenses the need for traces, new constructions, and ADJS. As in Pollard and Sag (1994), adjunct gaps start out lexically so that heads can only detect extracted adjuncts, not *in situ* ones, and cumulative phenomena are handled semantically in ways reminiscent of Sag (2005). By viewing the cumulative readings of extracted elements as an instance of a more general phenomenon, we will also be able to deal with other cases that arise beyond adjunct extraction.

3 A coordination-based proposal

Let us assume that there are no traces, and that adjuncts are not valents but rather modifiers in the usual sense. It could be that the cumulation of extracted adverbials is an instance of a more general phenomenon where the plurality-formation operation triggered by conjunction 'bleeds over' to certain unrealized dependents. Consider the data in (13), from Postal (1998, 136,160) and Kehler (2002, 125). Here, an extracted NP can denote a plurality composed of two individuals, each being linked to each verb in each conjunct. These data are relevant because each conjunct contains a different gap, and the two gaps are not fused together as a single entity. Rather, they might be cumulated into a complex entity (a conjoined NP) and permitted to percolate independently.

(13) a. [How many frogs]$_i$ and [how many toads]$_j$ did respectively Greg capture $\_{i}$ and Lucille train $\_{j}$?

b. [[Which pilot]$_i$ and [which sailor]$_j$] will Joan invite $\_{i}$ and Greta entertain $\_{j}$ (respectively)?

c. [[What book]$_i$ and [what magazine]$_j$] did John buy $\_{i}$ and Bill read $\_{j}$ respectively?

This is somewhat unexpected, because coordination is known to not allow different conjuncts to host different gaps, as shown in (14).

(14) *[A violin this well crafted]$_i$, even [the most difficult sonata]$_j$ will be easy to write $\_{j}$ and to play it on $\_{i}$.

One can argue that in (13) there is only one gap at the coordination level, and that this gap is linked to a pluralic filler. Each member of the plurality is predicated by a different VP conjunct. In (14) however, there are two fillers and thus each VP conjunct would have to contain a different gap. The latter is correctly ruled out if one assumes that SLASH values of conjuncts and mother node are structure-shared. Thus, it seems that as conjunction forms a plurality from the indices of the conjoined heads, the extracted dependents can be pluralized in a similar fashion. The cumulation of gaps is illustrated in Figure 4.[2]

S[SLASH {[INDEX $x + y$]}]

S[SLASH {[INDEX x]}]   S[SLASH {[INDEX y]}]

Figure 4: Conjunct nominal gap sharing with cumulation

If this analysis is on the right track, then one would expect to find the same phenomenon in constructions without *respectively*. However, the detection of such data is not easy because the gaps in such examples are preferentially interpreted non-cumulatively. In (15) I provide such data. These sentences can be interpreted as conveying that the plural NP filler corresponds to the union X ∪ Y, where X and Y are the entities extracted from different conjuncts.

(15) a. Setting aside illegal poaching for a moment, how many sharks$_{X+Y}$ do you estimate [[$\_{X}$ died naturally] and [$\_{Y}$ were killed recreationally]]?

b. The [ships$_{x+y}$ that [[a U-boat destroyed $\_{x}$] and [a kamikaze blew up $\_{y}$]]]] were $\left\{\begin{array}{l}\text{not insured}\\ \text{the Laconia and the Callaghan}\end{array}\right\}$.

[2]Chaves (2009) proposes a direct account of *respectively* readings that does not resort to any form of ‘conjunction reduction’, syntactically or semantically. Further research is forthcoming.

c. The houses$_{X+Y}$ [[the fire reduced to ash$_{}_X$] and [the flood leveled down $\_{}_Y$]] were near each other.

Now, the adjunct cumulation cases in Levine (2003) might be due to the same kind of phenomenon. This is illustrated in Figure 5. Each conjunct has a different adverbial gap, the index of each adverb matches the event that it modifies, and conjunction allows the indices of the adverbial gaps to be cumulated.

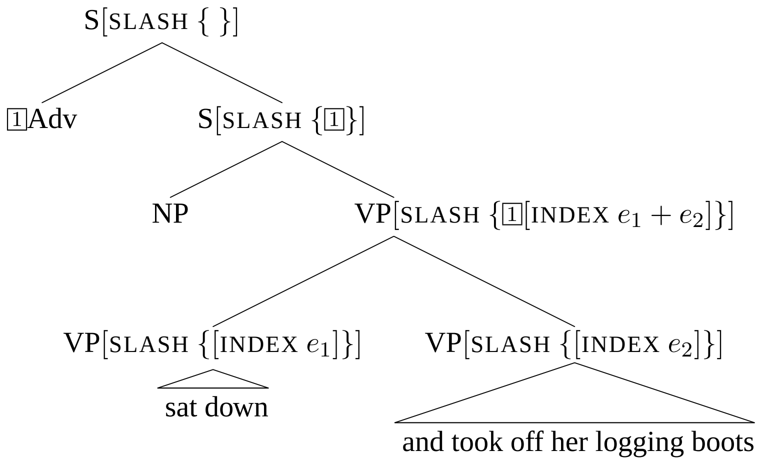


Figure 5: Conjunct adverbial gap sharing with cumulation

Put more in more general terms, in a coordinate structure with n displaced structures with indices $\alpha_1, ..., \alpha_n$, these can either be combined into one and the same entity $\alpha_1 = ... = \alpha_n$ or combined cumulatively into a complex entity, for example, a Linkean sum: $\alpha_1 + ... + \alpha_n$. In other words, the plurality-formation bleeds over to certain dependents. Because this mechanism is observed beyond adjunct extraction, it begs for a general account.

In this view where shared gaps in coordination can be cumulated, adjunction can operate as usual, and the only adjuncts that verbs need to have control over are the ones in SLASH, because of extraction pathway marking phenomena.

3.1 Other possible cases of cumulation in conjunction

Perhaps the cumulation phenomena observed above also arises in other kinds of dependents. For example, Vergnaud (1974), Abbott (1976), Jackendoff (1977) and others have noted a phenomenon where the same structure is cumulatively connected to different clauses. For example, in (16a) the plural NP *very different opponents* denotes a set of individuals some of which were defeated by John and some of which beat Mary. Similarly, in (16b) we do not know how much Fred spent nor how much Mia lost, although we know the total amount that Fred spent and Mia lost. In other words, these sentences are not equivalent to their counterparts in which the 'right node raised' NP is *in situ* after the first verb.[3]

(16) a. John DEFEATED and Mary LOST TO very different opponents.

b. Fred SPENT and Mia LOST a total of $10.000.

This process is sometimes not possible or highly marked. For example, reflexive expressions cannot be pluralized in this fashion:

(17) *John LOVES and Mary HATES themselves / each other.

This process is unique to plurality-forming conjunction. As Beavers and Sag (2004,66) note, disjunction does not allow cumulative readings:

(18) a. *John DEFEATED $_i$ or Mary LOST TO $_j$ [very different opponents]$_{i+j}$.

b. *Either Fred SPENT $_i$ or Mia LOST $_j$ [a total of $10.000]$_{i+j}$.

Yatabe (2003) argues that there are two kinds of RNR. One is prosodic and has no semantic effect, and another which is linearization-based and has a semantic effect. This distinction may be responsible for the contrast shown in (19) and (20). Prosodic RNR can be long-distance and can apply in non-coordinate contexts (as in (19a) and (20a); see Chaves and Sag (2007) for a recent discussion), whereas linearization RNR cannot be long-distance as shown in (19b) and (20b), and is restricted to conjunction.

(19) a. One man said that he LIKED _$_i$ and another even boasted that he ADORED _$_i$ [the woman in the commercial]$_i$.

b. ?*One man said that he LIKED _$_i$, and another even boasted that he ADORED _$_j$ [different women]$_{i+j}$.

(20) a. One man said that he HATED _$_i$ just because some other had boasted that he ADORED _$_i$ [the woman in the commercial]$_i$.

b. ?*One man said that he HATED _$_i$ just because some other had boasted that he ADORED _$_i$ [different women]$_{i+j}$.

Finally, it could be that the same cumulation phenomenon also occurs in extraposition phenomena, in the form of split antecedent relative clauses. Consider the data in (21), based in Ross and Perlmutter (1970). The relative clause is semantically linked to both subject NPs. It is unlikely that these are instances of RNR of the extraposed clause (e.g. [S ~~RelC~~] & [S RelC]) because no conjunct-final prosodic contrast is needed for examples like (21a), and because reflexives can be linked to the pluralized noun, as shown in (21b) (confront with (16)).

[3]Postal (1998) argues that cases like ?*the pilot claimed that the first nurse and the sailor proved that the second nurse were spies* also exhibit some form of summation/cumulation of the 'right-node raised' verbal structure. However, judgments are gradient and vary significantly, which lead Beavers and Sag (2004) to argue that these are quasi-sentences that result from performance effects. My account could in principle be extended to these data, along the lines pioneered by Yatabe (2002).

(21) a. A man entered the room and a woman left who were similar.

b. A man came into the store and a woman left right after him who used be in love with each other.

c.*A man entered the room or a woman left who were similar.

These cases can be accounted as follows. The nominal head that each extraposed relative clause is modifying is cumulated at the coordination level. One way to achieve this result is illustrated in Figure 6. Alternatively, one could also state this analysis in a more elegant way in terms of anchors (Kiss, 2005; Crysmann, 2004).

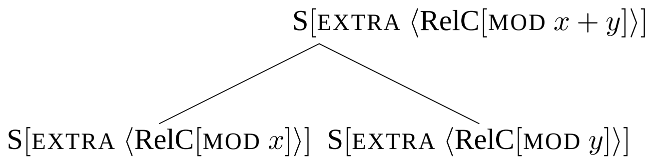


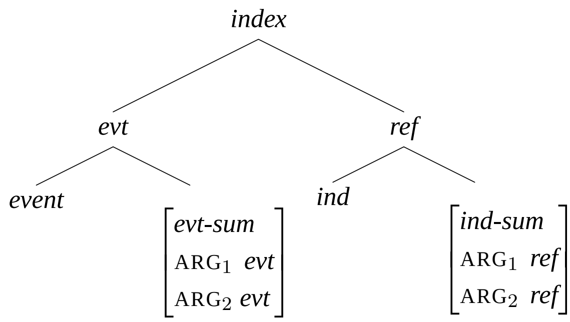
Figure 6: Conjunct relativized head sharing with cumulation

The remainder of the paper will flesh out an account of extraction that is compatible with gap cumulation, as well as with the cumulation of other delayed/displaced dependents such as RNR phrases and extraposed relative clauses.

3.2 HPSG formalization

The semantics of conjunction is a complex topic which I cannot address here fully, and so I will make the minimal assumptions needed for the purpose of this paper.[4] I start by allowing individual and event indices to be either Linkean sums '+' or atomic elements. The type hierarchy and signature are provided in (22).

(22)



Next, I define a non-deterministic relation '$\approx$' called *Integration* that maps a pair of lists onto one list. It allows for two different cases: in (23a) the arguments of the relation are simply structure-shared, and in (23b) we obtain the cumulation of two signs, by forming a new sign that denotes a plurality.

[4]For a more comprehensive semantic account of conjunction see Chaves (2007, Ch.5).

(23) a. Direct Integration (structure-sharing equality)

$(\boxed{1} \approx \boxed{1}) := \boxed{1}_{list}$

b. Cumulative Integration (summation):

$$\left(\left\langle\begin{bmatrix}\textsc{syn}\,|\,\textsc{loc} & \boxed{0}\\ \textsc{sem}\,|\,\textsc{index} & \boxed{1}\end{bmatrix}\right\rangle \approx \left\langle\begin{bmatrix}\textsc{syn}\,|\,\textsc{loc} & \boxed{0}\\ \textsc{sem}\,|\,\textsc{index} & \boxed{2}\end{bmatrix}\right\rangle\right) := \left\langle\begin{bmatrix}\textsc{syn}\,|\,\textsc{loc} & \boxed{0}\\ \textsc{sem} & \begin{bmatrix}\textsc{index} & \begin{bmatrix}\textsc{arg}_1 & \boxed{1}\\ \textsc{arg}_2 & \boxed{2}\end{bmatrix}\end{bmatrix}\end{bmatrix}\right\rangle$$

If we adopt the extraction account Ginzburg and Sag (2000), then we need only extend the conjunction with gap Integration. Since adjunct gaps reside in SLASH we can dispense ADJS and obtain the cumulation phenomena straight away. The rule is given below, in (24). For reasons of uniformity to be discussed below, the set-valued SLASH is replaced by a list-valued GAP feature, and I will use the feature geometry of Sag (2001). I also omit the full specification of KEY due to lack of space, but I am assuming basically the same as Bouma et al. (2001).

(24) CONJUNCTION CONSTRUCTION

$$\textit{conj-cx} \Rightarrow \begin{bmatrix}\textsc{mother} & \begin{bmatrix}\textsc{syn}\,|\,\textsc{gap} & \boxed{3} \approx \boxed{4}\\ \textsc{sem} & \begin{bmatrix}\textsc{index} & \begin{bmatrix}\textsc{arg}_1 & \boxed{1}\\ \textsc{arg}_2 & \boxed{2}\end{bmatrix}\end{bmatrix}\end{bmatrix}\\ \textsc{dtrs} & \left\langle\begin{bmatrix}\textsc{syn}\,|\,\textsc{gap} & \boxed{3}\\ \textsc{sem}\,|\,\textsc{index} & \boxed{1}\end{bmatrix}, \begin{bmatrix}\textsc{syn}\,|\,\textsc{gap} & \boxed{4}\\ \textsc{sem}\,|\,\textsc{index} & \boxed{2}\\ \textsc{crd-mrk} & \textit{conj}\end{bmatrix}\right\rangle\end{bmatrix}$$

I assume that the coordination construction is more general than the conjunction construction: the former requires LOCAL identity and has nothing to say about semantics, whereas the latter requires a right-marked conjunct with *and*, and yields a pluralic index from the indices of the conjuncts. Based on Beavers and Sag (2004), I assume coordination is binary branching and resorts to a feature CRD-MRK that identifies the coordination type (i.e. *conj*, *disj*, etc.).

We can, however, revise Ginzburg and Sag (2000) so that adjuncts are not complements. First, the SLASH-AMALGAMATION CONSTRAINT and the ARGUMENT STRUCTURE EXTENSION (see §2.1 above and Ginzburg and Sag 2000,169) are blended into a unique condition, in (25). Further constraints should pinpoint the exact class of extractable adjuncts. Then it will follow that $\boxed{h}$ has to be *verb* if the list of adjunct gaps is non-empty, ruling out adnominal adjunct extraction.

(25) Gap Amalgamation and Extension Condition

$$
word \Rightarrow \left[\begin{array}{ll}
\text{SYN} & \left[\begin{array}{l}
\text{LOC} \mid \text{HEAD } \boxed{h} \\
\text{GAP } \bigcup_{\oplus}(\boxed{1}\oplus\ldots\oplus\boxed{n})\ominus\boxed{0}\oplus \mathit{list}\left(\left[\text{MOD}\left\langle\begin{bmatrix}\text{SYN} \mid \text{L} \mid \text{HEAD } \boxed{h} \\ \text{SEM} \mid \text{KEY } \boxed{k}\end{bmatrix}\right\rangle\right]\right) \\
\text{BIND } \boxed{0}
\end{array}\right] \\
\text{SEM} \mid \text{KEY } \boxed{k} & \\
\text{ARG-ST} & \left\langle \left[\text{SYN} \mid \text{GAP } \boxed{1}\right], \ldots, \left[\text{SYN} \mid \text{GAP } \boxed{n}\right]\right\rangle
\end{array}\right]
$$

The relation $\bigcup_{\oplus}$ allows the amalgamation of gaps: $\bigcup_{\oplus}(\boxed{1}\bigcirc\boxed{2}\bigcirc\boxed{1}) = \boxed{1}\oplus\boxed{2}$. This operation takes as input a list, it splits the list nondeterministically into three sublists (two of which are structure-shared as $\boxed{1}$) and yields the append of the two remaining distinct lists. This allows some, all, or none of the gaps to be unified.

Non-scope bearing verb-modifying adjuncts possess an event index which is structure-shared with the event index of the verb they intersectively combine with. This is illustrated in (26a). Conjunction then is able to combine two such adjunct gaps into one gap with a summed event index. The cumulation of an ATB extracted PP adjunct is shown in Figure 7. I assume that the MRS representation of verbs includes an existentially quantified index, as discussed in §2.2.

This cumulation process occurs cross-categorially. With nominal gaps we obtain a sum of individual indices and with non-scope bearing verbal modifier gaps we obtain an event sum. As for extractable scope-bearing adverbs like *often*, I assume along with Sag (2005) that the adverb lexically outscopes the verb it modifies, as illustrated in (26b). Thus, in ATB extraction each verb heading each conjunct must be outscoped by the adjunct gap. See Bonami and Godard (2007) for other issues pertaining to scope surface order and scope ambiguities, as well as Crysmann (2004) for a more elaborate MRS account of intersective modification in German.

(26) a.
$$
\left[\begin{array}{ll}
\text{PHON} & \langle in \rangle \\
\text{SYN} & \left[\text{LOC}\left[\begin{array}{l}
\text{HEAD } \mathit{prep} \\
\text{MOD}\left\langle \text{VP}\begin{bmatrix}\text{INDEX } \boxed{e} \\ \text{LBL } \boxed{l}\end{bmatrix}\right\rangle \\
\text{SUBJ } \langle\ \rangle \\
\text{COMPS } \langle \text{NP}[\text{INDEX } \boxed{x}]\rangle
\end{array}\right]\right] \\
\text{SEM} & \left[\begin{array}{ll}
\text{INDEX } \boxed{e} & \\
\text{RELS} & \left\langle\begin{bmatrix}\text{LBL } \boxed{l} \\ \text{RELN } \mathit{in} \\ \text{ARG}_1 \ \boxed{e} \\ \text{ARG}_2 \ \boxed{x}\end{bmatrix}\right\rangle
\end{array}\right]
\end{array}\right]
$$

b.
$$
\left[\begin{array}{ll}
\text{PHON} & \langle often \rangle \\
\text{SYN} & \left[\text{LOC}\begin{bmatrix}\text{HEAD } \mathit{adv} \\ \text{MOD}\left\langle \text{VP}[\text{LBL } \boxed{l_1}]\right\rangle\end{bmatrix}\right] \\
\text{SEM} & \left[\begin{array}{ll}
\text{INDEX } \mathit{none} & \\
\text{RELS} & \left\langle\begin{bmatrix}\text{RELN } \mathit{often} \\ \text{ARG } \boxed{l_2}\end{bmatrix}\right\rangle \\
\text{HCONS} & \left\langle\begin{bmatrix}\mathit{qeq} \\ \text{HARG } \boxed{l_2} \\ \text{LARG } \boxed{l_1}\end{bmatrix}\right\rangle
\end{array}\right]
\end{array}\right]
$$

We can also extend the conjunction rule as in (27), so that split antecedent relative clauses and cumulation in linearization-RNR are captured. I follow in general

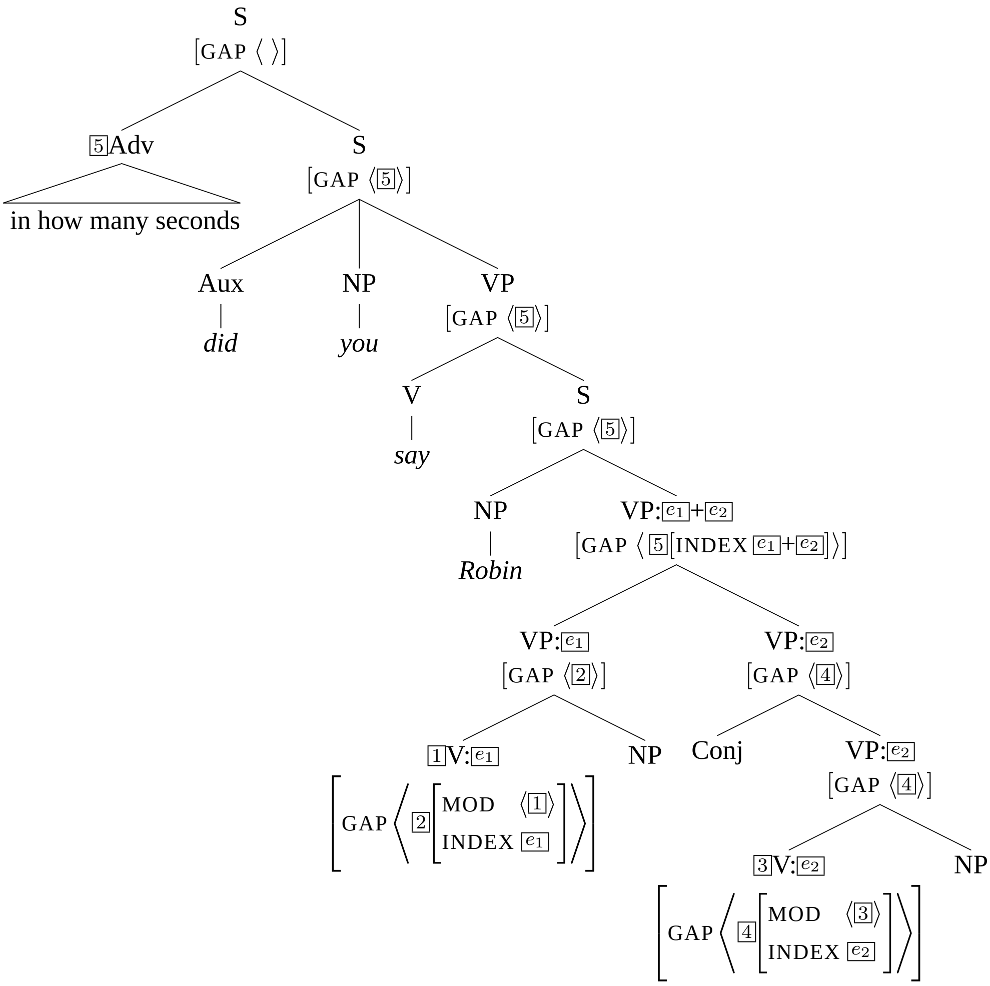


Figure 7: An example of VP conjunction with adverbial gap cumulation

terms Yatabe (2002) and Beavers and Sag (2004). Thus, the rightmost elements (if any) in DOM, and the MOD elements in extraposed structures can be cumulated.

(27) CONJUNCTION RULE (extended):

$$
\textit{conj-cx} \Rightarrow \begin{bmatrix} \text{MOTHER} & \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{GAP} & \boxed{3} \approx \boxed{4} \\ \text{EXTRA} & \langle ([\text{MOD } \boxed{5} \approx \boxed{6}]) \rangle \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{INDEX } \boxed{1}+\boxed{2} \end{bmatrix} \\ \text{DOM} & \boxed{A}_{ne-list} \oplus \boxed{C} \oplus \boxed{D}_{ne-list} \oplus (\boxed{B} \approx \boxed{E}) \end{bmatrix} \\ \text{DTRS} & \left\langle \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{GAP} & \boxed{3} \\ \text{EXTRA} & \langle ([\text{MOD } \boxed{5}]) \rangle \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{INDEX } \boxed{1} \end{bmatrix} \\ \text{DOM} & \boxed{A} \oplus \boxed{B} \end{bmatrix}, \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{GAP} & \boxed{4} \\ \text{EXTRA} & \langle ([\text{MOD } \boxed{6}]) \rangle \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{INDEX } \boxed{2} \end{bmatrix} \\ \text{DOM} & \boxed{C} \langle ([and]) \rangle \oplus \boxed{D} \oplus \boxed{E} \end{bmatrix} \right\rangle \end{bmatrix}
$$

4 Case marking of adverbial NPs

Adverbial NPs have been argued to receive case by essentially the same mechanism that assigns case to valents in a variety of languages. Przepiórkowski (1999) and others have argued that there is no satisfactory way to account for this unless adjuncts are taken to be complements. In this section I suggest an alternative that maintains the standard divide between arguments and modifiers.

Maling (1993) argues that case assignment in Finnish is structural, and that frequency adverbs pattern with arguments with regard to case. In a more recent and extensive study, Kiparsky (2001) argues that there is no direct way to determine the case marking of verbal dependents, and proposes to an Optimality Theory account where abstract and morphosyntactic case must be matched in an optimal way. Kiparsky (2001) uses features like H(IGHEST)R(OLE) $\pm$ and L(OWEST)R(OLE) $\pm$ to capture the various possible levels in the thematic hierarchy, in each level. Thus, in the morphological level case morphemes bear such features, at the syntactic level these features are assigned to the expressions according to the positions that they occupy, and finally, their abstract case reflects hierarchically organized theta-roles at Semantic Form. Case assignment is an optimal match between all three levels. In practice, abstract case features function as constraints on morphosyntactic case. A theta-role's abstract case must optimally match the morphosyntactic case. The account is centered around the idea that declarative sentences contain a 'pivot' position, which typically contains the grammatical subject (if there is one), or certain other elements in restricted conditions. The pivot is the highest direct argument that can be expressed. One of the most interesting aspects of pivots is that their effect is observed in arbitrary distances. If the matrix clause has a pivot, then a singular noun object of an infinitive complement is genitive whether the infinitive has a subject or not. The genitive object marking extends obligatorily down through a chain of such complements, and thus Kiparsky (2001,28) concludes that 'case marking constraints hold within the domain of finite clauses'.

One might import this account to HPSG by resorting to an ancillary function that computes the same conditions as the OT account. The phrasal rules specify the possible values for STRUCTURAL case while the value of LEXICAL case is specified by morphology. The question is, then, whether or not the mechanism that assigns structural case to complements can operate in the same way for adjuncts.[5]

(28) $\left[\text{CASE}\begin{bmatrix}\text{ABSTRACT} & [...] \\ \text{LEXICAL} & [...] \\ \text{STRUCTURAL} & [...]\end{bmatrix}\right]$

Possibly, a general principle like the one sketched in (29) would have access to all of the relevant information. In $\boxed{1}$ we have information about subjects (and more generally, external arguments via X-ARG), and about what kind of phrase is being

[5]Alternatively, one could encode the three dimensions in (28) as a multi-inheritance hierarchy with at least three partitions (although the two are not equivalent, as discussed in Müller (2001)).

considered (VP, S, finite, nonfinite, etc.), in DOM we have access to the pivot when $\boxed{1}$ is a clause, and in particular, nothing seems to preclude this principle from computing the case of an adverbial nominal by essentially the same means that the case of a complement is computed. The actual order in DOM plays only a minor role, since as Kiparsky (2001,15) notes, positional case is mostly redundant in Finnish because morphological case suffices, with the exception of oblique possessors and experiencers as direct arguments.

(29)
$$phrase \Rightarrow \begin{bmatrix} \text{SYN} \begin{bmatrix} \text{LOC } \boxed{1} \end{bmatrix} \\ \text{DOM } \boxed{2} \end{bmatrix} \wedge \text{ASSIGN-CASE}_F(\boxed{1},\boxed{2})$$

The general rule in (29) can access the necessary ingredients and establish the mappings between abstract, lexical, and structural case without giving up locality, and without having to assume that adjuncts are selected for. Finnish does not allow more then one object in VPs, and so the DOM value should be straightforward, with at most one subject, at most one object, one head, and any number of adjuncts.

In the case of Korean there appears to be significant speaker variation, as already noted in Maling (1989, ft.3). More recently, Jae Eun Jung (p.c) reports that 26 non-linguist native speakers residing in Seoul, with ages between 20 and 25, do not agree with the judgments in Maling (1989) and Maling et al. (2001). In this study, 50% of the informants preferred accusative case in the *ci*-passive sentences from Maling (1989,369) rather than nominative case. Similar results were obtained for the passives in Maling (1989,371). This part of the grammar is in flux, which makes it very difficult to draw any conclusions about how exactly case is assigned to frequency and duration adverbials. In deed, the exact conditions that regulate case assignment to nominal adverbials are not yet understood. As Wechsler and Lee (1996,636) write, they 'do not yet understand the factors conditioning the nominative/accusative split on durative adverbials'.

Still, the literature has converged on the following basic observations, mostly drawn from Maling (1989) and Maling et al. (2001). If the frequency adverbial nominal bears case, then in active transitive verbs both the object and the modifier are accusative, in stative verbs both are nominative, and in *ci*-passives both are nominative (but durative adverbials are accusative). In the case of *toy*-passives and *hi*-passives these are held to be ambiguous between active and passives, and thus the case marking follows as in the previous cases. There are some special cases as well, for example, in unergative verbs the adverbs can only bear accusative, and in nonagentive unaccusative verbs frequency adverbials bear nominative and durative bear accusative. In weather verbs and intransitive motion verbs both nominative and accusative are possible for frequency adverbs, which Maling (1989) attributes to structural ambiguity (in one case the adverb modifies the subject, and in the other case in modifies the verbal structure, with scopal semantic contrast). Wechsler and Lee (1996,640) propose that 'accusative is assigned to any case-bearing dependent with an external co-argument, and that nominative is assigned to any case-bearing dependent lacking an external co-argument' (an external argument in their terms is defined as the lexically distinguished argument that passivization suppresses).

The proper account of adverbial case assignment may hinge on semantic / pragmatic factors rather than on syntactic ones, but in what follows I will show how Wechsler and Lee's account can be formulated without giving up the distinction between adjuncts and valents. I start by assuming that Korean adverbial NPs adjoin to V – since their canonical position is immediately before the verb – and that scrambling is due to DOM linearization as in Kathol (2000). Given their position, these adjuncts have local access to all the relevant information for their case marking, namely, the VFORM and ARG-ST values. This is the same information that is needed at the word level to determine the case markings on valents.[6]

We can capture assignment conditions in a relation *Assign-Case*$_K$ that encodes the account in Wechsler and Lee (1996) without giving up the distinction between valents and adjuncts. The rule in (30a) applies to all words and computes the case assignment of valents (if there are any able to bear case). The rule in (30b) computes the case assignment of adverbial nominals from the verbal head that they combine with. Both (30a) and (30b) resort to one and the same assignment relation.

(30) a. *word* $\Rightarrow$

$$\begin{bmatrix} \text{SYN} \begin{bmatrix} \text{LOC} \begin{bmatrix} \text{HEAD } \boxed{1} \end{bmatrix} \end{bmatrix} \\ \text{ARG-ST } \boxed{2} \end{bmatrix} \wedge \text{Assign-Case}_K(\boxed{1},\boxed{2})$$

b. *adv-noun-lxm* $\Rightarrow$

$$\begin{bmatrix} \text{SYN } \boxed{3} \begin{bmatrix} \text{LOC} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \textit{noun} \end{bmatrix} \\ \text{MOD} \begin{bmatrix} \text{HEAD } \boxed{1}\textit{verb} \\ \text{ARG-ST } \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \wedge \text{Assign-Case}_K(\boxed{1},\boxed{2}\oplus\langle\boxed{3}\rangle)$$

For example, case assignment in active transitive verbs can be captured via a single condition (notice the accusative 'case spreading' from complements to adjuncts):

(31)
$$\text{Assign-Case}_K\left(\begin{bmatrix}\text{VFORM } \neg\textit{pass}\end{bmatrix}, \left\langle \text{NP}\begin{bmatrix}\text{CASE } \textit{nom} \\ \text{MOD } \textit{none}\end{bmatrix}, \text{NP}\begin{bmatrix}\text{CASE } \boxed{0}\textit{acc} \\ \text{MOD } \textit{none}\end{bmatrix}\right\rangle \oplus \textit{list}\left(\text{NP}\begin{bmatrix}\text{CASE } \boxed{0} \\ \text{MOD } \textit{verb}\end{bmatrix}\right)\right)$$

5 Conclusion

One view of adjunct extraction and cumulation assumes that gaps are inserted phrasally (either by traces or by a construction) and that verbs control adjuncts via a special feature ADJS. Another view assumes that gaps are lexically inserted, and that shared displaced dependents can be cumulated. In this paper I argue that the

[6]Some elements should be lexically unspecified and obtain structural case Kim and Sells (2007).

latter is done by coordination, and that cumulation is also observed in the extraction of nominal arguments. With minor modifications to Ginzburg and Sag (2000), my analysis dispenses traces, extra constructions, special assumptions about the scope of adjuncts, and the ADJS feature. Extraction pathway marking and case assignment to adverbial NPs can be done without blurring the distinction between complements and adjuncts, since verbs cannot access *in situ* adjuncts.

References

Abbott, Barbara. 1976. Right Node Raising as a Test for Constituenthood. *Linguistic Inquiry* 7, 639–642.

Bach, Emmon. 1986. The algebra of events. *Linguistics and Philosophy* 9, 5–16.

Beavers, John and Sag, Ivan A. 2004. Ellipsis and Apparent Non-Constituent Coordination. In Stefan Müller (ed.), *Proceedings of the HPSG-2004 Conference, Katholieke Universiteit Leuven*, pages 48–69, Stanford: CSLI Publications.

Bonami, Olivier and Godard, Danièle. 2007. Integrating Linguistic Dimensions: The Scope of Adverbs. In Stefan Müller (ed.), *The Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar*, pages 25–45, Stanford: CSLI Publications.

Bouma, Gosse, Malouf, Rob and Sag, Ivan A. 2001. Satisfying Constraints on Extraction and Adjunction. *Natural Language and Linguistic Theory* 1(19), 1–65.

Chaves, Rui P. 2007. *Coordinate Structures – Constraint-based Syntax-Semantics Processing*. Doctoral dissertation, University of Lisbon.

Chaves, Rui P. 2009. Reassessing 'respectively' readings, paper presented at the 83rd Annual Meeting of the LSA, San Francisco, California.

Chaves, Rui P. and Sag, Ivan A. 2007. Two kinds of ellipsis in English coordinate structures, *Under review for journal publication.*

Clements, George N., McClosey, James, Maling, Joan and Zaenen, Annie. 1983. String-vacuous rule application. *Linguistic Inquiry* 14, 1–17.

Copestake, Ann, Flickinger, Daniel, Sag, Ivan A. and Pollard, Carl. 2006. Minimal Recursion Semantics: An Introduction. *Journal Research on Language & Computation* 3(4), 281–332.

Crysmann, Berthold. 2004. Underspecification of Intersective Modifier Attachment: Some Arguments from German. In Stefan Müller (ed.), *Proceedings of the HPSG-2004 Conference, Center for Computational Linguistics, Katholieke Universiteit Leuven*, pages 378–392, Stanford: CSLI Publications.

Ginzburg, Jonathan and Sag, Ivan A. 2000. *Interrogative Investigations: the form, meaning and use of English interrogative constructions*. Stanford: CSLI Publications.

Hukari, Thomas E. and Levine, Robert D. 1995. Adjunct Extraction. *Journal of Linguistics* 31(2), 195–226.

Hukari, Thomas E. and Levine, Robert D. 1996. Phrase Structure Grammar: The Next Generation. *Journal of Linguistics* 32, 465–496.

Jackendoff, Ray. 1977. *$\overline{X}$-Syntax: A study of Phrase Structure*. the MIT Press.

Kathol, Andreas. 2000. *Linear Syntax*. Oxford University Press.

Kehler, Andrew. 2002. *Coherence, Reference, and the Theory of Grammar*. Stanford: CSLI Publications.

Kim, Jong-Bok and Sells, Peter. 2007. Two Types of Multiple Nominative Construction: A Constructional Approach. In Stefan Müller (ed.), *The Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar*, pages 364–372, Stanford: CSLI Publications.

Kiparsky, Paul. 2001. Structural Case in Finnish. *Lingua* 111, 315–376.

Kiss, Tibor. 2005. Semantic Constraints on Relative Clause Extraposition. *Natural Language and Linguistic Theory* 23(2), 281–334.

Lasersohn, Peter. 1995. *Plurality, Conjunction and Events*. Kluwer: Dordrecht.

Levine, R. and Hukari, T. 2006. *The Unity Of Unbounded Dependency Constructions*. CSLI Publications, Leland Stanford Junior University.

Levine, Robert D. 2003. Adjunct valents: cumulative scoping adverbial constructions and impossible descriptions. In Jongbok Kim and Stephen Wechsler (eds.), *The Proceedings of the 9th International Conference on Head-Driven Phrase Structure Grammar*, pages 209–232, Stanford: CSLI Publications.

Link, Godehard. 1998. *Algebraic Semantics in Language and Philosophy*. CSLI Lecture Notes No. 74.

Maling, Joan. 1989. Adverbials and structural case in Korean. In Susumu Kunu, Ik-Hwan Lee, John Whitman, Syng-Yun Bak, Young-Se Kang and Young joo Kim (eds.), *Harvard Studies in Korean Linguistics III*, pages 297–308, Cambridge Massachusetts.

Maling, Joan. 1993. Of Nominative and Accusative: The Hierarchical Assignment of Grammatical Case in Finnish. In Anders Holmberg and Urpo Nikanne (eds.), *Case and Order Functional Categories in Finnish Syntax*, pages 51–76, Dordrecht: Mouton de Gruyter.

Maling, Joan, Jun, Jongsup and Kim, Soowon. 2001. Case marking on duration adverbials revisited. In Ahn and Kim (eds.), *Selected Papers from the Twelfth International Conference on Korean Linguistics*, Seoul: Kyungjin Publishing.

McCloskey, James. 1979. *Model-Theoretic Semantics and Transformational Grammar*. Dordrecht: Reidel.

Müller, Stefan. 1999. *Deutsche Syntax deklarativ. Head-Driven Phrase Structure Grammar für das Deutsche*. Linguistische Arbeiten, No. 394, Tübingen: Max Niemeyer Verlag.

Müller, Stefan. 2001. Case in German – Towards an HPSG Analysis. In Walt Detmar Meurers and Tibor Kiss (eds.), *Constraint-Based Approaches to Germanic Syntax*, Studies in Constraint-Based Lexicalism, No. 7, pages 217–255, Stanford: CSLI.

van Noord, Gertjan and Bouma, Gosse. 1994. The Scope of Adjuncts and the Processing of Lexical Rules. In *Proceedings of COLING 94*, pages 250–256, Kyoto.

Oehrle, Richard T. 1987. Boolean properties in the analysis of gapping. In Geoffrey J. Huck and Almerindo E. Ojeda (eds.), *Syntax and Semantics 20: Discontinuous Constituency*, pages 201–240, Academic Press, Orlando, FL.

Pollard, Carl and Sag, Ivan A. 1994. *Head-driven Phrase Structure Grammar*. Chicago: University of Chicago Press and Stanford: CSLI.

Postal, Paul M. 1998. *Three investigations of extraction*. MIT: Cambridge, Massachusetts.

Przepiórkowski, Adam. 1999. On Case Assignment and "Adjuncts as Complements". In Gert Webelhuth, Jean-Pierre Koenig and Andreas Kathol (eds.), *Lexical and Constructional Aspects of Linguistic Explanation*, Chapter 15, pages 231–245, Stanford, CA: CSLI Publications.

Ross, J. R. and Perlmutter, D. M. 1970. Relative Clauses With Split Antecedents. *Linguistic Inquiry* 1.

Sag, Ivan A. 2000. Another argument against *wh*-trace. Jorge Hankamer Webfest, `http://ling.ucsc.edu/Jorge/sag.html`.

Sag, Ivan A. 2001. Six Dimensions of Natural Language Locality, talk given at the *8th International Conference on Head-driven Phrase Structure Grammar*, Trondheim.

Sag, Ivan A. 2005. Adverb Extraction and Coordination: a Reply to Levine. In Stefan Müller (ed.), *The Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar, Department of Informatics, University of Lisbon*, pages 322–342, Stanford: CSLI Publications.

Sag, Ivan A. and Fodor, Janet. 1994. Extraction without traces. In *Proceedings of the Thirtheenth West Coast Conference on Formal Linguistics*, pages 365–384, Stanford: CSLI.

Sato, Yo and Tam, Wai Lok. 2008. Towards a Unified Account of Adjuncts. In Stefan Müller (ed.), *The Proceedings of the 15th International Conference on Head-Driven Phrase Structure Grammar*, pages 173–188, Stanford: CSLI Publications.

Vergnaud, Jean-Roger. 1974. *French Relative Clauses*. Doctoral dissertation, MIT, Cambridge Massassuchets.

Wechsler, Stephen and Lee, Yae-Sheik. 1996. The Domain of Direct Case Assignment. *Natural Language and Linguistic Theory* 13(3), 629–664.

Yatabe, Sûichi. 2002. A linearization-based theory of summative agreement in peripheral node-raising constructions. In J.-B. Kim and S. Wechsler (eds.), *Proceedings of the 9th International Conference on Head-driven Phrase Structure Grammar*, pages 391–411, Kyung Hee University, Seoul.

Deriving Superficial Ergativity in Nias

Berthold Crysmann

Universität Bonn and Universität des Saarlandes

Proceedings of the HPSG09 Conference

Georg-August-Universität Göttingen, Germany

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

In this paper, I discuss the case and agreement system of Nias, a language that has been described as a marked-absolutive system by various authors (Donohue and Brown, 1999; Corbett, 2006; Cysouw, 2005; Handschuh, 2008; Wichmann, 2005). I shall argue in particular that the ergativity of this language is highly superficial in nature, showing that hypothesised marked-absolutive arguments fail to display typical subject properties. Extending the linking theory of ergativity by Manning (1994) and Manning and Sag (1999), which assumes an inverse linking pattern for transitive, I shall suggest that Nias transitives are best analysed as a Nominative-Accusative system, attributing the "ergative" split in Nias to an "inverse" linking of intransitives instead. Under this perspective, case, agreement, and word order will receive a natural explanation.

1 Case and Agreement in Nias

1.1 Case marking

Nias[1] distinguishes mainly two morphological cases in the nominal system: a morphologically zero-marked case, called the Ergative by some authors (Brown, 1997; Donohue and Brown, 1999), and a morphologically marked case, sometimes referred to as the Absolutive.[2] Case marking of lexical NPs in Nias is effected by initial segmental alternation (Brown, 2005). With pronominals, marked case is further differentiated into Absolutive and Genitive, the latter being used in possessive constructions and with most prepositions.

As depicted in table 1, Nias case marking on consonant initial lexical NPs is signalled by mutation, involving either voicing or trilling. For vowel-initial NPs, marked case is expressed by prefixation of /g/ or /n/, the choice being morphologically (not phonologically) conditioned (Brown, 2005).

Case assignment in Nias (Brown, 1997; Donohue and Brown, 1999) has repeatedly been assumed to belong to the ergative type. The main evidence for this

[†] The Nias data cited in this paper and the presentation of the basic empirical facts are based on Léa Brown's field work on the language, published in a series of papers (Brown, 1997; Donohue and Brown, 1999; Brown, 2005), as well as documented in her University of Sydney dissertation (Brown, 2001).

The analysis proposed here has been presented at the 4th Conference on Austronesian Languages and Linguistics, SOAS, London and the 16th Conference on HPSG. I would like to thank to the audiences at these tow venues for their stimulating questions and comments, in particular to Peter Sells, Peter Austin, Bill Palmer, Sebastian Nordhoff, Ileana Paul, Doug Arnold, Olivier Bonami and Ivan Sag. I am particularly indepted to Nikolaus Himmelmann for providing me with detailed comments, suggestions and criticism, and to my colleague Mats Exter for discussing the ideas proposed here already at an early stage. All remaining errors are of course mine.

[1] Nias is an Austronesian language spoken by over 600,000 speakers on the Barrier islands of Nias and Batu, off the Western coast of Sumatra.

[2] In more recent work, Brown (2001, 2005) has dropped the terms "absolutive" and "ergative" in favour of the descriptively neutral terms "mutated" and "unmutated".

unmutated	mutated
f	v
t	d
k	g
s	z [dʒ]
c [tʃ]	z [dʒ]
b	mb [ʙ]
d	ndr [d^{r}]

Table 1: Nias mutation

typological categorisation comes from the fact that case marking of the **S**(ole) argument in intransitives patterns with that of the **O**(bject) argument in transitives. The **A**(gent) of transitives, however, features case marking which is clearly distinct from the S(ole) argument of intransitives, yielding a partitioning characteristic of other ergative languages.

(1) manavuli sui [**n**-ama-da Tohönavanaetu] ba Maenamölö
return again MUT-father-1.PL.IN.GEN Tohönavanaetu LOC Maenamölö
'Ama Tohonavanaetu came back again to Maenamölö.' (Donohue and Brown, 1999)

(2) I-a [**mb**avi] [ama Gumi]
3.sg.RLS-eat MUT.pig father Gumi
'Father Gumi eats pig.' (Donohue and Brown, 1999)

As illustrated by the data above, mutated case is used to mark both **O** and **S**, arguments, whereas A arguments display zero case marking, a pattern that can be summarised as in table 2.

CASE	∅	MUT
Intr		S
Tr	A	P

Table 2: Nias case patterns

As pointed out by Donohue and Brown (1999), the case marking pattern observed in Nias is not an effect of surface adjacency, or even a pure surface phonological issue: as illustrated by the example below, assignment of mutated case applies even in the case of intervening obliques.

(3) I-be khö-nia **g**-ana'a.
3SG.RLS-give OBL-him MUT-gold
'He gave him (the) gold.' (Donohue and Brown, 1999)

The peculiar case assignment of Nias raises some typologically important issues:as stated by Donohue and Brown (1999), Nias constitutes an apparent exception to Greenberg's Universal 38:

"where there is a case system, the only case which ever has only zero allomorphs is the one which includes among its meanings that of the subject of the intransitive verb" (Greenberg, 1963)

Although Marked-S systems are indeed typologically rare, they are not unheard of: according to Dixon (1994), Marked Nominative systems can be found in the Yuman languages of Southern California, as well as with several languages of the Afroasitic family, mainly Cushitic and Omotic in Eastern Africa, as well as Berber Sasse (1984); Hayward (1990). In addition to these, marked nominative systems have also been reported for several languages of the unrelated Nilotic family (Andersen, 1988; Dimmendaal, 1985; Kiessling, 2007) found in close vicinity to Cushitic and Omotic, making this property qualify as an areal feature.

Marked-absolutive systems, by contrast, appear to be extremely rare: as far as I am aware, apart from Nias, only two languages from the Otomanguean family, namely Tlapanec (Wichmann, 2005) and Chinantec (Foris, 2000) have been argued to be of this type. However, in Tlapanec, evidence for Marked Absolutive is solely located in the system of cross-referencing pronominal affixes in this head-marking language. Thus, it appears that Nias is the only language with dependent marking for which an analysis in terms of Marked Absolutive has been advanced.

An important fact about Nias that should cast some initial doubt about Nias being a marked absolutive language is that morphological unmarkedness aligns pretty well with functional unmarkedness in this language: as discussed at length by Brown (1997), morphologically unmarked "ergative" case is also functionally unmarked. In particular, it is the form used in citation, for core arguments in relatives clauses and infinitivals, and for elliptical answers (see the examples below), .

(4) Intransitive

a. Q: hanata **z**i möi?
who MUT.REL go
'Who went?'

b. A: Ama Doli. / Möi **N**ama Doli.
Ama Doli go MUT.Ama Doli
'Ama Doli. / Ama Doli went.' (Brown, 1997)

(5) a. Q: haija ni-tagö?
what PASS-steal
'What did they steal?

b. A: Kefe- nia. / La-tagö **g**efe-nia.
what PASS-steal 3.P.RLS-steal MUT.money-POSS.3.S
'His money. / They stole his money. (Brown, 1997)'

Furthermore, topicalised preverbal constituents invariably surface with unmarked case.

(6) Si'o hö'ö ma=i-taru-'ö ba danö.
stick DIST PERF=3.S.RLS-plant-TR LOC MUT.ground
'That stick he planted in the ground.' (Brown, 2001)

1.2 Agreement

Nias, just like many Austronesian languages, recognises a major division in the Tense-Mood-Aspect system between Realis and Irrealis mood, a split which is also manifest in the agreement system.

Realis In the realis, verb agreement appears to follow, again, a superficial ergative divide: while **A** argument control verb agreement, both **O** and **S** arguments fail to do so.

(7) a. *I*-tolo **z**i'ila *ama-gu*
3SG.RLS-help MUT.village.advisor father-1SG.POSS
'My father helped the village advisors.' (Brown, 2003)

b. *La*-tolo **n**-ama-gu *si*-ila
3PL.RLS-help MUT-father-1SG.POSS village.advisor
'The village advisors helped my father.' (Brown, 2003)

(8) Mofanö **n**-ama-gu
leave MUT-father-1SG.POSS
'My father left.' (Brown, 2003)

As illustrated by the data above, **A**-arguments, which are unmarked for case, do control agreement on the verb, whereas **S** and **O** arguments, both featuring marked case, do not. As a result, transitives feature agreement morphology, whereas intransitives do not.

Irrealis Agreement in the irrealis, by contrast, does not align with the case system. While case assignment is entirely parallel to that found in the Realis, agreement on the verb is controlled by the highest role (**A** or **S**), irrespective of case marking.

(9) a. Gu-m-örö=e mana?
1.S.I-DYN-sleep=PTCL at.this.time
'I'm going to bed now, ok?' (Brown, 2001)

b. Ya-te-bato **d**eu
3.S.I-RES-stop MUT.rain
'The rain will stop.' (Brown, 2001)

(10) Ndra-m-a'ege-ö ndrao
3.P.I-I-laugh-TR MUT.1.S
'They will laugh at me.' (Brown, 2001)

1.3 Marked absolutive?

Summarising the empirical data, the characterisation of Nias as an ergative language is mainly supported by the alignment patterns: indeed, as far as case marking or agreement in the Realis are concerned, the language treats **S** and **O** arguments similarly, to the exclusion of **A** arguments. However, upon further scrutiny, it becomes apparent that morphologically (and functionally) marked "absolutive" arguments fail to show any properties of prototypical subjects. By contrast, supposedly "ergative", i.e. objective, arguments display all the prototypical properties of subjects, including exclusive control of agreement in the Realis, as well as appearing in the morphologically and functionally unmarked case. In essence, under the perspective of Nias as a Marked-Absolutive language, we are confronted with a typologically doubly marked system: not only that marked-absolutive systems, in general, are typologicaly rare, but also systems, in which the highest ranked case fails to control agreement: although Corbett (2006) tacitly adopts the marked-absolutive analysis of Nias, he still recognises "ergative"-controlled agreement as non-canonical.

Finally, pro-drop in Nias targets **A** arguments. Pronominal **A** arguments are realised by means of a cross-referencing prefix on the verb, the agreement prefix, whereas pronominal **S** and **O** arguments are expressed by means of an independent pronoun. Under the hypothesis that Nias is marked-absolutive, this would be quite a surprising fact, since it forces one to concede that pro-drop in this language makes exclusive reference to objects, again, a typologically rather marked property.

To conclude, Nias looks ergative, if we only consider the alignment of properties, but once we consider the properties themselves, an ergative analysis becomes less and less plausible: in essence, unmutated, supposedly "ergative" **A** arguments exhibit prototypical subject properties, as far as agreement and case are concerned, whereas mutated **O** or **S** arguments systematically lack both. Furthermore, unmarked case is shared between **A** arguments and topics, another property prototypical associated with subjects.

2 Linking

In his dissertation, Manning (1994) argues that syntactically ergative languages display a systematic split between subject properties: while surface-oriented processes, like case marking and agreement indeed follow an **S/O** pivot, other processes, like binding and control are actor-oriented. Building on a distinction between argument structure and valency (grammatical function), he suggests that these split properties can be accounted for, once we recognise two different notions of subjecthood: a-subjects, as thematically least oblique arguments, and surface grammatical subjects. The difference between syntactically accusative and syntactically ergative language is attributed to different linking patterns between these two representations. Accusative languages feature a direct linking between these two levels of representation, identifying the a-subject with subject grammatical function. Ergative languages, however, display an "inverse" linking for transitives, mapping the a-subject to direct object function, and the thematic object to subject function. This theory not only accounts for split subject properties in ergative languages, but also provides a convenient basis for case assignment in terms of grammatical function.

Within HPSG, Pollard and Sag (1994, ch. 9) have suggested to split the SUBCAT list into (at least) two valence lists, SUBJ and COMPS, following proposals by Borsley (1987). The SUBCAT list, being considered at the time a mere concatenation of valence lists was retained as the locus of Binding Theory. Manning and Sag (1999) argued that Manning's theory of ergativity can be straightforwardly integrated into HPSG by parametrising the mapping between argument structure (ARG-ST) and the valence lists SUBJ and COMPS.

In essence, the linking patterns suggested by Manning and Sag (1999) can be schematically represented as follows.

(11) Intransitive linking

Obliqueness:	ARG-ST:	$\langle$ S $\rangle$		
Grammatical Function:	SUBJ:	$\langle$ S $\rangle$	COMPS:	$\langle\,\rangle$

(12) a. Nominative-Accusative linking

Obliqueness:	ARG-ST:	$\langle$ A	,	O $\rangle$
Grammatical Function:	SUBJ:	$\langle$ A $\rangle$	COMPS:	$\langle$ O $\rangle$

b. Ergative-Absolutive linking

Obliqueness:	ARG-ST:	$\langle$ A	,	O $\rangle$
Grammatical Function:	SUBJ:	$\langle$ O $\rangle$	COMPS:	$\langle$ A $\rangle$

Just like in Manning's original proposal, the inverse linking characteristic of ergative languages will derive both the **S/O** alignment in the case system, as well as the split in subject properties.

2.1 Canonical agreement in Ergative languages

Manning's theory of ergativity in terms of "inverse" linking between argument structure and valence lists makes some interesting predictions for subject-verb agreement in ergative languages: since the notions of thematic (a-subject) and grammatical subjects (s-subject) do not coincide, we expect that agreement processes may either align with thematic rank, or with grammatical function and, therefore, case. Indeed, both these systems are actually attested.

The Daghestanian language Archi (Kibrik, 1994) represents an example of the latter type:

(13) Buwa d-arχaši d-i
mother.II.A **II**-lie.down **II**.be
'Mother is lying down.' (Kibrik, 1994)

(14) Buwa-mu b-ez diṯa<b>u $\bar{\chi}^{w}$alli a<b>u
mother.II-E **III**-1.S.D early<**III**> **bread.III.A** made<**III**>
'Mother made bread for me early.' (Kibrik, 1994)

This pattern is readily explained, if subject-verb agreement in this language is controlled by the surface subject, i.e. the SUBJ valency.

Intransitive			Transitive		
ARG-ST:	⟨ S ⟩		ARG-ST:	⟨ A , O ⟩	
SUBJ:	⟨ S ⟩	COMPS: ⟨ ⟩	SUBJ:	⟨ O ⟩	COMPS: ⟨ A ⟩

An example of the other type is contributed by Udi, another language of the Caucasus (Harris, 1984):

(15) zu a-r-e-zu k'wa
1.S.ABS hither-come-AOR-**1.S** home
'I came home.' (Harris, 1997)

(16) zu a-za-k'-sa šel läzätt'u pak.
1.S.ERG see$_1$-**1.S**-see$_2$-PRES good pretty garden.ABS
'I see a good, pretty garden.' (Harris, 1984)

As illustrated by the data above, the case system in Udi exhibits the typical ergative split. Subject-verb agreement, however, does not align with the case system, being uniformly controlled by the thematically highest role, i.e., the initial element on ARG-ST.

3 Nias as a nominative-accusative language

3.1 Outline of the analysis

In section 1, we have seen that **A** arguments of transitive verbs display all the typical properties of subjects: they receive morphologically and functionally unmarked case, they control agreement in the Realis, they undergo pro-drop, and they surface in peripheral position.

If this perspective of **A** arguments as surface subjects is correct, this means that Nias transitive verbs display a direct linking characteristic of nominative-accusative languages, and not an inverse linking.

(17) Nias direct transitive linking

Obliqueness:	ARG-ST:	⟨ A , O ⟩	
Grammatical Function:	SUBJ:	⟨ A ⟩	COMPS: ⟨ O ⟩

Assuming that this analysis is essentially on the right track, we are confronted with the following two remaining issues:

1. How to account for superficial ergativity?
2. How to account for lack of subject properties in intransitives?

As we shall see shortly, the answer to these questions rests on the analysis of intransitives: in particular, I shall propose that Manning's theory needs to be extended, recognising a second canonical linking pattern for intransitives, which assigns the **S** to object, rather than subject function.

Intransitive linking revisited Supporting evidence for such an extension comes from Split-S systems, i.e. languages that systematically differentiate unergtive and unaccusative verbs. One such language is related, near-by Acehnese.

According to Durie (1987), the distinction between actor and non-actor arguments is grammaticalised in the Acehnese agreement system: while actor arguments of transitive and intransitive verbs trigger obligatory agreement on the verb, realised as a proclitic, non-actor arguments, including **S** arguments of unaccusative verbs and **O** arguments of transitives, only trigger optional agreement marking, realised by an enclitic.

(18) a. (gopnyan) geu= jak
(3.S) 3.S= go
'(s)he goes' (Durie, 1987)

b. (gopnyan) rhët (=geuh)
(3.S) fall (=3.S)
'(s)he falls' (Durie, 1987)

(19) (gopnyan) ka lôn= ngieng (=geuh)
(3.S) CPL 1.S= see (=3.S)
'I saw him/her' (Durie, 1987)

In order to provide an account for Split-S systems and to capture the striking parallelism between non-actor **S** arguments with **O** arguments on the one side and between actor **S** arguments with **A** arguments of transitives on the other, we need to complement the intransitive linking pattern recognised in Manning and Sag (1999) with the following pattern, which maps the **S** argument of intransitives onto the COMPS valency list instead:

(20) "Inverse" intransitive linking

Obliqueness:	ARG-ST:	⟨ S ⟩		
Grammatical Function:	SUBJ:	⟨ ⟩	COMPS:	⟨ S ⟩

Besides grammaticalised mapping of **S** arguments to complement function, as witnessed by Acehnese, subject-less intransitives can also be observed in languages that otherwise canonically map **S** arguments of unaccusatives to SUBJ function (e.g. German).

(21) a. weil mich friert
because me.ACC freezes.3.SG
'because I am freezing'

b. weil ich friere
because I.NOM freeze.1.SG
'because I am freezing'

As witnessed by the example above, the sole argument of a German verb like **frieren** can be realised either as an accusative direct object complement, or as a nominative subject. If the **S** argument is realised as a nominative subject, as in the b. sentence above, the verb obligatorily agrees with it, if it is realised as a complement, i.e., if the verb is subject-less, default third singular agreement is chosen.

This fourth linking type, independently motivated by Acehnese unaccusatives, will form the basis of our reanalysis of the Nias case and agreement system, ultimately enabling us to reconcile the superficial ergative split with the observed lack of subject properties of intransitive **S** arguments.

(22) Nias linking patterns

	Intransitive				Transitive		
ARG-ST:	⟨ S ⟩			ARG-ST:	⟨ A	,	O ⟩
SUBJ:	⟨ ⟩	COMPS:	⟨ S ⟩	SUBJ:	⟨ A ⟩	COMPS:	⟨ O ⟩

Thus, in contrast to most systems, which generalise the unergative pattern to all intransitives, Nias chooses the other option licensed by Universal Grammar, namely generalising the unaccusative linking pattern.

Once we adopt this position, a straightforward account of the properties of the Nias case and agreement system falls readily into place: treating **S** arguments of intransitives as surface complements accounts both for their lack of subject properties (marked case, no agreement in the Realis, no pro-drop) and for the superficially ergative pattern, as these arguments are mapped onto exactly the same grammatical functions as **O** arguments of transitives. Adopting an "inverse" intransitive linking instead of an ergative-type inverse transitive linking, our analysis of Nias can do full justice to the subject properties of **A** arguments as essentially a Nominative-Accusative system.

Case and Agreement in the Irrealis The approach outlined so far can be straighforwardly applied to account for case and agreement marking in the Nias Irrealis as well. As we have seen in section 1 above, case marking patterns in the Irrealis are exactly parallel to those in the Realis. Since structural case assignment in a Manning-style linking theory applies on valence lists, we can conclude that the Realis/Irrealis alternation leaves the linking patterns unaffected.

With respect to agreement, however, we find considerable differences: while in the Realis, only **A** arguments (= surface subjects) control agreement and undergo pro-drop, both **A** and **S** arguments (= a-subjects), function as agreement controllers in the Irrealis. Likewise both can undergo pro-drop. Under the account presented here, the difference between Realis and Irrealis agreement patterns is captured by reference to the two different notions of subject. Recall further, that both types of agreement, that is agreement with s-subjects and agreement with a-subjects are cross-linguistically valid options.

Experiencer verbs (double mutation) The behaviour of experiencer-stimulus verbs ('like', 'be afraid of' etc.) also fits in quite neatly with this new perspective on Nias linking: in contrast to transitives, these verbs assign mutated case to both the experiencer and the stimulus, a fact that is easily derived, if we assume that these verbs pattern with intransitives (cf. Brown, 2001), as far as linking is concerned. As expected, agreement in the Realis is null.

(23) A-ta'u **mb**a'e **n**-ono matua
ST-fear MUT.monkey MUT.child male
'The monkey is afraid of the boy.' (Brown, 2005)

In the Irrealis, however, agreement morphology corresponds to the experiencer argument.

(24) Ndra-omasi **v**-a-maigi ono s=aßena tumbu.
3.P.I-like MUT-IPF-see child REL=just.now born
'They like to see the new born child.' (Brown, 2001)

Again, this is in line with our theory of case and agreement in Nias which states that Irrealis agreement should be independent of surface grammatical function, and therefore independent of case, whereas Realis agreement should always be controlled by an unmutated surface subject.

Word order The different status in terms of valency lists for mutated and unmutated arguments is further supported by word order facts. Umutated arguments in Nias surface in right-peripheral position, whereas mutated arguments appear closer to the verb.

According to HPSG's standard theory of subcategorisation (Pollard and Sag, 1994; Borsley, 1987), we actually expect SUBJ valencies to be realised outside head-complement structures, giving rise to a contoured phrase structure. Since unmutated **A** arguments are the only elements assigned to the SUBJ list under the current analysis, their peripheral realisation is actually predicted.

COMPS valencies, however, are saturated simultaneously by virtue of the Head-Complement Schema, giving rise to a flat phrase structure. Thus, if mutated **S** arguments are indeed complements they should, in principle, be able to surface in an internal position, intervening between the verb and other, more oblique complements. Again, this expectation is confirmed by A-subjects of experinecer verbs (Brown, 2001), which appear in internal position.

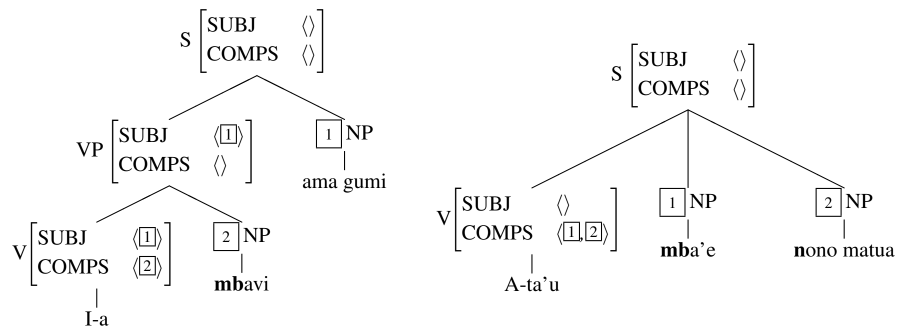


Figure 1: Peripheral vs. internal realisation

The perspective of Nias unmutated **A** arguments as nominative subjects, instead of ergative objects, also aligns quite well with typological observations regarding word order in Western Austronesian: As stated by Himmelmann (2005), VXS basic word order, together with evidence for a VP constituent is a common typological pattern in these languages.

3.2 Case assignment and agreement

In the previous section, we have seen how a change of perspective from Nias as an ergative language, to an accusative language with "inverse" linking of intransitives can account for the subject properties of unmutated **A** arguments (case, agreement, word order), the lack of such properties observed with unmutated **S** and **O** arguments. At the same time this shift in perspective models the superficial "ergativity" of the language, namely the similarity of **O** and **S** arguments, on the basis of their being non-subject complements.

In this section, we will develop the details of case assignment and agreement in Nias, systematically building on the linking suggested in the previous section.

3.2.1 Linking

The basis of our formal analysis of Nias case and agreement are the two linking patterns used to assign core arguments to grammatical functions. Oblique, non-core arguments are indiscriminately assigned to the COMPS list. Following Manning and Sag (1999), I shall assume that linking patterns are constraints on lexeme classes.[3] For our purposes, we shall capture the difference between core and oblique

[3] As a result, morphosyntactic rules will be able to derive non-canonical linkings.

arguments by reference to their case values.

Transistive verbs are characterised by having two core arguments on their ARG-ST list, whereas intransitive verbs only have one core argument. Both transitives and intransitives may specify additional oblique arguments.

(25) Direct transitive linking

$$\begin{bmatrix} \textit{lexeme} & \\ \text{ARG-ST} & \left\langle \boxed{1}\ \text{NP}[\textit{core}], \boxed{2}\ \text{NP}[\textit{core}] \mid \boxed{3}\ \textit{list}\left(\textit{XP[oblique]}\right) \right\rangle \end{bmatrix}$$
$$\rightarrow \begin{bmatrix} \text{SYNSEM}|\text{LOC}|\text{CAT}|\text{VAL} & \begin{bmatrix} \text{SUBJ} & \left\langle \boxed{1} \right\rangle \\ \text{COMPS} & \left\langle \boxed{2} \mid \boxed{3} \right\rangle \end{bmatrix} \end{bmatrix}$$

(26) “Inverse” intransitive linking

$$\begin{bmatrix} \textit{lexeme} & \\ \text{ARG-ST} & \left\langle \boxed{1}\ \text{NP}[\textit{core}] \mid \boxed{2}\ \textit{list}\left(\textit{XP[oblique]}\right) \right\rangle \end{bmatrix}$$
$$\rightarrow \begin{bmatrix} \text{SYNSEM}|\text{LOC}|\text{CAT}|\text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{COMPS} & \left\langle \boxed{1} \mid \boxed{2} \right\rangle \end{bmatrix} \end{bmatrix}$$

3.2.2 Case assignment

Default case Brown (2001, 2005) has shown convincingly that unmutated case is both morphologically and functionally unmarked: apart from being used in citations, it is the case found in elliptical answers, on predicate nominals, and on non-initial conjuncts in coordinate structures Finally, Nias recognises at least two structural domains where case distinctions for core arguments are neutralised. Topicalised (pre-verbal) core arguments invariably surface in unmutated case, regardless of grammatical function. Similarly, core arguments of dependent predicates invariably appear in the unmutated case, including **O** and **S** arguments (see section 3.3 below). Systematic case alternation of the type described above is restricted to the canonical, postverbal position of finite verbs.

Given the heterogeneity of environments in which unmutated case can surface and its unmarked status, I shall adopt Brown’s position and assume that unmutated case represents the default case marking in Nias.

In order to capture this intuition formally, we need to distinguish between morphosyntactic case and its morphological reflex (mutation).[4] The correspondence between the two will be captured by the following two implicational constraints, reminiscent of Feature Cooccurrence Restrictions (Gazdar et al., 1985):

(27) $\begin{bmatrix} \text{CASE} & \text{nom} \end{bmatrix} \rightarrow \begin{bmatrix} \text{MUT} & \text{-} \end{bmatrix}$

[4]Unless these two notions are represented as values of distinct features, purely morphological specifications will always be able to override the default constraint, without any syntactic licensing.

(28) $\begin{bmatrix}\text{CASE} & \text{acc}\end{bmatrix} \rightarrow \begin{bmatrix}\text{MUT} & +\end{bmatrix}$

Nominal inflectional rules merely register whether or not the noun is mutated. Nominative syntactic case is then specified as the default case in Nias:

(29) $\begin{bmatrix}\text{SYNSEM}|\text{L}|\text{CAT}|\text{HD} & \textit{noun}\end{bmatrix} \rightarrow \begin{bmatrix}\text{SYNSEM}|\text{L}|\text{CAT}|\text{HD}|\text{CASE} & \textit{/nom}\end{bmatrix}$

As depicted by the constraint above, default unmutated case is captured as a defeasible property of nominal signs: this is the case in which nominal expressions will surface, unless dictated otherwise by case assignment constraints.

Canonical case assignment Having established by way of linking constraints how core arguments are associated with grammatical functions, we can now restrict the assignment of structural case exclusively in terms of valence features.

(30)
$$\begin{bmatrix}\textit{word} \\ \text{SYNSEM}|\text{L}|\text{CAT}|\text{VAL} \begin{bmatrix}\text{SUBJ} & \left\langle \begin{bmatrix}\text{L}|\text{CAT}|\text{HD}|\text{CASE} & \textit{core}\end{bmatrix} \right\rangle\end{bmatrix}\end{bmatrix}$$
$$\rightarrow \begin{bmatrix}\text{SYNSEM}|\text{L}|\text{CAT}|\text{VAL} \begin{bmatrix}\text{SUBJ} & \left\langle \begin{bmatrix}\text{L}|\text{CAT}|\text{HD}|\text{CASE} & \textit{nom}\end{bmatrix} \right\rangle\end{bmatrix}\end{bmatrix}$$

As depicted above, SUBJ valencies are indiscriminately assigned unmutated case, while core arguments on COMPS receive mutated case, as shown below.

(31)
$$\begin{bmatrix}\textit{word} \\ \text{SYNSEM}|\text{L}|\text{CAT} \begin{bmatrix}\text{DEP} & - \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \langle ...\boxed{1} \begin{bmatrix}\text{L}|\text{CAT}|\text{HD}|\text{CASE} & \textit{core}\end{bmatrix} ...\rangle\end{bmatrix}\end{bmatrix}\end{bmatrix}$$
$$\rightarrow \begin{bmatrix}\text{SYNSEM}|\text{L}|\text{CAT}|\text{VAL} \begin{bmatrix}\text{COMPS} & \langle ...\ \boxed{1} \begin{bmatrix}\text{L}|\text{CAT}|\text{HD}|\text{CASE} & \textit{acc}\end{bmatrix} ...\rangle\end{bmatrix}\end{bmatrix}$$

Case assignment constraints apply to lexical signs of type *word*, i.e., maximal lexical signs. As a consequence, the case constraints will take into account the effects of valency-changing lexical rules.

Agreement As we have seen above, agreement patterns in Nias are sensitive to the major divide between Realis and Irrealis mood. Since Realis agreement is controlled by surface subjects, in line with unmutated case, we can straightforwardly constrain the verb's agreement feature to be reentrant with the INDEX feature of the SUBJ valency.

(32) Realis agreement (S-Subject)

$$\begin{bmatrix}\text{SYNSEM}\,|\,\text{LOC} & \begin{bmatrix}\text{CAT} & \begin{bmatrix}\text{HEAD} & \begin{bmatrix}\text{VFORM} & \textit{realis}\end{bmatrix} \\ \text{VAL}\,|\,\text{SUBJ} & \langle\begin{bmatrix}\text{LOC}\,|\,\text{CONT}\,|\,\text{HOOK}\,|\,\text{INDEX}\ \boxed{i}\end{bmatrix}\rangle\end{bmatrix}\end{bmatrix}\end{bmatrix}$$
$$\rightarrow \begin{bmatrix}\text{SYNSEM}\,|\,\text{LOC}\,|\,\text{CAT} & \begin{bmatrix}\text{HEAD}\,|\,\text{AGR}\ \boxed{i}\end{bmatrix}\end{bmatrix}$$

Irrealis agreement, which is controlled by the thematically highest core argument is determined by the INDEX of the first member on ARG-ST, the A-subject.

(33) Irrealis agreement (A-subject)

$$\begin{bmatrix}\text{SYNSEM}\,|\,\text{LOC} & \begin{bmatrix}\text{CAT} & \begin{bmatrix}\text{HEAD} & \begin{bmatrix}\text{VFORM} & \textit{irrealis}\end{bmatrix} \\ \text{ARG-ST} & \langle\begin{bmatrix}\text{LOC}\,|\,\text{CONT}\,|\,\text{INDEX}\ \boxed{i}\end{bmatrix}, \ldots\rangle\end{bmatrix}\end{bmatrix}\end{bmatrix}$$
$$\rightarrow \begin{bmatrix}\text{SYNSEM}\,|\,\text{LOC}\,|\,\text{CAT} & \begin{bmatrix}\text{HEAD}\,|\,\text{AGR} & \boxed{i}\end{bmatrix}\end{bmatrix}$$

As can be easily verified, the set of constraints proposed thus far derive the basic case and agreement properties of Nias. What may not be so evident is that the current theory already covers case assignment to topicalised constituents, which appear in preverbal, rather than the canonical post-verbal position. Assuming a standard HPSG approach to Nias topicalisation in terms of lexical extraction rules, the relevant valency will have already been removed from either SUBJ or COMPS at the point where word-level case assignment rules apply. Thus, in the absence of local case assignment constraints, topicalised core arguments are free to receive default case.

3.3 Complex predicates

So far, we have only considered the case and agreement properties of basic finite verbs in the Realis and Irrealis. In this last section we will extend our approach to complex predicates used for the expression of progessives and purposives.

Imperfective constructions Besides the major system of Realis vs. Irrealis marking, verbs in Nias can also be inflected for Imperfective. As documented by Brown (2005), the language employs two distinct markers for this category, an infix *-um-* and the prefix *maN-*, the latter being used for transitives. Agreement in the Imperfective is always zero. Another peculiarity that sets the Imperfective apart from other verb forms is that both core arguments of transitive verbs appear with mutated case.

(34)

Man-uri	zawi	ya
IPF-keep.alive	MUT.cattle	MUT.3SG

'He keeps cattle.' (Brown, 2005)

In the context of our approach, we can readily account for zero agreement and double mutation by means of a valence-changing lexical rule along the following lines:

(35)
$$\begin{bmatrix} \text{PH} & \boxed{0} \\ \text{SYNSEM}|\text{LOC}|\text{CAT} & \begin{bmatrix} \text{VAL} & \begin{bmatrix} \text{SUBJ} & \left\langle \boxed{1} \right\rangle \\ \text{COMPS} & \left\langle \boxed{2} \mid \boxed{3} \right\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix}$$
$$\mapsto \begin{bmatrix} \text{PH} & \langle \text{maN} \rangle \oplus \boxed{0} \\ \text{SYNSEM}|\text{LOC}|\text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{VFORM} & \textit{imperf} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \left\langle\, \right\rangle \\ \text{COMPS} & \left\langle \boxed{2}, \boxed{1} \mid \boxed{3} \right\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The result of rule application will be a derived subject-less representation akin to that of experiencer verbs.

Progressive Alongside their independent use, imperfective verb forms also feature in two complex constructions, the progressive and the purposive (Brown, 2005). The progressive is formed by means of the verb *lau* ‘do’ typically taking an imperfective complement.

(36)

I-lau	t<um>ataro	ba	n-ora	n-omo	*ama-gu*
3.S.RLS-do	IPF-sit	LOC	MUT-step	MUT-house	father-POSS.1.S

‘My father is siting on the door step.’ (Brown, 2001)

(37)

I-lau	ma-makha	balale	*ina-gu*
3.S.RLS-do	IPF-weave	basket	mother-POSS.1.S

‘My mother is weaving a basket.’ (Brown, 2001)

In contrast to the Imperfective, the progressive “auxiliary” agrees with the raised **A** or **S** argument of the imperfective complement. Besides controlling agreement on the auxiliary, the raised argument exhibits further prototypical s-subject properties, namely unmutated case and (right-)peripheral surface position. Non-raised arguments receive default unmutated case, which is characteristic of dependent contexts (Brown, 2001, 2005).

I therefore propose the following lexical entry for the progressive raising verb *lau*:

(38)
$$
\begin{bmatrix} \text{PH} & \langle \text{lau} \rangle \\ \text{SYNSEM}\,|\,\text{L}\,|\,\text{CAT} & \begin{bmatrix} \text{HD} & \textit{verb} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \left\langle \boxed{1} \begin{bmatrix}\text{L}\,|\,\text{CONT}\,|\,\text{HOOK}\,|\,\text{INDEX} & \boxed{i}\end{bmatrix} \right\rangle \\ \text{COMPS} & \left\langle \boxed{2} \begin{bmatrix} \text{L} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{DEP} & + \\ \text{HD}\,|\,\text{VFORM} & \textit{imprf} \\ \text{COMPS} & \langle \boxed{1} \rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix}\text{HOOK}\,|\,\text{XARG} & \boxed{i}\end{bmatrix} \end{bmatrix}\end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix}
$$

As shown above, *lau* raises the unsaturated valency of its complement's highest argument onto its own SUBJ list. The restriction that raising can only target the highest argument of the verb is captured by reference to the XARG hook feature (Copestake et al., 2001), which points to the index variable of the verb's least oblique argument in the semantic representation (MRS; Copestake et al., 2005).

A brief note on the assignment of default case: as captured by the lexical entry of *lau* above (and, for that matter, that of purposive *möi* below), the feature structure of the dependent imperfective complement is restricted to be [DEP +]. Since our constraints on mutated case assignment given in the preceding subsection is conditioned on [DEP -], core arguments on the dependent verb's COMPS list will actually be exempt from structural case assignment, receiving default nominative case instead.

"Purposive clauses" The second complex predicate involving imperfective verbs as complements are so-called purposive clauses (see again Brown, 2001, 2005 for an overview of the construction).

(39)	Möi	**ndrao**	ma-mili	eu	s=o-guna	ba-omo
	go	1.S.MUT	IPF-choose	wood	REL=HAVE-use	LOC-house

'I'm going to choose the wood for the house.' (Brown, 2001)

(40)	*Gu*-möi	manai	böli-nia
	1.S.IRR-go	IPF.get	price-POSS.3.S

'I'll go and get the money for it.' (Brown, 2001)

While non-raised complements of the dependent imperfective verb appear again with unmutated (default) case, the raised **A** or **S** argument is assigned mutated case. In addition to case, the raised argument in this construction displays all the other properties characteristic of surface objects, which clearly set it apart from the raised argument in the progressive: as illustrated by the examples above, the raised argument controls agreement in the Irrealis, but not in the Realis, suggesting that the argument is raised to COMPS, not SUBJ. Obligatory internal realisation further conforms the non-subject status of the raised argument.

As captured by the following lexical entry for *möi*, I suggest that this verb raises a distinguished argument of its imperfective complement onto its COMPS and ARG-ST list.

(41)
$$\begin{bmatrix} \text{PH} & \langle \text{möi} \rangle \\ \text{SYNSEM}|\text{L}|\text{CAT} & \begin{bmatrix} \text{HD} & \textit{verb} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle\,\rangle \\ \text{COMPS} & \left\langle \begin{array}{l} \boxed{1}\begin{bmatrix}\text{L}|\text{CONT}|\text{HOOK}|\text{INDEX} & \boxed{i}\end{bmatrix}, \\ \boxed{2}\begin{bmatrix}\text{L} & \begin{bmatrix}\text{CAT} & \begin{bmatrix}\text{DEP} & + \\ \text{HD}|\text{VFORM} & \textit{imprf} \\ \text{COMPS} & \langle\boxed{1}\rangle\end{bmatrix} \\ \text{CONT} & \begin{bmatrix}\text{HOOK}|\text{XARG} & \boxed{i}\end{bmatrix}\end{bmatrix}\end{bmatrix} \end{array} \right\rangle \end{bmatrix}\end{bmatrix} \\ \text{ARG-ST} & \langle \boxed{1},\boxed{2}\rangle \end{bmatrix}$$

To summarise our discussion of complex predicates, progressives exhibit the same clustering of S-subject properties characteristic of simple predicates, namely peripheral realisation (VOS word order), agreement in the Realis, and unmarked case. Likewise, raising to COMPS in the purposive construction replicates the clustering of non-subject properties already observed with experiencer verbs, namely internal realisation, lack of agreement in the Realis, and marked, mutated case.

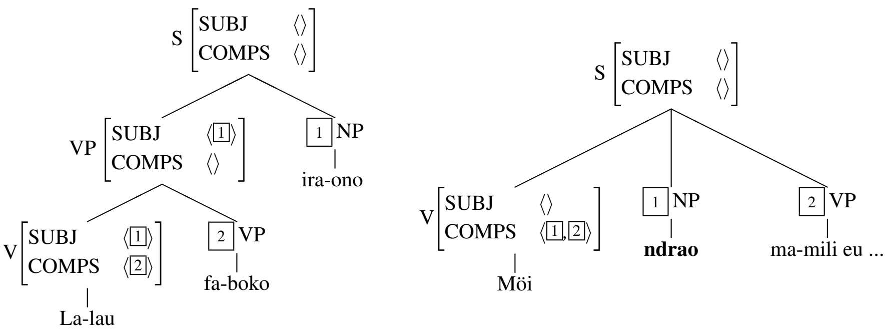


Figure 2: Peripheral vs. internal realisation of raised arguments

4 Conclusion

In this paper I have proposed an analysis of the Nias case and agreement system according to which transitives display a direct Nominative-Accusative linking pattern, whereas the sole argument of intransitives is mapped to direct object function (member of COMPS). Thus, unlike true ergative systems, which display an inverse linking of transitive core arguments, the alignment of **S** and **O** arguments in Nias is derived from an “inverse” intransitive linking. In contrast to most previous approaches, which highlight the typological rarity of "marked absolutive" systems (Donohue and Brown, 1999; Corbett, 2006; Cysouw, 2005; Wichmann, 2005; Handschuh, 2008), the current analysis not only accounts for the superficially "ergative" alignment pattern, but also locates prototypical subject properties (agreement, unmarked case, external surface position) with the notion of grammatical subject. Under the alternative view, namely that of a Marked-Absolutive system, the apparent lack of subject properties of supposedly "absolutive" arguments remains a complete miracle.

On the basis of the Nias data, I have argued that the theory of argument structure-valence correspondence developed by Manning (1994); Manning and Sag (1999) should be extended with an alternative "inverse" linking patterns for intransitive verbs which assigns the sole core argument of intransitive verbs to COMPS valence list, rather than SUBJ. This move not only paves the way for a straightforward analysis of Nias case and agreement in terms of grammatical function, but was also shown to be independently motivated by Split-S systems like Acehnese, as well as lexical subject-less verbs in German.

References

Andersen, T. 1988. Ergativity in Päri, a Nilotic OVS Language. *Lingua* 75, 285–324.

Borsley, R. 1987. *Subjects and Complements in HPSG*. Stanford: CSLI.

Brown, Lea. 1997. Nominal Mutation in Nias. In Cecilia Odé and Wim Stokhof (eds.), *Proceedings of the 7th Intl' Conference on Austronesian Linguistics*, pages 395–413, Amsterdam: Editions Rodopi B.V.

Brown, Lea. 2001. *A Grammar of Nias Selatan*. Ph. D.thesis, University of Sydney.

Brown, Lea. 2003. Nias: An exception to universals of argument marking, paper presented at the Association of Linguistic Typology 5, Sardinia, 15–18 Sep.

Brown, Lea. 2005. Nias. In Alexander Adelaar and Nikolaus Himmelmann (eds.), *The Austronesian Languages of Asia and Madagaskar*, pages 562–589, London: Routledge.

Copestake, Ann, Flickinger, Dan, Pollard, Carl and Sag, Ivan. 2005. Minimal Recursion Semantics: an introduction. *Research on Language and Computation* 3(4), 281–332.

Copestake, Ann, Lascarides, Alex and Flickinger, Dan. 2001. An Algebra for Semantic Construction in Constraint-based Grammars. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics, Toulouse, France*.

Corbett, Greville. 2006. *Agreement*. Cambridge: Cambridge UP.

Cysouw, Michael. 2005. Marked Absolutive and Marked Nominative Case Systems in Synchronic and Diachronic Perspective. In *Grammatik und Verarbeitung verbaler Argumente. Antrag auf Förderung einer DFG-Forschergruppe*, pages 89–110, Leipzig: Universität Leipzig, MPI für Kognitions- und Neurowissenschaften, MPI für evolutionäre Anthropologie.

Dimmendaal, Gerrit J. 1985. Prominence Hierarchies and Turkana Syntax. In Gerrit J. Dimmendaal (ed.), *Current Approaches to African Linguistics*, volume 3, pages 127–148, Dordrecht: Foris.

Dixon, R. M. W. 1994. *Ergativity*. Cambridge: Cambridge UP.

Donohue, Mark and Brown, Lea. 1999. Ergativity: Some additions from Indonesia. *Australian Journal of Linguistics* 19(1), 57–76.

Durie, M. 1987. Grammatical Relations in Acehnese. *Studies in Language* 11, 365–399.

Foris, David Paul. 2000. *A grammar of Sochiapan Chinantec*. Studies in Chinantec Languages, No. 6, Arlington: SIL International and The University of Texas at Arlington.

Gazdar, Gerald, Klein, Ewan, Pullum, Geoffrey K. and Sag, Ivan. 1985. *Generalized Phrase Structure Grammar*. Oxford: Basil Blackwell.

Greenberg, Joseph. 1963. Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. In *Universals of Human Language*, pages 58–90, Cambridges, MA: MIT Press.

Handschuh, Corinna. 2008. Split Marked-S Case Systems. In Marc Richards and Andrej L. Malchukov (eds.), *Scales*, Linguistische Arbeitsberichte, No. 86, pages 73–82, Leipzig: Universität Leipzig.

Harris, Alice. 1984. Case Marking, Verb Agreement, and Inversion in Udi. In David M. Perlmutter and Carol G. Rosen (eds.), *Studies in Relational Grammar 2*, pages 243–258, Chicago: University of Chicago Press.

Harris, Alice. 1997. What's in a Word? The Problem of Endoclisis in Udi, Paper presented atEastern States Conference on Linguistics (ESCOL), Yale University, CT.

Hayward, R. J. 1990. Notes on the Zayse Language. In R. J. Hayward (ed.), *Omotic Language Studies*, pages 210–355, London: SOAS.

Himmelmann, Nikolaus. 2005. The Austronesian Languages of Asia and Madagaskar: Typological Characteristics. In AlexanderAdelaar and Nikolaus Himmelmann (eds.), *The Austronesian Languages of Asia and Madagaskar*, pages 110–181, London: Routledge.

Kibrik, Aleksandr E. 1994. Archi. In Rieks Smeets (ed.), *Indigenous Languages of the Caucasus IV: North-East Caucasian Languages II: Presenting the Three Nakh Languages and Six Minor Lezgian Languages*, pages 297–365, Delmar, NY: Caravan Books.

Kiessling, Roland. 2007. The "Marked Nominative" in Datooga. *JALL* 28, 149–191.

Manning, Christopher and Sag, Ivan. 1999. Dissociations between Argument Structure and Grammatical Relations. In Andreas Kathol, Jean-Pierre Koenig and Gert Webelhuth (eds.), *Lexical and Constructional Aspects of Linguistic Explanation*, Studies in Constraint-based Lexicalism, Stanford: CSLI Publications.

Manning, Christopher D. 1994. *Ergativity: Argument Structure and Grammatical Relations*. Ph. D.thesis, Stanford.

Pollard, Carl and Sag, Ivan. 1994. *Head–Driven Phrase Structure Grammar*. Stanford: CSLI and University of Chicago Press.

Sasse, Hans-Jürgen. 1984. Case in Semitic, Cushitic, and Berber. In J. Bynon (ed.), *Current Progress in Afro-Asiatic Linguistics*, pages 111–126, Amsterdam: Benjamins.

Wichmann, Søren. 2005. Tlapanec cases. In Rosemary Beam de Azcona and Mary Paster (eds.), *Conference on Otomanguean and Oaxacan Languages, March 19-21, 2004, University of California at Berkeley*, pages 133–145, Berkeley: University of California at Berkeley.

French VN lexemes: morphological compounding in HPSG

Marianne Desmets
U. Paris Ouest/LLF

Florence Villoing
U. Paris Saint-Denis/SFL

Proceedings of the HPSG09 Conference

Georg-August-University at Göttingen, Germany

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

Although the original framework of HPSG is mostly compatible with independent theoretical claims or analyses in lexeme base morphology (Anderson 1992, Aronoff & Fudeman 2004, Beard 1995, Booij 2005, Carstairs-McCarthy 1992, Fradin 2003, Haspelmath 2002, Matthews 1991, Plag 2003, for example), so far, most morphological research in morphology has been done on inflexional phenomena (Orgun & Inkelas 2002, Bonami & Boyé 2006), and few on derivational morphology has been addressed by only a few (Koenig 1999, Riehemann 1998). Yet, we believe it is worth investigating how the formal and theoretical apparatus of HPSG deals with capturing multilevel constraints that apply in the lexeme formation of French Verb-Noun nominal compounds in French (, such as as GRILLE-PAIN, (lit. grill-bread, 'toaster'), PERCE-OREILLE, (lit. pierce-ear, 'earwig'), TOURNEVIS, (lit. turn-screw, 'screwdriver'), or LÈCHE-VITRINE, (lit. lick-window, 'window-shopping'), can be captured by the formal and theoretical apparatus of HPSG. Contrary to the view what has often expressed in the pastbeen said, we argue that VN lexemes formation comes under is subject to morphological constraints rather than to but not under syntactic mechanisms. Our analysis integrates VN lexemes into a multiple-dimensional typed- hierarchy of lexemes and provides an account for of semantic generalizations involved in different types of lexeme formation (compounding, derivation, and conversion).

Morphological compounding is a mechanism of lexeme formation that has been studied less within HPSG compared to derivational and inflexional phenomena. In this paper, we propose a morphological treatment of French Verb-Noun compounds (as in 1), which have been frequently considered as lexicalized syntactic phrases in the literature. We present an HPSG analysis[1] that integrates compounding in a general lexeme typed-hierarchy, and captures some generalities about the semantics needed in most deverbal lexeme formations, in particular, in VN lexemes, derived lexemes, and convert lexemes.

(1) a. GRILLE-PAIN[2] (lit. grill-bread, 'toaster')
b. PERCE-OREILLE (lit. pierce-ear, 'earwig')
c. TOURNEVIS (lit. turn-screw, 'screwdriver')
d. LÈCHE-VITRINE (lit. lick-window 'window-shopping')

1 Why VN compounds are not syntactic formations

In the literature, French VN compounds are commonly considered as syntactic formations (Di-Sciullo & Williams 1987, Barbaud 1994, Lieber 1992, Zwanenburg 1992, among others). However, following Corbin (1992), Fradin (2005) and Villoing (2009), we argue that VN compounds do not

[1] We would like to thank the members of the 'Lectures in HPSG' seminar at Paris-Diderot University, specially Anne Abeillé, Olivier Bonami, Danièle Godard, Bernard Fradin and Françoise Kerleroux for their remarks and substantive suggestions; we would also like to thank the members of the HPSG'09 conference for their interesting questions.

[2] By convention, lexemes are in small capitals.

show the properties expected of lexicalized syntactic phrases, a fact which argues in favor of the idea that they are formed morphologically rather than syntactically.

First, compounding as the morphological formation of lexemes does not typically involve functional words. Determiners, prepositions, and pronouns (including clitic forms, which are inflexional forms, see Miller & Sag 1997) never realize in a compound. Remarkably, the type of nouns selected by the VN compounding rule always appears with a determiner in the corresponding sentence (cf. 2), while determiners never realize with the N in VN nominals, cf. (3).

(2) Cet objet grille **le** pain./*Cet objet grille pain.
lit. this object grill **the** bread. / this object grill bread

(3) GRILLE-PAIN$_{\mathbf{N}}$ (lit. grill-bread, 'toaster')
Cet objet est un grille-pain / *Cet objet est un grille-**le**-pain.
lit. this object is a grill-bread. / *This object is a grill **the** bread

In contrast, syntactic lexicalization of verb phrases, including those involving the same categories (V and N) as VN compounds (cf. 5), do characteristically preserve functional words of the original syntactic phrase, including prepositions (4a), pronouns (4b,c), and determiners (5):

(4) a. BOIT-**SANS**-SOIF$_{\mathbf{N}}$ (lit. drinks-without-thirst, 'drunkard')
b. RENDEZ-**VOUS**$_{\mathbf{N}}$ (lit. go-you, 'appointment')
c. SOT-**L**'**Y**-LAISSE $_{\mathbf{N}}$ (lit. silly-it-there-leaves, 'chicken oyster')

(5) a. TROMPE **L**'OEIL$_{\mathbf{N}}$ (lit. deceives-the-eye, 'trompe l'oeil')
b. TROMPE-**LA**-MORT$_{\mathbf{N}}$ (lit. deceives-the-death, 'daredevil')
c. CRÈVE-**LA**-FAIM$_{\mathbf{N}}$ (lit.dies-the-hunger,'beggar, destitute person')
d. PUE-**LA**-SUEUR $_{\mathbf{N}}$ (lit. stinks the perspiration, 'poor laborer')

In addition, if VN compounds were lexicalized phrases, inherent reflexive pronouns that are obligatorily realized in the syntax would be expected to appear. However, this is not the case, as demonstrated in (6):

(6) a. Il **se** casse la tête. (lit. he $_{\text{REFLX}}$ breaks the head)
b. C'est un casse-tête. (lit. it's a break-head)
c. *C'est un **se** casse-tête (lit. it's a $_{\text{REFLEX}}$ break-head)

Second, lexicalized phrases preserve in their structure the original SVO word order of the source sentence, as in (7). In light of this property, it is interesting to observe that many Verb-Noun compounds cannot appear as such in a sentence, specifically because the N realized on the right of the verb does not satisfy the syntactic constraints on the realization of the semantic arguments of the verb. The N of the VN lexemes in (8), for example, is understood as an agent and would be realized in a sentence as a subject on the left of the verb.

(7) a. **JE** NE SAIS QUOI$_N$ (lit. I don't know what, 'something')
b. **JE** SAIS TOUT$_N$ (lit. I know all, 'smart-aleck' 'know all')

(8) a. HURLE-**LOUP**$_N$ (lit. howl-wolf, toponym)
b. GOBE-**MOUTON**$_N$ (lit. swallow sheep, kind of poisonous plant)
c. PISSE-**CHIEN**$_N$ (lit. pee-dog, 'type of plant')

In lexeme compounds, the semantic relations between the verb and noun is not absolutely uniform, nor as predictable as it would be in a syntactic structure. While the N in a VN compound most frequently denotes the patient of the verb (cf. 9a), it can also denote other roles, such as the agent (cf. 8), spatial localization (cf.9b), or temporal localization (cf. 9c):

(9) a. OUVRE-**BOÎTE**$_N$ (lit. open-tin, 'tin opener')
b. TRAÎNE-**BUISSON**$_N$ (lit. hang around on-bush, 'animal')
c. REVEILLE-**MATIN**$_N$ (it. wake up-morning, 'alarm clock')

In fact, this relative plasticity of argument interpretation is a characteristic of lexeme compounding, and contrasts with the limited range of interpretation exhibited by the argument structure of a given verb in a sentence. As another illustration of this phenomenon, we observe that the resulting interpretation of a VN compound may also vary, even for a given verb-noun semantic relation. For example, among the patient relations in (10), VN (10a) denotes a patient, VN (10b) denotes an event, and VN (10c) an agent (and others may denote an instrument, or a localization).

(10) a. GOBE-MOUTON$_N$ (lit. swallow-sheep, 'poisonous plant')
b. LECHE-VITRINE$_N$ (lit. lick-window, 'window shopping')
c. GRATTE-PAPIER$_N$ (lit. scratch-paper, 'pen pusher')

As VN compounds do not exhibit syntactic constraints that are preserved in lexicalized phrases, we conclude that these compounds are morphological constructs in French (Corbin 1992, Villoing 2003, Fradin 2005).

2 The lexeme properties of VN compounds

2.1 General properties

Verb-Noun compounds are nominals. As morphological constructs, they are formed of two lexemes: a verbal base-lexeme and a nominal base-lexeme. The semantics of the whole compound (S3, in Table 1) involves the semantics of the base-lexemes AND the semantics of the morphological rule.

LEXEME 1	LEXEME 2		LEXEME 3
F1: x	F2: y	⇔	F3: xy
Cat 1: V	Cat 2: N		Cat 3: N
S1	S2		S3

Table (1): The morphological French VN lexeme compound formation pattern, where F = phonology; Cat = syntactic category; S = semantics

We observe that the VN rule has two possible semantic outputs: event-denoting nominals as in (11) or object-denoting nominals as in (12); the latter may denote humans as in (12a), instruments as in (12b), or spatial localizations as in (12c).

(11) (*faire du*) LECHE-VITRINE$_N$ (lit. (to do some) lick-window, 'window shopping')

(12) a. GRATTE-PAPIER$_N$ (lit. scratch-paper, 'pen pusher')
b. GRILLE-PAIN$_N$ (lit. grill-bread, 'toaster')
c. COUPE-GORGE$_N$ (lit. cut-throat, 'dangerous back alley')

2.2 VN compounds: a property of Romance languages

VN compounding is characteristic of Romance languages (see examples in Italian (13) and Spanish (14)). This process is much less productive in Germanic languages, which typically employs another compounding process, the so-called "synthetic compounding", combining two nouns, the second of which is deverbal ([NV-er]$_N$; cf. examples in English (15) and Dutch (16)).

(13) a. SPREMI$_V$-LIMONI$_N$ (lit. press-lemon, 'lemon squeezer')
b. ROMPI$_V$COLLO$_N$ (lit. break-neck, 'daredevil')

(14) a. LANZA$_V$COHETES$_N$ (lit. throw-rocket, 'rocket launcher')
b. COME$_V$CURAS$_N$ (lit. eat-priest, 'anticlerical')

(15) a. TRUCK$_N$-DRIVER$_N$
b. DISH$_N$-WASHER$_N$
c. WHALE$_N$-HUNTING$_N$

(16) a. BRAND$_N$-BLUSSER$_N$ (lit. fire-extinguisher, 'extinguisher')
b. GIF$_N$-MENGER$_N$ (lit. poison-mixer, 'poisoner')
c. GRAPPEN$_N$-MAKER$_N$ (lit. jokes-maker, 'comedian')

2.3 VN compounds as word forms

As word forms, VN compounds have all the expected syntactic functions of nominals. They can function as be objects , such as in (16a), or subjects, such as in (16b).

(16) a. Paul a acheté un **grille-pain**. (Obj: Spec+N)
lit. Paul bought a grill-bread ('toaster')

b. Le **grille-pain** est cassé. (Subj: Spec+N)
lit. The grill-bread ('toaster') is broken.

The same properties are observed for event-denoting VNs, as in (17a-c):

(17) a. Le **lèche-vitrine** est mon loisir préféré (Subj: Spec+N)
lit. the window-shopping is my favorite hobby
b. Marie adore le **lèche-vitrine** (Obj: Spec+N)
lit. Mary loves window-shopping
c. Lola fait du **lèche-vitrine** (Obj : Spec indef +N)
lit. Lola goes window-shopping

The semantics of the compounding rule allow object-denoting VNs to be predicative (18a, b) or attributive (18c):

(18) a. Paul a acheté du papier **tue-mouche**.
lit. Paul bought some kill-fly paper ('flypaper')
b. Ce couloir a trois portes **coupe-feu**.
lit. This corridor has three cut-fire doors ('firebreak door')
c. Pierre est **rabat-joie**.
lit. Peter is reduce-joy ('spoilsport').

In this case, the modified N (*papier* in (18a), *portes* in (18b)) or the subject (*Pierre* in (18c)) is the Proto-agent of the verbal base-lexeme (in the sense of Dowty 1991). In the predicative use, the paper is seen as the killer of flies (18a) and the door as the one that cuts fire (18b). In the attributive use, Pierre, a human, is seen as the one who causes the reduction of joy (18c).

Event-denoting VNs can also have attributive or predicative uses, since French allows the construction of VNs denoting a property from an event, cf. (19) and (20). But, this is neither direct nor systematic, and requires some semantic accommodation.

(19) a. On part pour une journée **lèche-vitrine**. (web)
lit. We are going for a day window-shopping

(20) Il est très **baise-main**.
lit. He is very kiss-hand

2.4 The phonological properties of VN compounds

Many discussions have focused on the nature of the verb in VN compounds, especially on the question whether it is a stem or a word-form (see Villoing 1999 for an overview). Since VN verb forms are not marked for inflection, we consider them stems. As for their phonological properties, we follow

Bonami & Boyé's (2003) account of verbal inflection in French. In their perspective (following, among others, Aronoff 1994 for Latin conjugation), verbal lexemes are associated in the lexicon with a vector of different possible phonological representations. These phonological representations are distinct stems, which Bonami & Boyé (2003) call the "stem space".

Lexeme	**Stem 1** (PRST. SG)	**Stem 2** (PRST.3.PL)	**Stem 3** (PRST. 1/2 PL IMPARF.)
BOIRE 'to drink'	/bwa/	/bwav/	/byv/

Table (2): The phonological verb stem of the verb BOIRE

Each lexical or inflectional morphological rule selects for a specific stem as input. From the possible stems of the verb, the VN compounding rule always selects for stem 1. The verb lexeme SOUTENIR, for example, has at least two stems /s u t ə n/ and /s u t ʒ ɛ̃/; the rule selects for the first, which is also used to form the present singular.

Verb lexeme	**Stem 1** (PRST SG)	**VN compound**	
COUPER 'to cut'	/k u p/	COUPE-PAPIER$_N$	lit. cut-paper, 'paper knife'
LECHER 'to lick'	/l ɛ ʃ/	LÈCHE-VITRINE$_N$	lit.lick-window, 'window-shopping'
ESSUYER 'to wipe'	/ɛ s ɥ i/	ESSUIE-GLACE$_N$	lit.wipe-window, 'windshield wiper'
OUVRIR 'to open'	/u v r/	OUVRE-BOÎTE$_N$	lit. open-tin, 'tin opener'
SOUTENIR 'to support'	/s u t j ɛ̃/	SOUTIEN-GORGE$_N$	lit.support-bosom, 'bra'
TORDRE 'to wring'	/t ɔ r/	TORD-BOYAU$_N$	lit.wring-gut, 'rotgut'

Table (3): The phonological verb stem of VN compounds

The first stem is the default phonological stem for all verbs involved in the VN compounding rule, while the default stem for derivational rules is commonly stem 3, used for the present plural or for the perfect tense.

Verb lexeme	**Stem 1** (PRST SG)	**VN compound**	**Stem 3** (PERFECT)	**Deverbal nouns**
ESSUYER 'to wipe'	/e s ɥ i/	ESSUIE-GLACE$_{N}$	/e s ɥ i j/	ESSUYAGE 'drying up' ESSUYEUR 'dryer'
SOUTENIR 'to support'	/s u t j ɛ̃/	SOUTIEN-GORGE$_{N}$	/s u t ə n/	SOUTENABLE 'bearable' SOUTENANCE 'academic defense'
TORDRE 'to wring'	/t ɔ r /	TORD-BOYAU$_{N}$	/t ɔ r d/	TORDABLE '°wringable' TORDEUR 'wringer'

Table (4): The phonological verb stem of VN compounds and deverbal nouns

The noun can, in most cases, be analyzed as a stem, but may sometimes look like a word form marked for plural:

(21) a. ESSUIE-**MAINS**$_{N}$ (lit. dry-hands, 'hand towel')
b. PRESSE-**FRUITS**$_{N}$ (lit. press-fruits, 'squeeze')
c. PROTÈGE-**YEUX**$_{N}$ (lit. protect-eyes, 'eye mask')

We believe that this is not syntactic marking, but an inherent inflection (such as described by Booij, 1996). Inherent inflection is required by the semantics and not by the syntax. The choice of singular or plural marking by the rule does not really change the semantics of the whole VN.

2.5 The semantic properties of VN

2.5.1 The semantics of the verbal base-lexeme

The verbal base-lexeme of a VN is dynamic (following Vendler 1967 and Dowty 1979). Stative verbs are, therefore, bad candidates for VN compounding:

(22) a. ?? Paul est un véritable **sait**-latin.
lit. Paul is a true know-Latin
b. ?? Le Béluga, les **aime**-caviar russes en sont fous.
approx. The Beluga, Russian love-caviars are crazy about it

Most verb bases are transitive and present an agent/patient relation. Therefore, unaccusative verbs (23) and unergative verbs (24) are typically bad candidates as well:

(23) a. ?? °**ARRIVE**-TRAIN$_{N}$ (lit. arrive-train)
b. ?? °**TOMBE**-PLUIE$_{N}$ (lit. fall-rain)

(24) a. ?? °**ABOIE**-CHIEN$_N$ (lit. bark-dog)
b. ?? °**JONGLE**-CLOWN$_N$ (lit. juggle-clown)

Nevertheless, some VN compounds are built on unaccusative or unergative verb bases; in this case, the interpretation is causative, through an agent participant variable added by the rule (see 37-40 below).

2.5.2 The semantics of the nominal base-lexeme

In most cases, the semantic role of the noun-base is the Proto-patient[3] argument of the verbal base-lexeme, as in (25). So, the noun base denotes what is affected by the process described by the verb. In rare cases, it may be understood as another argument: agent, spatial or temporal localization. All the possibilities are summed up in Table (5).

(25) a. LÈCHE**-VITRINE**$_N$ (lit. lick-window, 'window-shopping')
b. OUVRE**-BOÎTE**$_N$ (lit. open-tin, 'tin opener')
c. GRATTE**-PAPIER**$_N$ (lit. scratch-paper, 'pen pusher')
d. COUPE**-GORGE**$_N$ (lit. cut-throat, 'dangerous back alley')
e. TROTTE-**BÉBÉ**$_N$ (lit. toddle-along-baby, 'baby walker')

Patient[4]	Agent	Location	Temporal
LÈCHE-VITRINE **OUVRE-BOÎTE** **GRATTE-PAPIER** **COUPE-GORGE**	GOBE-MOUTON HURLE-LOUP PISSE-VACHE TROTTE-BÉBÉ	TRAÎNE-BUISSON	RÉVEILLE-MATIN

Table (5): The semantic role of the N in a VN compound

2.5.3 Semantic properties of the whole VN

As we said above, VN compounding has two possible types of semantic output: event-denoting nominals and object-denoting nominals. Event-denoting nominals, as in (26), denote a subset of events:

(26) a. LECHE-VITRINE$_N$ (lit. lick-window, 'window-shopping')
b. REMUE-MENAGE$_N$ (lit. move-household, 'commotion')

---

[3] The Proto-patient, as well as the Proto-agent, are defined according to the criteria given by Dowty (1991) and Davis & Koenig (2000)
[4] Boldface indicates the most common interpretation.

Object-denoting nominals are obtained from two different types of semantic rules.

• The first rule operates the abstraction of a participant variable of the verbal base-lexeme. In this case, VN compounds are mostly interpreted as the Proto-agent, as in (27):

(27) a. OUVRE-BOITE$_N$ (lit. open-tin, 'tin opener')
b. REVEILLE-MATIN$_N$ (lit.wake up-morning, 'alarm clock')
c. GRATTE-PAPIER$_N$ (lit. scratch-paper, 'pen pusher')
d. GARDE-COTE$_N$ (lit. watch-coast, 'coastguard')

But, in a few other cases, it may also be a patient (cf. 28) or a spatial localization (cf. 29) , as noted above. Table (6) sums up the various denotation types available for object-denoting VNs that correspond to the abstraction of a variable.

(28) a. GOBE-MOUTON$_N$ (lit. swallow-sheep, 'kind of poisonous plant')
b. BROUTE-BIQUET$_N$ (lit. graze-kid (young goat), 'honeysuckle')
(29) a. COUPE-GORGE$_N$ (lit. cut-throat 'dangerous back alley')
b. HURLE-LOUP$_N$ (lit. howl-wolf, 'toponym')

(proto)Agent	**(proto)Patient**	**Location**
GRATTE-PAPIER	GOBE-MOUTON	HURLE-LOUP
TRAÎNE-SAVATE	BROUTE-BIQUET	COUPE-GORGE
OUVRE-BOÎTE		GARDE-MEUBLE.
RÉVEILLE-MATIN		

Table (6): The semantic role of VN compounds (select a participant)

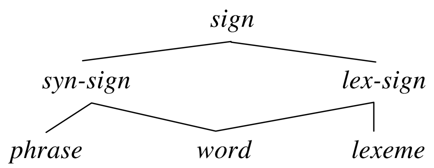
• The second semantic rule involved in object-denoting VNs adds an agent participant variable to the verbal base-lexeme relation, *via* a causative relation in the case of a non-agentive verb base-lexeme (inaccusative), as in (30), or an instrumental relation in the case of unergative verbal bases, as in (31):

(30) a. COULE-SANG$_N$ (lit. flow-blood, 'type of plant')
b. SAUTE-BOUCHON$_N$ (lit. jump-cork, 'champagne')
(31) a. PISSE-CHIEN$_N$ (lit. pee-dog, 'type of plant')
b. TROTTE-BÉBÉ$_N$ (lit.toddle-along-baby, 'baby walker')

3 Analysis

3.1. A type hierarchy for morphologically complex lexemes

In Bonami & Boyer's (2006) *sign* type hierarchy that we choose to use[5], lexemic properties of a *lex-sign* (i.e., words and lexemes) are expressed *via* the attribute MORPHOLOGICAL-DAUGHTERS. This allows us to express that a word is a syntactic sign with a lexemic identity. This hierarchy also allows the distinction between words and lexeme signs, stipulating that PHONOLOGY is an attribute of syntactic signs (i.e. phrases and words), whereas the phonological identity of *lexeme* signs is expressed *via* the STEMS feature (see table (7) below).



Fig(1). Bonami & Boyé's (2006) *sign* type hierarchy

TYPE	CONSTRAINT	ISA
syn-sign	[PHON *phon*]	*sign*
lex-sign	[M-DTRS *list*(*lexeme*)]	*sign*
phrase	[DTRS *list*(*syn-sign*)]	*syn-sign*
word	[M-DTRS <*lexeme*>]	*syn-sign & lex-sign*
lexeme	[STEMS *stem-space*]	*lex-sign*

Table (7): Constraints on the *sign* type hierarchy

The lexicon of languages builds lexemes by different means; this includes a widespread distinction (in French, as in other European languages) between simple lexemes (*simplex*) and morphologically complex ones. We propose to account for this variety of organization by using a further dimension of classification, called FORMATION, in addition to the PART-OF-SPEECH and VALENCE dimensions, see Fig. (2).

Lexemes with a complex morphology (*morph-complex-lex*) are classified into compound, derived and converted lexemes[6]. This analysis is based on several recent works in morphology; in particular, we integrate the results of Tribout (forthcoming) on converted lexemes, Fradin & Kerleroux (2002) and

---

[5] Our analysis could also be mapped onto the SBCG framework (Sag, 2007), considering *vn-lex* as a type of construct.

[6] We believe that inflected signs are *syntactic-signs*, hence, INFLECTION should be a dimension of *word* type hierarchy.

Kerleroux (2004) on derived lexemes with the suffix -*eur*, Namer & Villoing (2008) on lexemes with the suffix -*oir(e)*, Ferret, Soares & Villoing (2009) on lexemes with the suffix -*age*, Plénat (2005) on lexemes with the suffix -*ette*, and Roché (2003) on lexemes with suffix -*on*.

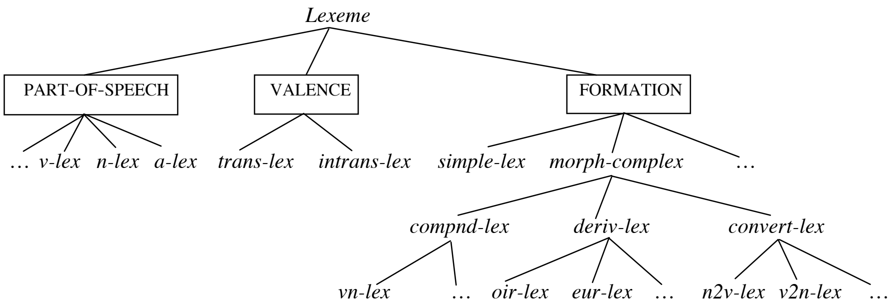


Fig (2). A multi-dimensional *lexeme* type hierarchy

3.2. Semantic rules available for deverbal lexeme formation

What emerges from these analyses is the fact that semantic rules involved in the formation of deverbal lexemes have much in common, whether these are compound, derived or converted lexemes. First, these always involve the semantic argument structure of the verb base. To account for this factor, we propose to use a type hierarchy for semantic roles *à la* Davis and Koenig (2000), as follow:

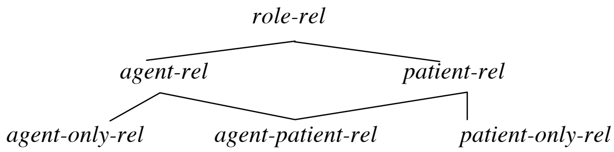


Fig (3). The *role-relation* (partial) type hierarchy

Second, two general semantic patterns are evident: deverbal lexemes may denote an event (or a set of events) or a referential index. The latter may be abstracted from the semantic argument structure of the verb base or be an additional argument. These general semantic patterns are captured in the *complex-nominal-relation* type hierarchy we propose:

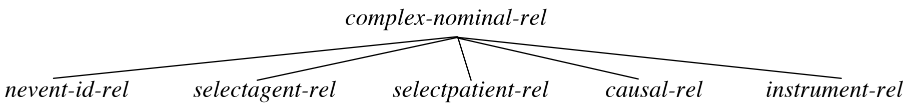


Fig (4). The *complex-nom-rel* (partial) type hierarchy

Constraints on *complex-nominal-relation,* given in (32), are rather general, since each specific lexeme imposes its own particular semantics. *Nominal-event-id-relation* is an identity relation that takes an *austinian* as an argument[7] whose SITUATION index, corresponding to the index of the verb base, is identified with the INSTANCE value of the relation. This relation applies to event nominalizations in general: in addition to VN lexemes (like LÈCHE-VITRINE 'window shopping' or SAUTE-MOUTON 'leapfrog'), it also applies to converted lexemes (NAGE 'swimming', CHUTE 'downfall'), and to derived lexemes with the suffix *-ette* (BRONZETTE 'sunbathing'), with the suffix *-age* (BALAYAGE 'sweeping') or with the suffix *-on* (PLONGEON 'dive').

(32) *nevent-id-rel* => $\begin{bmatrix} \text{INST} & [1] \\ \text{ARG} & \textit{austinian}[\text{SIT}\ [1]] \end{bmatrix}$

selectagent-rel => $\begin{bmatrix} \text{INST} & [1] \\ \text{ARG} & \mid \text{NUCL}\ [\text{AGT}\ [1]] \end{bmatrix}$

selectpatient-rel => $\begin{bmatrix} \text{INST} & [1] \\ \text{ARG} & \mid \text{NUCL}\ [\text{PAT}\ [1]] \end{bmatrix}$

causal-rel => $\begin{bmatrix} \text{INST} & [1] \\ \text{ARG} & \textit{austinian} \end{bmatrix}$

instrument-rel => $\begin{bmatrix} \text{INST} & [1] \\ \text{ARG} & \mid \text{NUCL}\ [\text{AGT}\ [2]] \end{bmatrix}$

Selectagent-rel, selectpatient-rel, and *selectloc-rel* are relations in which a particular semantic argument is abstracted from a verb base relation. It may denote an agent, as in the VN lexemes GRATTE-PAPIER ('pen pusher'*),* GARDE-BARRIÈRE ('gate keeper'), and in derived lexemes with the suffix *-eur* (MARCHEUR 'walker', CHANTEUR 'singer'), with the suffix *-on* (GROGNON 'grumbling', BROUILLON 'draft'), or in converted lexemes (JUGE 'judge', GARDE 'guard'). It may denote a patient, as in the VN lexeme GOBE-MOUTON ('kind of poisonous plant'), in derived lexemes with the suffix *-oir* (TIROIR 'drawer'), with the suffix *-ette* (SUCETTE 'lollypop'), with the suffix *-on* (NOURRISSON 'infant', SUÇON 'hickey'), or in converted lexemes (AFFICHE 'poster', PARCOURS 'route'). The abstracted semantic argument may also denote the localization of an event (or a set of events), as in the VN lexeme GARDE-MEUBLE ('storage'), in derived lexemes with the suffix *-oir* (LAVOIR 'wash house', FUMOIR 'smocking room'), in lexemes with the suffix *-ette* (BUVETTE 'taproom', CACHETTE 'hiding place'), or in converted lexemes (INSTITUT 'institute', ARRIVÉE 'arrival').

The *causal-relation* and *instrumental-relation* are mostly used in cases where an 'external' agent is added to the argument structure of a verb base. The

---

7 We borrow the *austinian* type from Ginzburg and Sag (2000).

causal-relation adds a cause argument, mostly to a *patient-only* type of verb base, like the VN lexeme COULE-SANG ('plant', lit. leek–blood). The *instrumental-relation* adds an argument understood as an instrument or a mean to a verb base that has an *agent-rel* type of relation, like the VN lexeme TROTTE-BÉBÉ ('baby walker'), or in derived lexemes with the suffix *-oire* (PASSOIRE 'strainer', MACHOIRE 'jaw'), with the suffix *-eur* (CHARGEUR 'cartridge', INTERRUPTEUR 'switch'), with the suffix *-ette* (ALLUMETTE 'matchstick'), or with the suffix *-on* (GUIDON 'handlebar', TORCHON 'dishcloth').

3.3. A type hierarchy for VN compounds

As expected, the first partition of the *vn-lexeme* type hierarchy we propose is between event-denoting nouns (*nevent-vn-lex*) and object denoting nouns (*nobj-vn-lex*). There are three subtypes of *nobj-vn-lexeme*: *agent-vn-lex*, *patient-vn-lex* and *localization-vn-lex*. Many VN lexemes are of the general *agent-vn-lex* type (that selects the agent argument of a transitive verb base), and some other *agent-vn-lex* are built by adding a cause or an instrument argument to the argument structure of the verb base.

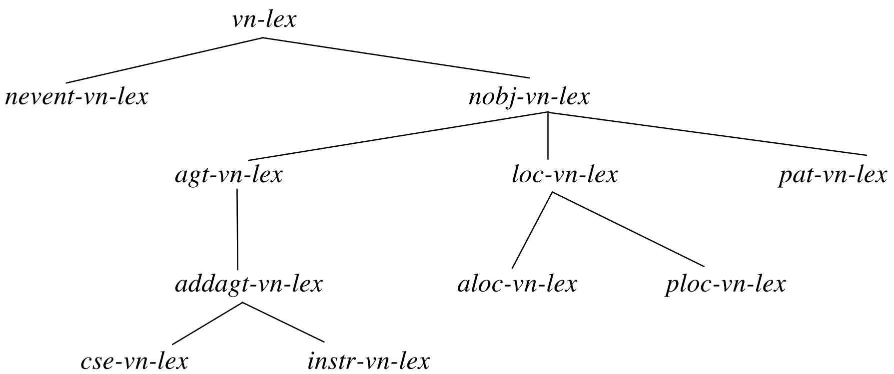


Fig (5). The *vn-lexeme* type hierarchy

The following table lists these subtypes with examples of VN lexemes:

nevent-vn-lex	*agt-vn-lex*	*pat-vn-lex*	*aloc-vn-lex*	*ploc-vn-lex*	*cse-vn-lex*	*instr-vn-lex*
LECHE-VITRINE	GRATTE-PAPIER GRILLE-PAIN	GOBE- MOUTON	HURLE-LOUP	GARDE-MEUBLE	COULE-SANG	TROTTE-BEBE

Table (9). Illustration of the different types of VN lexemes

Before looking at the detailed constraints on *vn-lex*, we must discuss the fact that the inheritance principled hierarchy allows us to adequately express the repartition between common and particular properties among the different VN lexemes we analyzed. And yet, the descriptive generalities hierarchically ordered (as in fig. 5) fail to match the intuitive or desired picture according to which the most productive VN types should be ranked higher than the less productive ones. Indeed, considering the productivity of these lexemes, *agent-vn-lex* should be the highest super type, or be the default type. However, that would lead to problems of descriptive congruency in the representation of other types (*nvent-vn-lex*, *loc-vn-lex* and *pat-vn-lex*) as subtypes. A possible solution, to be explored in later work, would be to include a PRODUCTIVITY feature with a variable of weight as value, that would integrate results obtained from a robust corpus study of VN lexeme productivity, based on Baayen's (1992 and 2008) methods.

Constraints associated with the different *vn-lexeme* subtypes integrate the general semantic rules proposed earlier. The constraints in (33) stipulate that a lexeme *vn-lexeme* is a noun with a *complex-nominal-relation* and two morphological daughters: a verbal base-lexeme and a nominal base-lexeme. The verbal base has a *dynamic-rel* semantic relation and, by default, it has a *patient-relation*, which means it minimally has a patient argument, and may possibly have an agent-patient relation. In addition, the value of its semantic argument PATIENT is, by default, the same as the INDEX of the nominal base-lexeme. Stem phonology is preceded by concatenation, in the standard way.

(33) *vn-lexeme* =>

$$\begin{bmatrix} \text{STEMS} & [\text{SLOT1 [3]} \oplus \text{[4]}] \\ \text{SYNSEM} & \begin{bmatrix} \text{CAT} & [\text{HEAD } \textit{noun}] \\ \text{CONT} & \begin{bmatrix} \text{IND [1]} \\ \text{RESTR} \left\{ \begin{bmatrix} \textit{cplx-nom-rel} \\ \text{INST [1]} \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix} \\ \text{M-DTRS} & \langle\ \textit{v-lex} \begin{bmatrix} \text{STEMS [SLOT1 [3]]} \\ \text{SS} \mid \text{CONT} \left[\text{NULC } \textit{dynamic-rel} / \begin{bmatrix} \textit{patient-rel} \\ \text{PAT [2]} \end{bmatrix}\right] \end{bmatrix}, \\ & \quad \textit{n-lex} \begin{bmatrix} \text{STEMS [SLOT1 [4]]} \\ \text{SS} \mid \text{CONT [IND [2]]} \end{bmatrix} \rangle \end{bmatrix}$$

In (34), constraints on *nominal-event-vn-lexeme* (*nevent-vn-lex*) stipulate that the value of the *austinian* ARGUMENT of its *relation* is the same as the CONTENT value of the verb base.
In (35), *nominal-object-vn-lex* (*nobj-vn-lex*) has, by default[8], an *austinian* ARGUMENT in its set of RESTRICTION, whose value is the same as the CONTENT value of the verb base.

---

8 This default is overridden by *cse-vn-lex* and *instr-vn-lex*, which needs to introduce an additional agent argument here, *via* a *causal-relation* or an *instrumental-relation* respectively,

(34) *nevent-vn-lexeme* =>

$$\begin{bmatrix} \text{CONT} & \begin{bmatrix} \text{RESTR} & \left\{ \begin{bmatrix} \textit{nevent-id-rel} & \\ \text{INST} & [1] \\ \text{ARG} & [5]\textit{austinian}[\text{SIT}\ \ [1]] \end{bmatrix} \right\} \end{bmatrix} \\ \text{M-DTRS} & < \textit{v-lex} \begin{bmatrix} \text{CONT} & [5] \end{bmatrix}, [\ ] > \end{bmatrix}$$

(35) *nobj-vn-lexeme* =>

$$\begin{bmatrix} \text{CONT}\,|\ \text{RESTR}\ \{\ [\text{ARG}\ /\ [5]\textit{austinian}\ ]\ \} \\ \text{M-DTRS}\ \ < \textit{v-lex}\ \ [\text{CONT}\ [5]\ ]\ \ > \end{bmatrix}$$

In (36), constraints on *agent-vn-lex* state it has a *selectagent-relation* in its set of restrictions.

(36) *agent-vn-lexeme* => $[\text{CONT}\,|\,\text{RESTR}\ \{\ \textit{selectagent-rel}\}\ ]$

The *addagent-vn-lex* type, in (37), is created for descriptive purposes more than for strictly formal needs, since it does not add any specification at its own level, but gives rise to two subtypes, *cse-vn-lex* (cf. 38) and *instr-vn-lex* (cf. 39). The property both these types have in common is the selection of the agent argument of an intermediary semantic relation, *causal-rel* and *instrumental-rel* respectively, that takes the CONTENT of the verb base as its second ARGUMENT.

(37) *addagent-vn-lexeme* => *agt-vn-lex*

(38) *cse-agt-vn-lex* =>

$$\begin{bmatrix} \text{CONT}\,|\,\text{RESTR} & \left\{ \begin{bmatrix} \textit{selectagent-rel} & \\ \text{INST}\ [1] & \\ \text{ARG}\,|\,\text{NUCL} & \begin{bmatrix} \textit{causal-rel} & \\ \text{AGT}\ \ [1] & \\ \text{ARG}\ \ [5] & \begin{bmatrix} \textit{pat-only-rel} \\ \text{PAT}\ \ [2] \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\} \\ \text{M-DTRS} & < \quad [\text{CONT}\ [5]\ ] \quad , [\text{IND}\ [2]\ ]\ > \end{bmatrix}$$

(39) *instr-agt-vn-lexeme* =>

$$\begin{bmatrix} \text{CONT}\,|\,\text{RESTR} & \left\{ \begin{bmatrix} \textit{selectagent-rel} & \\ \text{INST}\ [1] & \\ \text{ARG}\,|\,\text{NUCL} & \begin{bmatrix} \textit{instrumental-rel} & \\ \text{AGT}\ \ [1] & \\ \text{ARG}\ \ [5] & \begin{bmatrix} \textit{agent-rel} \\ \text{AGT}\ \ [2] \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\} \\ \text{M-DTRS} & < \quad [\text{CONT}\ [5]\ ] \quad , [\text{IND}\ [2]\ ]\ > \end{bmatrix}$$

VN lexemes that express spatial localization (*loc-vn-lex*, in 40) have either an intransitive verb base (*aloc-vn-lex*, in 41) or a transitive one (*ploc-vn-lex*, in 42):

---

and doing so, embeds the *austinian* argument that corresponds to the content of the verb base.

(40) *loc-vn-lexeme* =>

$$\left[\begin{array}{l} \text{CONT} \mid \text{RESTR} \left\{ \left[\begin{array}{ll} \multicolumn{2}{l}{\textit{selectloc-rel}} \\ \text{INST} & [1] \\ \text{ARG} \mid \text{NUCL} & [\ \text{LOC}\ [1]\ ] \end{array}\right] \right\} \end{array}\right]$$

(41) *aloc-vn-lexeme* =>

$$\left[\begin{array}{l} \text{M-DTRS} < \left[\text{CONT} \mid \text{RESTR} \left\{ \left[\text{ARG} \mid \text{NUCL} \left[\begin{array}{ll} \text{AGT} & [2] \\ \text{LOC} & [1] \end{array}\right] \right] \right\} \right], [\,] > \end{array}\right]$$

(42) *ploc-vn-lexeme* =>

$$\left[\begin{array}{l} \text{M-DTRS} < \left[\text{CONT} \mid \text{RESTR} \left\{ \left[\text{ARG} \mid \text{NUCL} \left[\begin{array}{ll} \text{AGT} & [3] \\ \text{PAT} & [2] \\ \text{LOC} & [1] \end{array}\right] \right] \right\} \right], [\,] > \end{array}\right]$$

VN lexemes that denote a patient are not productive, and not numerous. But when they correspond to a type of VN, the constraints needed are as below:
(43)
pat-vn-lexeme =>

$$\left[\begin{array}{l} \text{CONT} \mid \text{RESTR} \left\{ \left[\begin{array}{ll} \multicolumn{2}{l}{\textit{selectpat-rel}} \\ \text{INST}\ [1] & \\ \text{ARG} \mid \text{NUCL} & \left[\begin{array}{ll} \text{AGT} & [2] \\ \text{PAT} & [1] \end{array}\right] \end{array}\right] \right\} \\ \text{M-DTRS} < [\ ], [\text{IND}\ [2]\ ] > \end{array}\right]$$

The figures below illustrate different lexical entries: in (44), LÈCHE-VITRINE is a type of *vevent-vn-lex*; in (45), GRATTE-PAPIER is an *agent-vn-lex*; in (46), HURLE-LOUP is a toponym, a type of *aloc-vn-lex*; in (47), COULE-SANG is a *cse-vn-lex*; and in (48), TROTTE-BÉBÉ is an *instr-vn-lex*.

(44) *LÈCHE-VITRINE* —>

$$\left[\begin{array}{ll} \text{STEMS} & [\text{SLOT1}\ [3] \oplus [4]\ ] \\ \text{SYNSEM} & \left[\begin{array}{ll} \text{CAT} & \mid \text{HEAD}\ \textit{noun} \\ \text{CONT} & \left[\begin{array}{l} \text{IND}\ \ [1] \\ \text{RESTR} \left\{ \left[\begin{array}{ll} \multicolumn{2}{l}{\textit{lèche-vitrine-rel}} \\ \text{INST} & [1] \\ \text{ARG} & [5]\ [\text{SIT}\ [1]] \end{array}\right] \right\} \end{array}\right] \end{array}\right] \\ \text{M-DTRS} & < \textit{v-lex} \left[\begin{array}{ll} \text{STEMS} & [\ \text{SLOT1}\ \ [3]/\text{lèche}/\ ] \\ \text{SS} \mid \text{CONT}\ [5] & \left[\text{NULC}\ \ \textit{lèche-rel}\ [\text{AGT}\ \textit{index}, \text{PAT}\ [2]]\right] \end{array}\right], \\ & \quad \textit{n-lex} \left[\begin{array}{ll} \text{STEMS} & [\ \text{SLOT1}\ \ [4]/\text{vitrine}/\ ] \\ \text{SS} \mid \text{CONT} & [\text{IND}\ \ \ [2]] \end{array}\right] > \end{array}\right]$$

(45) *GRATTE-PAPIER* —>

STEMS [SLOT1 [3] $\oplus$ [4]]
SYNSEM [CAT | HEAD *noun*
CONT [IND [1]
RESTR { [*gratte-papier-rel*
INST [1]
ARG [5]] }]]
M-DTRS < *v-lex* [STEMS [SLOT1 [3]/gratte /]
SS | CONT [5] [NULC *gratte-rel* [AGT [1], PAT [2]]]],
n-lex [STEMS [SLOT1 [4]/ papier /]
SS | CONT [IND [2]]] >

(45) *HURLE-LOUP* —>

STEMS [SLOT1 [3] $\oplus$ [4]]
SYNSEM [CAT | HEAD *noun*
CONT [IND [1]
RESTR { [*hurle-loup-rel*
INST [1]
ARG [5]] }]]
M-DTRS < *v-lex* [STEMS [SLOT1 [3]/ hurle /]
SS | CONT [5] [NULC *hurle-rel* [AGT [2], LOC [1]]]],
n-lex [STEMS [SLOT1 [4]/ loup /]
SS | CONT [IND [2]]] >

(45) *COULE-SANG* —>

STEMS [SLOT1 [3] $\oplus$ [4]]
SYNSEM [CAT | HEAD *noun*
CONT [IND [1]
RESTR { [*coule-sang-rel*
INST [1]
ARG [AGT [1]
ARG [5]]] }]]
M-DTRS < *v-lex* [STEMS [SLOT1 [3]/ coule /]
SS | CONT [NULC [5]*coule-rel* [PAT [2]]]],
n-lex [STEMS [SLOT1 [4]/ sang /]
SS | CONT [IND [2]]] >

(46) *TROTTE-BÉBÉ* —>

$$
\begin{bmatrix}
\text{STEMS} & [\text{SLOT1 } [3] \oplus [4]\,] \\
\text{SYNSEM} & \begin{bmatrix} \text{CAT} & |\ \text{HEAD } \textit{noun} \\ \text{CONT} & \begin{bmatrix} \text{IND} & [1] \\ \text{RESTR} & \left\{ \begin{bmatrix} \textit{trotte-bébé-rel} & \\ \text{INST} & [1] \\ \text{ARG} & \begin{bmatrix} \text{AGT} & [1] \\ \text{ARG} & [5] \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix} \\
\text{M-DTRS} & < \textit{v-lex} \begin{bmatrix} \text{STEMS} & [\text{ SLOT1 } [3]/\text{ trotte }/\,] \\ \text{SS}\,|\,\text{CONT} & [\text{NULC } [5]\ \textit{trotte-rel}\ [\text{ AGT } [2]\,]\,] \end{bmatrix}, \\
 & \textit{n-lex} \begin{bmatrix} \text{STEMS} & [\text{ SLOT1 } [4]/\text{ bébé }/\,] \\ \text{SS}\ |\ \text{CONT} & [\text{IND } [2]] \end{bmatrix} >
\end{bmatrix}
$$

4 Conclusion

We have presented here a formalized account of French Verb-Noun compounds, in line with the morphological analysis proposed in Villoing (2009). Our analysis integrates *vn-lexeme* types into the general *lexeme* typed-hierarchy, under a FORMATION dimension that allows the expression of a general classification among lexemes. We expect that the question of lexeme productivity may be solved by the integration of specific features into lexeme entries, as the result of a corpus study of VN productivity modeled on the methods of Baayen (2008). Moreover, we have shown that the fact HPSG allows semantics to be encoded as an independent resource is an advantage in capturing the general semantic patterns that are involved in the formation of several (de)verbal lexemes. In fact, there are other systematic lexical variations, which do not come under morphology, that also involve some of the general semantic types of relations we propose here. The very productive inchoative/causative verb pattern (TO INCREASE intrans/trans), for example, involves the *causal-relation*. Consequently, it is worth considering semantics as a *lexical-sign* dimension of classification in itself, as a way to encode in the hierarchy the fact that some semantic relations are lexically productive rules, available both for words and lexemes.

References

Anderson, S. R. 1992. *A-Morphous Morphology*, Cambridge, Cambridge University Press.

Aronoff, M. 1994. *Morphology by itself*, Cambridge, MA, The MIT Press

Aronoff, M. & Fudeman, K. 2004. *What is Morphology?*, Blackwell Publishing- Fundamentals of Linguistics.

Baayen, H. R. 1992. “Quantitative Aspects of Morphological Productivity”. *Yearbook of Morphology* 1991, 109-149.

Baayen, H. R. 2008. “Corpus linguistics in morphology: Morphological productivity”. In Lüdeling A. & Kytö M. eds, *Corpus Linguistics. An International Handbook*, Berlin, New York, Mouton de Gruyter, pp. 899–919

Barbaud, P. 1994. « Conversion syntaxique ». *Lingvisticæ Investigationes*, XVIII (1), pp. 1-26.

Beard, R. 1995. *Lexeme-Morpheme Base Morphology*. SUNY Press, Albany, NY.

Booij, G. 1986. “Form and Meaning in Morphology: the case of Dutch "agent-nouns”. *Linguistics* 24:503-518.

Booij, G. 2005, *The Grammar of Words: An Introduction to Linguistic Morphology*, Oxford, Textbooks in Linguistics.

Bonami, O. & Boyé, G. 2003. “La construction des paradigmes”. *Journée de Morphologie Non-Conventionnelle*, Toulouse, Décembre 2003.

Bonami, O. & Boyé, G. 2006. Deriving inflectional irregularity. *Proceedings of the 13th International Connference on HPSG*. Stanford: CSLI Publications.

Carstairs-McCarthy, A. 1992. *Current Morphology*. Routledge.

Corbin, D. 1992. “Hypothèses sur les frontières de la composition nominale”, *Cahiers de grammaire*, 17, pp. 26-55.

Davis, A. R. & Koenig, J.-P. 2000. “Linking as constraints on word classes in a hierachical lexicon”. *Language* 76:56-109.

Di Sciullo, A-M. & Williams E. 1987. *On the definition of Word*. MIT Press, Cambridge.

Dowty, D. 1979. *Word Meaning and Montague Grammar*. Reidel Publishing Company, Dordrecht.

Dowty, D. 1991. « Thematic Proto-roles and Argument Selection ». *Language*, 67(3), pp. 547-619.

Ferret, K., Soare, E. & Villoing, F. 2009. “Noms en *–age* et noms en *–ée*: un autre cas de compétition affixale?”, *JENOM* 2, 18-19 juin 2009, Université de Lille.

Fradin, B. 2003. *Nouvelles approches en morphologie*, Paris, Puf.

Fradin, B. 2005. On a semantically grounded difference between derivation and compounding'. In Morphology and its Demarcations, Dressler W. U., D. Katovsky & F. Rainer (eds) Amsterdam / Philadelphia: John Benjamins.

Fradin B. & Kerleroux, F. 2003. “Troubles with lexemes”, in *Proceedings of the 3rd Mediterranean Meeting of Morphology*, G. Booij, J. De Cesaris, S. Scalise, & A. Ralli eds., Barcelona, IULA-Universitat Pompeu Fabra ;

Haspelmath, M. 2002. *Understanding Morphology*, Londres, Arnold.

Kerleroux, F. 2004. “Sur quels objets portent les opérations morphologiques de construction ?”, *Lexique* 16, 85-124.

Koenig, J.-P. 1999. *Lexical Relations*. CSLI publications: Stanford.

Lieber, R. 1992. *Deconstructing Morphology : Word Formation in Syntactic Theory*. The University of Chicago Press, Chicago/London.

Matthews, P.H. 1974 [reedition 1991]. *Morphology, an introduction to the theory of word-structure*, Cambridge, Cambridge University Press.

Miller, P. H. and Sag, I. A. 1997. French clitic movement without clitics or movement. *Natural Language and Linguistic Theory* 15.3: 573-639.

Namer, F. & Villoing, F. 2008. “Interpréter les noms déverbaux : quelle relation avec la structure argumentale du verbe de base ? Le cas des noms en *–oir* du français”, *Congrès Mondial de Linguistique Française*, Paris, 9-12 juillet 2008.

Orgun Cemil Orhan & Sharon Inkelas. 2002. “Reconsidering bracket erasure”, *Yearbook of Morphology* 2001:115-146.

Plag, I. 1999. *Morphological Productivity: Structural constraints in English Derivation*. Berlin and New York: Mouton de Gruyter.

Plénat, M. 2005. “Brèves remarques sur les déverbaux en *–ette*”, in Lambert & H. Nølke (eds), *La syntaxe au coeur de la grammaire*. Recueil offert en hommage pour le 60e anniversaire de Claude Muller, Rennes, Presses Universitaires de Rennes, 245-258.

Rieheman, S. 1998. “Type-Based Derivational Morphology”. *Journal of Comparative Germanic Linguistics 2*: 49-77.

Roché, M. 2003. “De la “bonne formation” des dérivés en *–on*”, *Cahiers de Grammaire* 28, 91–112.

Sag, I. A. 2007. *Sign-based Construction Grammar: An Informal Synopsis*. Unpublished ms., Stanford University.

Tribout, D. (forthcoming). *La conversion en français*. PhD Dissertation. University Paris Diderot

Vendler, Z. 1967. *Linguistics in Philosophy*. Holt, Rinehart & Winston, Ithaca, New York.

Villoing, F. 2003. “Les mots composés VN du français : arguments en faveur d'une construction morphologique”, *Cahiers de Grammaire* 28 “Morphologie et Lexique”, pp. 183-196.

Villoing, F. 2009. “Les mots composés VN”, *Aperçus de morphologie du français*, B. Fradin, F. Kerleroux, M. Plénat (éds), Saint-Denis, Presses Universitaires de Vincennes, pp. 175-197.

Zwanenburg, W. 1992. “Compounding in French”, *Rivista di Linguistica* 4-1: 221-240.

Inflectional Morphology in Turkish VP Coordination

Antske Fokkens, Laurie Poulson and Emily M. Bender

Universität des Saarlandes and University of Washington

Proceedings of the HPSG09 Conference

Georg-August-Universität Göttingen, Germany

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

We address three properties of Turkish morphology and VP coordination: the identification of tense and aspect values across conjuncts, the optional omission of affixes on non-final conjuncts coordinated with the word *ve* and the obligatory sharing of scopal modals across conjuncts in coordination structures with the affix *-ip*. For the modals in an *-ip* structure, we propose an analysis that uses syntactic features to trigger the application of a construction at the level of the coordinated VP introducing the scopal predications. Our analysis is implemented in a small HPSG grammar and tested against datasets confirming the functionality and consistency of the analysis.

1 Introduction

This paper presents an analysis of the interaction between verbal morphology and VP coordination in Turkish. There are three properties of Turkish VP coordination of particular interest: the identification of tense, aspect and modality values across the conjuncts, the phenomenon of suspended affixation wherein affixes may be dropped from earlier conjuncts, and a coordination structure that seems to require an analysis in terms of phrasal affixes and thus seems to challenge the notion of lexical integrity. This phrasal affixation is illustrated in example (1), where the meaning of the sentence, with *-mEli* 'must' taking wide scope over the coordination, seems to suggest that *-mEli* is attached to the whole coordinated VP.

(1)	Çocuk-lar	film	izle-yip	pizza	ye-meli-ler.
	child-PL	movie	watch-COORD	pizza	eat-NEC-3PL
	"The children must watch a movie and eat pizza."				

This paper is also an example of grammar engineering for linguistic hypothesis testing (Bender, 2008), in the sense that we have built a grammar fragment for Turkish that encodes our analyses and verified its behavior over a group of testsuites. These testsuites contain 163 examples, including 96 culled from the literature and an additional 67 we developed and checked with 2-5 native-speaker consultants. The grammar was developed on the basis of the LinGO Grammar Matrix customization system (Bender et al., 2002; Bender and Flickinger, 2005; Drellishak and Bender, 2005),[1] and both the grammar and the testsuites are available for download.[2] Consistent with other Matrix-derived grammars, our grammar

[†]We would like to thank Cagatay Demiralp, Engin Ural and Huseyin Mergan, as well as two additional consultants for their help with the data and their patience. Anonymous reviewers and the audience at the HPSG 2009 conference provided useful comments, which helped to improve this paper. Naturally, all remaining errors are our own. We also thank the IRTG and PIRE for funding a two month stay at the University of Washington. This material is based upon work supported by the National Science Foundation under Grant No. 0644097. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

[1]http://www.delph-in.net/matrix/customize/matrix.cgi

[2]http://www.delph-in.net/matrix/turkish

fragment for Turkish produces semantic representations in the format of Minimal Recursion Semantics (MRS, Copestake et al. 2005) and is compatible with the LKB (Copestake, 2002).

This paper tests the following hypotheses:

(i) Obligatory matching of tense and aspect between VP conjuncts can be modeled through structure sharing of features on the event variable.

(ii) The same structure-sharing plus a lexical rule licensing the partially-inflected forms and additional constraints on the coordination rules can account for most suspended affixation facts.

(iii) When scopal affixes (of necessity and ability) are shared among conjuncts, a constructional account along the lines of Tseng 2003 can resolve the apparent violation of lexical integrity.

(iv) The above hypotheses can be implemented in a mutually consistent fashion, which is furthermore consistent with analyses of word order and other phenomena required to parse the sentences in the testsuite.

§2 provides background information on verbal morphology in Turkish and the set of morphological rules we created using the Matrix customization system. §3 describes *ve* coordination, the tense and aspect matching that it requires, and the phenomenon of suspended affixation, along with our analysis of these facts. §4 describes our analysis of another coordination construction, this time marked by an affix *-ip* on the verb of each non-final conjunct. This second construction is of particular interest because it includes apparent phrasal affixes, our analysis of which is given in §4.3. §5 situates our analyses with respect to related work, including Broadwell's (2008) LFG analysis of related facts in Turkish and Tseng's (2003) analysis of apparent phrasal affixes in French.

2 Verbal Morphology in Turkish

2.1 Properties of Turkish Verbs

This section presents an overview of morphemes that may be added to the stem and presents conditions on completeness and well-formedness of the verbs in order to provide background for the analysis of suspended affixation and inflectional marking of coordination in §§3-4. The description is based on, among others, Kornfilt 1997, Lewis 1967, Sezer 2001 and Kabak 2007.

The distinction between derivational and inflectional morphemes is not clearcut in Turkish. Traditionally, morphemes that can be followed by the infinitive marker *-mEk* are considered derivational. According to this definition, Turkish has the following derivational morphemes: *-DIr/t* (causative), *-Il* (passive), *-mA*

Table 1: Inflectional Morpheme Slots

1		2		3		4
-DI	direct past	*-(i)DI*	direct past	*-(i)sE*	conditional	AGR-*k*
-sE	conditional	*-(i)sE*	conditional	*-(i)mIş*	reported past	AGR-*z*
-mIş	reported past/ present perfect	*-mIş*	reported past			
-Iyor	continuous					
-yEcEG	future					
-Ir/-Er	aorist					
-mEli	necessitive					
-mEkte	continuous					

(negation), -(*y*)*A* (abilitative) and -(*y*)*Abil* (abilitative).[3] In addition to the derivational morphemes, there are four slots that may host an inflectional morpheme. The inflectional morphemes are presented in Table 1. A finite verb must bear an inflectional marker from slot 1 and an agreement marker (slot 4). At least one inflectional marker must be phonologically overt (Kabak, 2007).[4,5]

Turkish has two paradigms of agreement markers: the *k*-paradigm which co-occurs with definite past and conditional (-*DI* and -*sE*, respectively) and the *z*-paradigm which co-occurs with all other TAM[6] morphemes. Which paradigm is used depends on the last TAM morpheme attached to the verb. §2.2 describes the morphological analysis that we obtained from the Matrix customization system and how we adapted this analysis in order to accommodate the selection of the different agreement paradigms.

2.2 Verbal Morphology with Lexical Rules

The analysis of basic Turkish morphology we propose makes use of the morphotactic infrastructure added to the Matrix customization system by O'Hara (2008), which provides implementations for some wide-spread phenomena in morphology. The grammar created with the Matrix customization system only requires minor changes for the basic morphology to work.

[3]We adopt the convention of using capital letters to represent phonemes whose realization depends on vowel or consonant harmony.

[4]Some linguists assume that secondary tense markers are hosted by an auxiliary suffix -*i*/(*y*) (see Lees 1962 and Sezer 2001, among others), though this suffix has also been analyzed as a phonological element (Erguvanlı-Taylan, 1999). Our analysis is compatible with either view.

[5]We noticed in our data that the plural morpheme does not always follow the order of the slots presented above, though we have not found mention of this in the literature. For present purposes, we assume that this variability in morpheme ordering is a morphophonological property, and we abstract away from it in our implementation; our testsuites regularize examples to follow the canonical order as presented in Table 2.1.

[6]Henceforth, the term TAM morphemes refers to all inflectional morphemes in slots 1-3.

The morphotactic infrastructure allows the grammar engineer to define multiple morphological "slots" for each stem type or set of stem types. It provides implementations for optional and obligatory morphemes that may add syntactic and semantic features to the derived form. It also allows lexical rules to require preceding slots or to force following slots, as well as to forbid other slots from appearing. These properties are enforced by binary features on the verb that are related to specific morphological slots and registered under the feature TRACK. TRACK is appropriate for lexical rules and lexical items, but not for phrases.

These binary features work as follows. If, for instance, an optional morpheme2 requires morpheme1 in order to be licensed, bare verbs will carry a feature [MORPHEME2 $-$]. The lexical rule associated with morpheme1 turns this value into $+$, which allows the (otherwise prohibited) morpheme2-rule to apply.

When filling out the Matrix customization questionnaire, we defined nine morphological slots for verbs: five slots for derivational morphemes, three slots for TAM-markers and a slot for agreement markers. In the current version of the grammar, the derivational slots are placeholders, providing only the form of the morphemes and not the associated morphosyntactic or semantic constraints. This is because the Matrix customization system does not currently support the morphosyntax and semantics of causatives or other morphemes that add predicates, nor can it handle negative affixes that are not word-final. These facts could of course be handled by extending the starter grammar. However, because most of the derivational affixes do not have an impact on our analysis, we decided to leave the implementation of these morphemes for future work. The only exception is the derivational morpheme -(*y*)*Abil*: its behavior in *-ip* coordination forms one of the main points of discussion in this paper, and we implemented our analysis of it as an extension to the grammar produced by the customization system. This analysis is discussed in detail in §4.

The morphotactic infrastructure in the customization system does provide most necessary features to implement the inflectional morphology in our verb forms. The library permits the association of features related to tense, aspect and mood as well as subject agreement on verbs. The only phenomenon that is not supported by the current customization system is the interaction of the two agreement paradigms with different inflectional morphemes. In this case, we have morphemes which fill the same obligatory slot but which interact in different ways with preceding morphemes. In order to account for the different agreement paradigms, we created two subtypes of *agreement-lexical-rule*, and distinguished them with the binary feature AGR-PARADIGM, which we added to TRACK. The morphemes in each TAM-slot have two subtypes as well: one for the so-called "true" tenses *-DI* and *-sE*, and one for the other morphemes appearing in the same slot. Rules inheriting from the former type turn AGR-PARADIGM to k, whereas rules inheriting from the latter assign it the value z. The value of AGR-PARADIGM controls which agreement rule applies.

The analysis described above ensures that the right morphology is present on independent finite verb forms. In what follows, we present two structures that correspond to VP coordination in English. In these structures, the morphological requirements on a non-final conjunct differ from those on independent verbs.

3 Coordination with *ve*

Turkish has several structures that correspond largely to VP coordination in English. Namely, the suffix *-ip*, the coordination word *ve*, the coordination clitic *de*, and simple juxtaposition (Lewis, 1967). In this paper, we consider the structures with the suffix *-ip* and the word *ve*, as in examples (2) and (3).

(2) Çocuk-lar film izle-**yip** pizza yi-yor-lar-dı.
child-PL movie watch-COORD pizza eat-CONT-3PL-PAST
"The children were watching a movie and eating pizza."

(3) Çocuk-lar film izli-yor **ve** pizza yi-yor-lar-dı.
child-PL movie watch-CONT and pizza eat-CONT-3PL-PAST
"The children were watching a movie and eating pizza."

According to the native Turkish speakers consulted, both of these coordination structures share the property that all conjuncts must have the same tense, aspect and mood even though they may be only overtly marked on final conjuncts. The difference between these two structures lies in the morphological requirements on the first conjunct. The verb marked with *-ip* in example (2) may not bear any other markers. On the other hand, the progressive marker *-yor* is obligatorily repeated in the *ve* structure. In example (3), two of the three suffixes are only marked on the final verb. Additional inflection markers may be present on the preceding conjunct, as long as they are also found on the following conjunct. This reflects the phenomenon often referred to as "suspended affixation". In the rest of this section, we provide a more detailed description of VPs coordinated with *ve*, and propose an analysis for suspended affixation. We take up *-ip* coordination in §4.

3.1 Shared TAM Features

As mentioned above, speakers reject expressions where VPs are coordinated that do not have the same tense, as in example (4). If tense and aspect marking is the same, any two VPs can be coordinated using *ve*.[7]

(4) * Çocuk-lar film izli-yor-du ve pizza yi-yecek.
child-PL movie watch-CONT-PAST and pizza eat-FUT
"The children were watching a movie and will eat pizza" (intended)

[7]The data presented in examples (5) and (6) was provided to us by a native speaker, and rated as acceptable by two others. One of the native speakers we consulted, however, did not accept any of these examples, stating that the plural agreement marker is missing on the verb. See §3.2 for more remarks on the subject.

(5) Çocuk-lar film izli-yor-du ve pizza yi-yor-du.
child-PL movie watch-CONT-PAST and pizza eat-CONT-PAST
"The children were watching a movie and eating pizza"

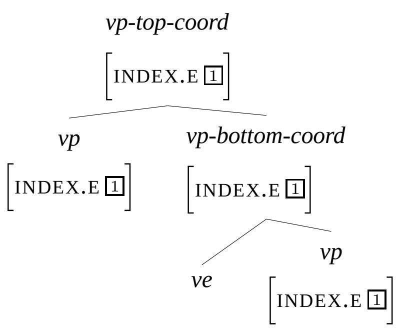
(6) Çocuk-lar film izli-yecek ve pizza yi-yecek.
child-PL movie watch-FUT and pizza eat-FUT
"The children will watch a movie and eat pizza"

We assume that this required identity of TAM morphemes is a semantic constraint (i.e. coordinated VPs must express events taking place in the same time, with the same mood, aspect, etc.), and implement it via a sharing of semantic features.

Just like our analysis of verbal inflection, the coordination analysis here builds upon the implementation of coordination defined through the Matrix customization system (Drellishak and Bender, 2005). Through the customization system, we derived an implementation of polysyndetic coordination, with coordination marker *ve*. This was later manually extended to also include the customization system's implementation of monosyndetic coordination, in order to account for some of the examples found in Kabak 2007.

Following general practice in MRS (Copestake et al., 2005), the event variable of the elementary predication introduced by a verb is also "published" through the verb's INDEX value. Furthermore, this INDEX value is shared with larger constituents that are projections of that verb, and thus the coordination construction has access to the information it needs to ensure matching of event features across conjuncts. The Matrix coordination analysis assumes that a coordinated structure consists of a *bottom-coord-phrase* combining the coordination marker with the right element of the coordination and a *top-coord-phrase* that adds the left conjunct, as in (7). In the Matrix definition of basic coordinated verb phrases, the TAM features of the coordinated phrase are identical to those of the right conjunct. Semantically ill-formed structures (i.e. structures in which left and right conjunct have a different TAM interpretation) can easily be excluded by sharing the TAM features of the left conjunct as well. With this additional constraint, unification fails when left and right conjunct provide conflicting semantics. The tree in (7) provides a simplified example of a VP *ve* VP coordination.

(7)



3.2 Suspended Affixation

In §3.1, we saw that verbs must bear the same tense and aspect markers in order to form a coordinated VP. However, if we look at (3), repeated as (8) below, it is possible to coordinate the forms *izli-yor* and *yi-yor-lar-dı*, despite the fact that only the last form bears a past tense marker.

(8) Çocuk-lar film izli-yor ve pizza yi-yor-lar-dı.
child-PL movie watch-CONT and pizza eat-CONT-3PL-PAST
"The children were watching a movie and eating pizza."

We see the sharing of tense and aspect information in (8) as well. Here, *izli-yor* is interpreted as if it also bore the past tense and agreement markers visible on the second form. If the past tense marker is only present on the first verb of the VP coordination, the sentence becomes unacceptable, as in example (9):

(9) * Çocuk-lar film izli-yor-du ve pizza yi-yor.
child-PL movie watch-CONT-PAST and pizza eat-CONT
"The children were watching a movie and eating pizza."

Since Lewis 1967, this phenomenon has been known as "suspended affixation". Suspended affixation also occurs in nominal coordination where case and number marking are shared. Even though only VP coordination is discussed in this work, the proposed analysis easily extends to NP-coordination.

In verbal structures, suspended affixation does not allow arbitrary strings to be omitted. Rather, as argued in Kabak 2007, a form exhibiting suspension of affixes is acceptable only if it constitutes a morphological word, i.e., a word able to stand in isolation. According to Kabak, morphological words end in "terminal morphemes"; agreement morphemes and aspect and modality morphemes are "terminal".[8] These terminal aspect and modality morphemes are all of the slot 1 morphemes in Table 2.1 except *-DI* and *-sE*.

For instance, in example (10), suspended affixation is not possible. It can only be interpreted as two coordinated sentences. Interpreting the first verb with no agreement marking, i.e. without a null 3SG morpheme, is not possible as the verb must end in a terminal morpheme and so cannot end in *-DI*. In contrast, in example (11), the first verb is interpreted as undergoing suspended affixation since *-yor* is a terminal morpheme. Therefore both verbs are understood to have the same subject.

(10) Film izle-di-∅ ve pizza ye-di-m
movie watch-PAST-3SG and pizza eat-PAST-1SG
"(S)he watched a movie, and I ate pizza."

(11) Film izli-yor ve pizza yi-yor-um.
movie watch-CONT and pizza eat-CONT-1SG
"I am watching a movie and eating pizza."

[8] The affix *-ip*, discussed in §4 also functions as a terminal morpheme.

Speakers have a strong preference for coordinated VPs over coordinated sentences with pro-drop. Example (10) was judged "not nice" and one of our speakers even rated it "ungrammatical". This preference may explain why none of the speakers consulted could interpret *izli-yor* as a fully inflected form of third person singular in example (11).

To our knowledge, the work presented by Kabak (2007) provides the most detailed and precise description of suspended affixation available. In the data we collected from native speakers, however, another issue emerged that was not evident in Kabak's data. Three of our four native speakers accepted the example in (12).

(12) Çocuk-lar film izli-yor-dı-∅ ve pizza yi-yor-lar-dı
Child-PL movie watch-CONT-PAST-3 and pizza eat-CONT-3PL-PAST
"The children were watching a movie and eating pizza."

(12) is an apparent counter-example to Kabaks's generalization about the forms that can appear with suspended affixation, as it ends with *-DI*. However, these speakers appear to treat the -∅ marker as unmarked for number, even in non-coordinated contexts, like (13).[9] Thus Kabak's generalization can be maintained.

(13) Çocuk-lar film izle-r-∅
Child-PL movie watch-AOR-3
"The children watch a movie"

One puzzle remains, however, and is illustrated in (14). The speakers we consulted interpreted this example as having two distinct subjects, but if the -∅ third-person marker is underspecified for number, a same-subject reading should be available.

(14) Çocuk-lar film izli-yor-lar-dı ve pizza yi-yor-dı-∅
Child-PL movie watch-CONT-3PL-PAST and pizza eat-CONT-PAST-3
"The children were watching a movie and he was eating pizza."

Perhaps it is possible to account for this with an appeal to pragmatics, where the marking on the first conjunct is taken as contrastive. Alternatively, a syntactic account in terms of including a feature [LAR *luk*] registering presence of overt plural markers could account for this data. The coordination construction can then exclude structures where the left-hand daughter is [LAR +] and the right-hand daughter [LAR −]. This analysis works in similar ways as that of multiple suspended affixation explained in §3.4, but is relatively inelegant. We leave the resolution of this issue to future work.

3.3 Analysis of Suspended Affixation

The analysis of coordination presented in §3.1 does not accommodate suspended affixation, since only verbs bearing agreement markers are considered words. In

[9] The one speaker who rejected (12) also rejected (13).

order to account for examples such as (11), we introduce a lexical rule, called the *non-final-conjunct-rule*, that changes verbs bearing a morpheme from slot 1 into words, without adding any further inflection. It takes a verbal form ending in a terminal TAM morpheme as its daughter and creates a word that must be the left daughter of a coordinated structure. The rule sketched in (15) below.

(15) $\begin{bmatrix} \text{SYNSEM.LOCAL.CAT.HEAD.NONFIN-CONJ} + \\ \text{DTR } \textit{term-morph-infl-lex-rule} \end{bmatrix}$

The constraint on the DTR value ensures that this rule may only take as input forms ending with a slot 1 morpheme other than *-DI* and *-sE*; the type *term-morph-infl-lex-rule* is a supertype to all lexical rules that introduce such slot 1 morphemes. When the rule in (15) applies, it creates a word which is underspecified for TENSE and ASPECT, making it compatible with values for these features "unified in" from the right hand conjunct in a coordination structure. The other rules that take slot 1 morphemes as input are the ordinary rules for slots 2 and 3. When these rules apply, the resulting form is not restricted to be a left conjunct and it is given specific values for TENSE and/or ASPECT. In this way, we capture Kabak's generalization that there are two paths for a lexeme to become a well-formed morphological word, through the *non-final-conjunct-rule* or through the slot 2 and 3 rules.

As shown in (15), we posit a head feature NONFIN-CONJ, which takes values of type *luk*.[10] *Luk* is a supertype of *boolean* and *na* (*not-applicable*). The *na* value allows us to distinguish coordinated structures from non-coordinated structures, and facilitates the analysis of suspended affixation in multiple coordination (§3.4). The subtypes of *boolean* are used to distinguish verbs that are marked as non-final conjuncts ([NONFIN-CONJ $+$]) and exclude them from the head daughter position of subject-head phrases ([NONFIN-CONJ *na-or-*$-$]) and the right conjunct of coordinated structures ([NONFIN-CONJ $-$]). The value of left conjuncts in these structures is unrestricted, since suspending affixes is an optional process. (16) illustrates this analysis of binary VP coordination with *ve*.

(16)

vp-top-coord $\begin{bmatrix} \text{INDEX.E } \boxed{1} \\ \text{HEAD.NONFIN-CONJ } \textit{na} \end{bmatrix}$

- *vp* $\begin{bmatrix} \text{INDEX.E } \boxed{1} \end{bmatrix}$
- *vp-bottom-coord* $\begin{bmatrix} \text{INDEX.E } \boxed{1} \\ \text{HEAD.NONFIN-CONJ } \boxed{2} - \end{bmatrix}$
  - *ve*
  - *vp* $\begin{bmatrix} \text{INDEX.E } \boxed{1} \\ \text{HEAD.NONFIN-CONJ } \boxed{2} \end{bmatrix}$

[10]In using the type *luk*, we follow the English Resource Grammar (Flickinger, 2000).

3.4 Suspended Affixation with Multiple Conjuncts

In addition to the data presented in Kabak (2007), we looked at structures with more than two conjuncts. In this case, suspended affixation can apply as long as the verb is not preceded by a fully inflected verbal form that is part of the same VP coordination. In fact, speakers prefer expressions where suspended affixation has applied to all but the last verb. The examples below illustrate cases of well- and ill-formed structures with multiple conjuncts.

(17) Çocuk-lar kitap oku-yor-lar-di, film izli-yor-lar-di ve
child-PL book read-CONT-3PL-PAST movie watch-CONT-3PL-PAST and
pizza yi-yor-lar-di.
pizza eat-CONT-3PL-PAST
"The children were reading a book and watching a movie and eating pizza"

(18) ? Çocuk-lar kitap oku-yor, film izli-yor-lar-di ve pizza
child-PL book read-CONT movie watch-CONT-3-PL-PAST and pizza
yi-yor-lar-di.
eat-CONT-3PL-PAST
"The children were reading a book and watching a movie and eating pizza"

(19) Çocuk-lar kitap oku-yor, film izli-yor ve pizza
child-PL book read-CONT movie watch-CONT and pizza
yi-yor-lar-di.
eat-CONT-3PL-PAST
"The children were reading a book and watching a movie and eating pizza"

(20) * Çocuk-lar kitap oku-yor-lar-di, film izli-yor ve pizza
child-PL book read-CONT-3PL-PAST movie watch-CONT and pizza
yi-yor-lar-di.
eat-CONT-3PL-PAST
"The children were reading a book and watching a movie and eating pizza"

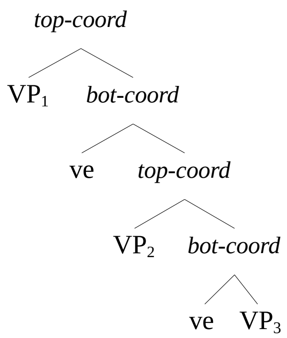
The data above suggest the following generalizations:[11]

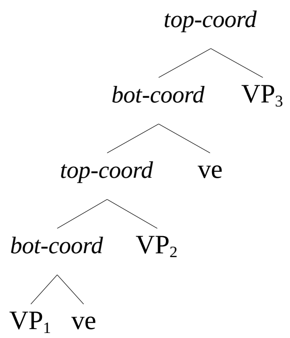
(i) The final VP of a coordinated structure must be fully inflected.

(ii) Fully inflected VPs may not precede VPs that exhibit suspended affixation within a coordinated structure.

Coordination structures provided by the Matrix customization system are right-branching. This is problematic for generalization (ii) above. Consider the right-branched structure in example (21). The data shows that if affixes on VP_2 are suspended, VP_1 may not be fully inflected, but we cannot pass the value of NONFIN-CONJ from VP_2 to the coordinated VP above it, because outside of multiple coordination, that coordinated VP behaves as if it is [NONFIN-CONJ *na*].

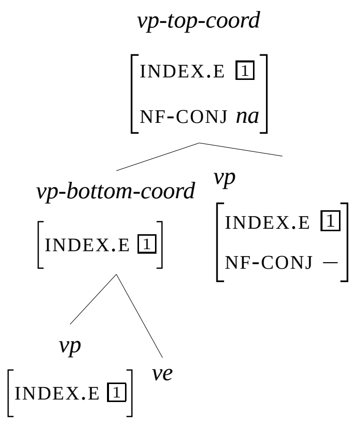
[11]Speakers have different intuitions on this data. Some only accept (19) and (17). Others say that none of the examples is "completely ungrammatical". All speakers agree, however, that the order of acceptability is clear: (19)> (17) > (18) > (20)

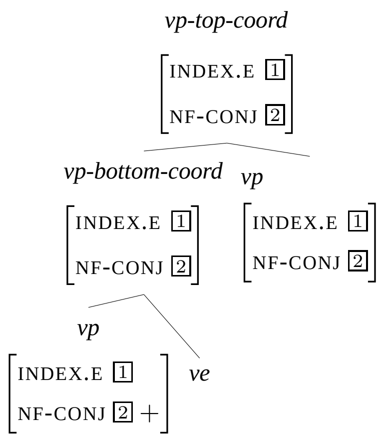
(21) 

(22) 

A more natural approach may be to assume that the morphology of VP_1 may pose restrictions on following conjuncts. Compare the structure in example (22) to the one represented in example (21). In (22), the right daughter of a well-formed embedded VP can determine restrictions on the rest of the structure. This allows us to impose the restriction that a VP that has suspended affixes may only serve as the right conjunct if the left conjunct has suspended affixes as well. The resulting VP coordination of two such VPs bears the value [NONFIN-CONJ +] and must occur as the left daughter of a coordinated VP itself. A fully inflected VP, on the other hand, may always be the right conjunct in a coordinated VP. Because the resulting coordination is [NONFIN-CONJ *na*], it can never become left conjunct when the right conjunct exhibits suspended affixation.

The analysis we assume requires two coordination constructions: one for left conjuncts that exhibit suspended affixation, and one for left conjuncts that do not. The trees in (23) and (24) represent the two constructions.

(23) 

(24) 

Changing VP coordination to a left-branching structure seems natural for Turkish, since it is a language that generally prefers left-branching structures. It also provides further insight in typological properties of coordination structures. Drellishak and Bender (2005) assume that a cross-linguistic analysis of coordination could make do with right-branching structures only, and suggest that the only structures a right-branching approach would exclude are unattested examples such as "*conj* A B C" (*Ibid.*, p.18). Multiple *ve*-coordination reveals an unforeseen case where left-branching seems required. This is because of the double role suspended affixation and complete inflection play in the well-formedness conditions of the complete coordination. On the one hand, the presence of full inflection on the final conjunct is a well-formedness condition that must be encoded on the final structure, so that the coordinated VP can be combined with other elements in the sentence. On the other hand, this same property poses restrictions internal to the coordinated VP, which requires this information to be shared among the (non-final) conjuncts. In a right-branching structure, the final conjunct is the most embedded phrase within the coordination. Relevant information must thus be passed up through the entire coordination construction in order to appear on the resulting coordinated VP. This makes it impossible to share information between phrases that are added to the coordination structure later on, if they appeal to the same feature. When using left-branching coordination, on the other hand, this problem is avoided: relevant information can be passed up directly from the VP that was added to the structure last, allowing the final conjunct to provide relevant information concerning the entire VP. At the same time, restrictions that are internal to VP coordination can be handled by the interaction between *vp-top-coord* and *vp-bottom-coord*.

3.5 Summary

This section has presented an analysis of *ve* coordination and suspended affixation. The analysis accounts for the matching of tense, aspect and modality features across the conjuncts in *ve* coordination structures as well as the potential for affixes to be "dropped" from left-hand conjuncts. In addition, our analysis extends to coordination of more than two conjuncts with *ve* and captures the facts about the distribution of suspended affixation in these constructions.

4 The -*ip* Structure

In this section, we discuss the other coordination structure of interest: coordination marked with the suffix -*ip*. As with *ve* coordination, the semantics associated with the inflection marked on the final conjunct are shared with any other conjuncts. In contrast to the *ve* structure, -*ip* is a suffix on the verb in the left conjunct and it cannot co-occur with any inflectional morphology. This section provides a description of our analysis of the -*ip* structure as a coordination relation. In addition, we provide a brief discussion of the consequences for this approach of an alternative analysis of -*ip* as a "converb" marker.

4.1 Affixal VP-Coordination

In order to implement the *-ip* coordination relation, we follow Drellishak and Bender's (2005) analysis for the Trans-New Guinea language Ono wherein a feature registers the presence of marking that is relevant for VP coordination. This feature allows the VP to become part of a coordinated structure through a unary rule. The Turkish *-ip* suffix is, similarly, a VP coordination marker attached to a word. While this option is not directly provided by the Matrix customization system, the relevant constraints can be added to an analysis provided by the customization system in a straight-forward way. In Ono, the marked form was the right conjunct. In our case, where the left conjunct is marked, we needed to change coordination into a left-branching structure, as was done for *ve*-coordination.

The suffix *-ip* cannot occur with any other inflectional morphemes but can be added directly to the stem or to derivational morphemes. According to our analysis, it is therefore added to the verb at the first slot for inflectional suffixes, creating a word. This lexical rule changes the value of a feature IP to $+$. In all other cases, this feature will have the value $-$. The coordination structure that creates *-ip* coordination only takes left daughters that are VPs and marked [IP $+$], the resulting structure is [IP $-$] again, as illustrated in (25). Note that values related to tense, aspect and mood are shared among the conjuncts, just as for *ve* coordination.

(25)

vp-top-coord
$\begin{bmatrix} \text{HEAD.IP} \; - \\ \text{INDEX.E} \; \boxed{1} \end{bmatrix}$

vp-ip
$\begin{bmatrix} \text{HEAD.IP} \; + \\ \text{COMPS} < > \\ \text{INDEX.E} \; \boxed{1} \end{bmatrix}$

vp-bottom-coord
$\begin{bmatrix} \text{HEAD.IP} \; - \\ \text{INDEX.E} \; \boxed{1} \end{bmatrix}$

vp
$\begin{bmatrix} \text{HEAD.IP} \; - \\ \text{INDEX.E} \; \boxed{1} \end{bmatrix}$

4.2 Converb Marker

Some linguists consider verbs marked by *-ip* "converbs" (Tikkanen, 2001), though in descriptive literature (Lewis, 1967) it is generally treated as a coordination marker. Empirical studies have, to our knowledge, not yet settled this matter; the definition of "converb" is not clear-cut and the importance of the "modifying" character of converbs is debated. Johanson (1995) argues that there are both modifying and non-modifying converbs in Turkish, where non-modifying converbs are distinguished semantically in that they depict "events of equal narrative status" (*Ibid.*, p.322). The difference between these and coordinated clauses relates to the information structure of the clause; converbs may express information that is in focus.

Another difference between modifying and non-modifying converbs in Turkish lies in their interaction with the scope of the main verb. Whereas modifying converbs fall outside of the scope of tense, aspect and modal markers of the main clause, these do have scope over non-modifying converbs that precede them, as illustrated in (26) from Johanson 1995, p.323.

(26)	Herkes	çik-ip	'Ben	Türk-üm	di-yebil-meli
	everybody	come.out-CONV	I	Turk-COP.1.SG	say-POSSIB-NEC.3SG

"Everybody should be allowed to step forth and [should be allowed to] say that he is a Turk."

According to Johanson's definition, the *-ip* structures discussed in this paper should be considered non-modifying converbs. This would mean that their interpretation would be that of events with narrative status equal to that of main verbs. This is exactly what the coordination analysis above provides. The only difference between a coordinated structure and a non-modifying converb structure is the subordinate character of the latter. However, because this is not represented in the final semantic interpretation of the sentence there does not seem to be a reason to propose an analysis that is radically different from the one that is proposed above, except for perhaps changing the names of the phrases used to *converb* rather than *coordinated*. One could also extend the analysis to incorporate the correct information structure, though this is beyond the scope of the present study.

In sum, whether one considers the *-ip* structure as a converbial structure or as a coordinate structure depends on the criteria that are used to distinguish the two. We take the final semantic representation, which is compatible with the coordination account, as the primary consideration and use it as the basis for our analysis.

4.3 Shared Scopal Morphemes

Whereas the question of whether *-ip* marks converbs or coordinated structures is, in our opinion, not of crucial importance, one observation mentioned by Johanson (1995) is particularly relevant here: Verbs bearing *-ip* fall under the scope of the verb they precede. Though we are not aware of accounts that discuss this matter in detail, this property is mentioned by several authors of Turkish grammars. Our data does confirms this observation concerning the wide scope of the suffix *-mEli*. In addition, we found that the suffix -(*y*)*Abil* has scope over the entire coordinated structure when it appears only on the right conjunct. Consider (27) and (28):

(27)	Çocuk-lar	film	izle-yip	pizza	ye-meli-ler.
	child-PL	movie	watch-COORD	pizza	eat-NEC-3PL

"The children must watch a movie and eat pizza."

(28)	Çocuk-lar	film	izle-yip	pizza	yi-yebil-ir-ler.
	child-PL	movie	watch-COORD	pizza	eat-ABIL-AOR-3PL

"The children can watch a movie and eat pizza."

The analysis of *-ip* structures described above handles part of the shared interpretation between the verbs: information regarding tense, aspect and mood are stored as features that are part of the verb's event variable, which is identified across conjuncts. However, *-mEli* and -(*y*)*Abil* contribute information that is usually handled in terms of (scopal) elementary predications: necessity and ability, respectively. Thus, it is more surprising to see this information shared across conjuncts.[12] In §4.4, we demonstrate that a constructional analysis can provide the right semantics for *-ip* structures in which these scopal morphemes occur.

4.4 A Constructional Analysis

If we assume that -(*y*)*Abil* and *-mEli* are scopal and treat them as predicate introducing morphemes, we cannot obtain the correct interpretation of coordinated VPs by simply sharing the value of TAM features across both events. Nor can we just allow the semantics of these morphemes to attach "low"; instead of merely the second verb, the suffixes too must have scope over the entire coordinated VP. This seems to suggest that these affixes attach to phrases rather than words, but "phrasal affixes" would violate the assumption of lexical integrity, which is generally held in HPSG. Instead, we propose a constructional solution, in the spirit of the analysis that Tseng (2003) proposes for apparent phrasal affixes in French.

Both -(*y*)*Abil* and *-mEli* contribute a HEAD feature, each of which is referenced by a special construction that takes a VP daughter and adds the appropriate semantics. The AVM in (29) below provides a simplified representation of the unary *ability-phrase-rule*.

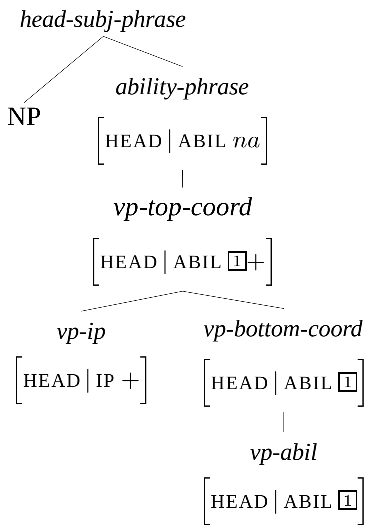
(29)
$$\begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} \\ \text{ABIL} & \textit{na} \end{bmatrix} \\ \text{VAL} & \boxed{1} \end{bmatrix} \\ \text{C-CONT} & \begin{bmatrix} \text{RELS} & \left\langle \begin{bmatrix} \textit{rel} \\ \text{PRED} & \text{“\_}\textit{abil\_rel}\text{”} \\ \text{ARG1} & \boxed{2} \end{bmatrix} \right\rangle \\ \text{HCONS} & \left\langle \begin{bmatrix} \textit{qeq} \\ \text{HARG} & \boxed{2} \\ \text{LARG} & \boxed{3} \end{bmatrix} \right\rangle \end{bmatrix} \\ \text{ARGS} & \left\langle \begin{bmatrix} \text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} \\ \text{ABIL} & + \end{bmatrix} \\ \text{VAL} & \boxed{1}\begin{bmatrix} \text{SUBJ} \langle [\,] \rangle \end{bmatrix} \end{bmatrix} \\ \text{CONT} \mid \text{HOOK} \mid \text{LTOP}\ \boxed{3} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

This non-branching construction licenses a VP node over any VP with the feature declaration [ABIL $+$], and its purpose is to insert the *_abil_rel* predication into

[12]Other derivational morphemes seem not have this property. According to Lewis (1967), the negation morpheme *-mA* also has wide-scope in the *-ip* structure, but none of the speakers we consulted got this reading.

the semantics. This predicate is specified in the C-CONT (construction content) feature of the construction, following standard MRS practice for semantically contentful constructions. It further specifies that the local top handle of the daughter VP is the argument of the introduced predicate.[13] In order to ensure that this construction only applies outside (and not within) VPs coordinated with *-ip*, the mother is marked [ABIL *na*] and the rule licensing the right-hand daughter of an *-ip* structure requires [ABIL *bool*]. A similar construction is posited for *-mEli*, with an associated feature NECESS, subject to analogous constraints. We ensure that the relevant construction fires if the morphology is present by requiring the value *na* for both of these features in the *head-subj-phrase*. The tree below illustrates the workings of the ABIL feature in an *-ip* coordination.

(30)



4.5 Summary

This section has presented an analysis of *-ip* coordination. Our analysis handles the following facts: In *-ip* coordination, non-final conjuncts must be marked with *-ip*, which is incompatible with any other inflectional morphology. Information expressed by inflectional morphemes on the final conjunct (including tense and aspect information) is interpreted as shared with all conjuncts. Our analysis handles this sharing through the same identification of TAM features as in *ve* coordination. In addition, when the final conjunct bears the affixes -(*y*)*Abil* or *-mEli*, these are interpreted as taking wide scope over the whole coordinated VP. We assume these affixes correspond to scopal elementary predications in the semantics and we propose an analysis where the affixes contribute only syntactic features, which then trigger the application of a construction at the level of the coordinated VP introducing the scopal predications into the semantics.

[13] This argument relation is mediated by the "equal modulo quantifiers" (*qeq*) handle constraint, to allow quantifiers to scope in between, while maintaining the scopal relationship where the *_abil_rel* outscopes the verb's (or verbs') predication(s).

5 Related Work

In this section, we situate our analysis with respect to related work. First, in §5.1, we contrast our analysis to the LFG account of Broadwell (2008). Then, in §5.2, we describe how our account of these Turkish facts is broadly similar to Tseng's (2003) account of a very different phenomenon in French.

5.1 Suspended affixation in LFG

To our knowledge, the only other formal account of suspended affixation in Turkish is the LFG account of Broadwell (2008). Broadwell applies Westcoat's (2002) notion of "relaxed lexical integrity" which allows a single (morphological) word to represent two adjacent c-structure nodes, even if the c-structure nodes do not form a constituent. On this analysis, the affixes that are shared between two or more conjuncts represent independent c-structure nodes attaching to the entire coordination. They are associated with special "instantiation" rules which allow them to be co-instantiated with the final word of the nearest conjunct.

Broadwell considers an analysis similar to ours as an alternative to the "co-instantiation" approach. On this alternative analysis, the affixes are part of the final conjunct, which bears special functional equations propagating its values for the features expressed in the affixes to the coordinate structure as a whole. Broadwell argues against this analysis on the basis that it requires the stipulation that the special annotations appear on the rightmost conjunct. On the "co-instantiaton" analysis, the location of the affixes within the coordinate structure can be seen to follow from the general head-final property of Turkish.

However, we argue that lexical integrity is not something to give up lightly. Furthermore, our analysis allows us to capture the similarity between required matching of tense and aspect morphology when it is overt and required matching of tense and aspect values when the morphology is not present on a non-final conjunct. In addition, we note that Broadwell proposes a flat (symmetrical) structure for coordination, whereas we follow a binary-branching analysis. On a binary branching analysis, it is less surprising that one conjunct should have special properties. Finally, Broadwell notes that his syntactic analysis cannot capture Kabak's morphological generalization about which affixes can be suspended, and appeals instead to an external morphological filter. We conclude that our account seems preferable in that it allows us to handle the data in more detail while simultaneously preserving lexical integrity.

5.2 Phrasal affixes in French

Tseng (2003) posits a very similar solution to ours for what appears to be a very different problem. In particular, he is addressing the apparent contradiction between the phonological and syntactic status of the formatives *le, de* and *à* in French. These elements (one determiner and two prepositions) are functors, which we would expect to combine respectively with an N′ or an NP, but phonologically (and, Tseng argues, morphologically) they combine with the first word in that N′ or NP. Since this first word need not be the head, the syntactic and semantic information that the functors require is not available to them locally.

Tseng's solution is to more-or-less freely attach *le*, *de* and *à* as prefixes. The morphophonological information associated with these morphological rules handles contextual variation in the form of the prefix, while the morphosyntactic effect of the rule is to encode information about the affix in an EDGE feature. The EDGE feature is propagated up the periphery of the constituent and is finally interpreted by a unary rule which builds an NP out of an N′ or a PP out of an NP, according to the information stored in EDGE.

The analysis proposed in this paper of -(*y*)*Abil* and -*mEli* in -*ip* coordination differs from Tseng (2003) in that it does not refer to the feature EDGE and the values of our phrasal affix features are less complex: they merely register presence of particular morphemes. It would be possible to adapt this analysis and make it more similar to Tseng's account of phrasal affixes in French, with the additional advantage that we would only use one feature (EDGE) rather than two (ABIL) and (NECESS). Fundamentally, however, our analysis is exactly parallel to Tseng's, in positing a pair of rules, one morphological and one syntactic, in order to handle apparent phrasal affixes without sacrificing lexical integrity. The fact that the same analytical device can handle such superficially different phenomena speaks to its generality while also raising interesting questions about the typology of phrasal affixes. When are such paired rules required, and why are they not more common?

6 Conclusion

This paper presented three phenomena related to the morpho-syntax of Turkish VP coordination. First, our data showed that tense, aspect and modality marking on coordinated VPs must be identical. We proposed an analysis that models this by sharing the value of event semantics on both VPs.

The second phenomenon we discussed is that of suspended affixation. This paper introduced new observations related to plural markers and coordination with multiple coordinands. §3.2 and §3.4 presented analyses for binary and multiple coordination, respectively. The latter showed that the restrictions on multiple coordinands require left branching coordination structures, contra the claim in Drellishak and Bender 2005.

The only other formal analysis of suspended affixation that we are aware of is described in Broadwell (2008). In §5.1, we discussed this alternative account and argue that our proposal is superior because (i) it can account for the morphosyntactic properties of the phenomenon as described by Kabak and (ii) it respects lexical integrity.

Finally, we discussed coordinated structures that make use of the suffix -*ip*. The alternative view that -*ip* is a converb marker was discussed, and it was argued that treating these verbs as converbs does not simplify the analysis, nor lead to a more accurate semantic representation of the sentence. We presented data that shows that markers on the verb that follows the VP marked with *-ip* scope over both VPs and therefore seem to attach to a phrase rather than a word. This would violate HPSG assumptions on lexical integrity. We show, however, that the data can be analyzed with the help of a construction.

All the analyses presented in this paper have been implemented in a small grammar fragment. In addition to presenting the phenomena and their analyses, we also indicated how the analyses were implemented with help of the Matrix customization system. This had two main benefits: First, it allowed us to test both the accuracy of our analyses and whether they could be implemented in a mutually consistent fashion. Second, it allowed us to test the cross-linguistic applicability and utility of the Grammar Matrix. On the one hand, the Grammar Matrix customization system supported the creation of this paper: it allowed us to quite quickly produce a grammar testing our hypotheses, which confirmed its applicability and utility. On the other hand, our implementations pointed to a typological fact that had not been foreseen in building the coordination library of the Matrix: namely that morphological properties may require left-branching coordination structures.

References

Bender, Emily, Flickinger, Dan and Oepen, Stephan. 2002. The Grammar Matrix: An Open-Source Starter-Kit fir the Rapid Development of Cross-linguistically Consistent Broad-Coverage Precision Grammars. In *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th Conference on Computational Linguistics*, pages 8 – 14, Taipei, Taiwan.

Bender, Emily M. 2008. Grammar Engineering for Linguistic Hypothesis Testing. In Nicholas Gaylord, Alexis Palmer and Elias Ponvert (eds.), *Proceedings of the Texas Linguistics Society X Conference: Computational Linguistics for Less-Studied Languages*, pages 16–36, Stanford: CSLI Publications.

Bender, Emily M. and Flickinger, Dan. 2005. Rapid Prototyping of Scalable Grammars: Towards Modularity in Extensions to a Language-Independent Core. In *Proceedings of the 2nd International Joint Conference on Natural Language Processing IJCNLP-05 (Posters/Demos)*, Jeju Island, Korea.

Broadwell, George Aaron. 2008. Turkish Suspended Affixation Is Lexical Sharing. In *Proceedings of the LFG08 Conference*.

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

Copestake, Ann, Flickinger, Dan, Pollard, Carl and Sag, Ivan. 2005. Minimal Recursion Semantics: An Introduction. *Research on Language and Computation* 3(4), 281–332.

Drellishak, Scott and Bender, Emily. 2005. A Coordination Module for a Crosslinguistic Grammar Resource. In Stefan Müller (ed.), *Proceedings of the HPSG05 Conference*, Stanford: CSLI Publications.

Erguvanlı-Taylan, Eser. 1999. Review: Turkish by Jaklin Kornfilt. *Anthropological Linguistics* 41(2), 253 – 258.

Flickinger, Dan. 2000. On Building a More Efficient Grammar by Exploiting Types. *Natural Language Engineering* 6 (1) (Special Issue on Efficient Processing with HPSG), 15 – 28.

Johanson, Lars. 1995. On Turkish Converb Clauses. In Martin Haspelmath and Ekkehard König (eds.), *Converbs in Cross-linguistic Perspective*, pages 313 – 349, Berlin: Mouton de Gruyter.

Kabak, Bariş. 2007. Turkish Suspended Affixation. *Linguistics* 45(2), 311–347.

Kornfilt, Jaklin. 1997. *Turkish*. New York: Routledge.

Lees, Robert B. 1962. A Compact Analysis for the Turkish Personal Morphemes. In Nicholas Poppe (ed.), *American Studies in Altaic Linguistics*, volume 13 of *Indiana University Uralic and Altaic Series*, The Hague, the Netherlands: Mouton.

Lewis, Geoffrey. 1967. *Turkish Grammar*. New York: Oxford University Press, second edition.

O'Hara, Kelly. 2008. *A Morphosyntactic Infrastructure for a Grammar Customization System*. Masters Thesis, University of Washington.

Sezer, Engin. 2001. Finite inflection in Turkish. In Eser Erguvanlı Taylan (ed.), *The Verb in Turkish*, Chapter 1, pages 1 – 46, Philadelphia: John Benjamins.

Tikkanen, Bertil. 2001. Converbs. In Martin Haspelmath and Wulf Oesterreicher (eds.), *Language Typology and Language Universals: An International Handbook*, volume 2, Chapter 83, Berlin, Germany: W. de Gruyter.

Tseng, Jesse. 2003. Phrasal Affixes and French Morphosyntax. In G. Jäger, P. Monachesi, G. Penn and S. Wintner (eds.), *Proceedings of Formal Grammar 2003*, pages 177–188, Stanford: CSLI Publications.

Westcoat, Michael T. 2002. *On Lexical Sharing*. Ph. D.thesis, Stanford University.

Towards an analysis of the adverbial use of German interrogative *was* (‘what’)

Anke Holler

University of Göttingen

Proceedings of the HPSG09 Conference

Georg-August-Universität Göttingen, Germany

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

The paper discusses the so-called adverbial use of the *wh*-pronoun *was* ('what'), which establishes a non-standard interrogative construction type in German. It argues that the adverbial use of *was* ('what') is based on the lexical properties of a categorically deficient pronoun *was* ('what'), which bears a causal meaning. In addition, adverbial *was* ('what') differs from canonical argument *was* ('what') as it is analyzed as a functor which is generated in clause-initial position.

By means of empirical facts mainly provided by d'Avis (2001) it is shown that *was* ('what') behaves ambivalently regarding the *wh*-property: On the one hand, *was* ('what') can introduce an interrogative clause, but on the other hand it cannot license *wh*-phrases in situ. While formally analyzing the data against the background of existing accounts on *wh*-interrogatives couched in the framework of Head-driven Phrase Structure Grammar, an analysis is developed that separates two pieces of information to keep track of the *wh*-information percolating in an interrogative clause. Whereas the WH-value models *wh*-fronting and pied-piping phenomena, the QUE value links syntactic and semantic information and thus keeps track of *wh*-in-situ phrases.

1 Introduction

Interrogative constructions always have been of great interest to linguists, and thus, it is not surprising that the analysis of *wh*-interrogatives also gained a lot of attention in the framework of Head-driven Phrase Structure Grammar (HPSG). There exists by now a significant amount of HPSGian work on *wh*-interrogatives dealing with relevant syntactic and semantic phenomena such as *wh*-fronting, *wh*-embedding, and *wh*-scope assignment.

With Ginzburg and Sag (2000), who provided a comprehensive account of a wide range of interrogative constructions in English, the core problems concerning the analysis of *wh*-interrogatives in a constraint-based framework seemed to be tackled. Nevertheless, some central issues are still open and worth to be discussed. van Eynde (2004), for instance, has shown that Ginzburg and Sags' treatment of pied piping as a non-local dependency faces a number of problems. He therefore proposes to restrict the percolation of the *wh*-property within a *wh*-interrogative phrase by treating pied piping locally. In this article I will contribute another fact to the discussion that suggests a revision of the bookkeeping mechanism of the *wh*-property percolating within a *wh*-interrogative clause. By analyzing the German non-standard adverbial *was*-construction I will argue that there exist deficient *wh*-phrases whose behaviour necessitates in a constraint-based grammar a separated representation of information on syntactic *wh*-fronting on the one hand and information on the realization of *wh*-in-situ phrases on the other hand.

[†]I would like to thank two anonymous reviewers and the audience of the HPSG09 conference for valuable comments. I am also grateful to Marianne Desmets, Danièle Godard, Jacob Maché, Alexandr Rosen, Ivan Sag, and Manfred Sailer for data contributions and helpful discussions.

The adverbial *was*-construction under discussion is empirically well-known from German grammar writing, and has been grammatically described thoroughly by d'Avis (2001), who couched his analysis in the framework of Government and Binding. An example of the considered construction is given in (1), which is taken from d'Avis (2001).

(1) Was schlägst du denn schon wieder den Hund?
What beats you PART PART again the dog
'Why are you beating the dog again?'

The adverbial *was*-construction is peculiar mainly because it is introduced by a *wh*-phrase *was* ('what') which does not function as a subject or object phrase. Instead, *was* ('what') is used similar to an adverbial phrase bearing the meaning of *why*. Thus, the non-standard *wh*-construction in (1) has a causal interrogative meaning although it contains no overt *wh*-phrase canonically possessing a causal lexical meaning.

In this article I will argue by means of empirical data basically provided by d'Avis (2001) that the adverbial use of the *wh*-phrase *was* ('what') follows from a categorial deficiency of the pronominal element *was* ('what'). In addition, I will formally analyze the data against the background of existing HPSG accounts on *wh*-interrogatives, thereby showing that none of the previous proposals couched in HPSGian theoretical terms is sufficient to account for the deficiency of adverbial *was* ('what') and the respective German non-standard *wh*-interrogative construction type.

The article is structured as follows. Firstly, I will introduce the empirical properties of adverbial *was* ('what') and the corresponding non-standard *wh*-interrogative construction. Secondly, I will briefly present recent HPSG approaches to *wh*-interrogatives, and I will show that all of them are not appropriate to deal with *wh*-clauses introduced by adverbial *was* ('what'). Thirdly, I will provide evidence that adverbial *was* ('what') behaves like a categorically deficient pronoun (cf. Cardinaletti and Starke (1999); Abeillé and Godard (2003)), and I will finally develop a new analysis covering the presented empirical facts, thereby arguing that an adequate account of non-standard uses of *was* ('what') requires a better differentiation between syntactic *wh*-fronting on the one hand and handling *wh*-in situ on the other hand.

2 Empirical facts

The non-canonical adverbial *was*-construction exemplified by (1) is distinguished by two obvious properties: (i) The *wh*-pronoun *was* ('what') heads a *wh*-phrase that has to be realized in clause-initial position, and (ii) *was* ('what') bears the meaning of *why* instead of the meaning of *what* and is, thus, used like a causal adverb. The things are even worse: Although *was* ('what') contributes a causal meaning, it does not behave like a standard interrogative adverb either. Thus, an obvious analysis that treats *was* ('what') as a semantically ambiguous pronoun simply having two different lexical meanings cannot cope with the empirical facts, which I will present next in more detail.

2.1 Empirical properties of the German adverbial *was*-construction

As d'Avis (2001) observed, the non-standard *was* ('what') differs from an ordinary *wh*-word in several aspects, particularly with respect to coordination, extraction, and *wh*-in situ phenomena.

First of all, adverbial *was* ('what') contrasts to any standard *wh*-phrase as it cannot be realized in situ. As example (2a) illustrates, adverbial *was* ('what') is restricted to a clause-initial position (the so-called Vorfeld). Contrary to this, *warum* ('why') can be used in situ without any problems as (2b) demonstrates.

(2) a. * Wann trifft sich Maria was mit ihrem Exmann?
when meet REFL Maria what with her divorcé

b. Wann trifft sich Maria warum mit ihrem Exmann?
when meet REFL Maria why with her divorcé
'When does Maria meet her divorcé for which reason?'

In addition, adverbial *was* ('what') cannot be realized in reprise questions, although standard *wh*-phrases such as *warum* ('why') are completely fine in such a context. This is illustrated by (3b) versus (3c).

(3) a. Hans will sich scheiden lassen, weil seine Frau zu viel arbeitet.
Hans wants REFL divorce let because his wife too much works
Hans wants to divorce because his wife works too much.'

b. * Hans will sich WAS scheiden lassen?
Hans wants REFL what divorce let

c. Hans will sich WARUM scheiden lassen?
Hans wants REFL why divorce let
'Hans wants to divorce WHY?'

The reverse side of the just mentioned properties of adverbial *was* ('what') seems to be that *was* ('what'), contrary to *warum* ('why'), cannot license another *wh*-phrase in situ. Whereas the multiple *wh*-question in (4a) is completely out, the one in (4b) is perfect.

(4) a. * Was spielt sich wer denn so auf?
what act REFL who PART that way up

b. Warum spielt sich wer denn so auf?
why act REFL who PART that way up
'Why does who act that way up?'

Furthermore, (5) demonstrates that adverbial *was* ('what') cannot be extracted out of a complement clause although *warum* ('why') allows this extraction (at least in some dialects of German).

(5) a. * Was glaubst du, dass Otto den Hund t geschlagen hat?
what believe you that Otto the dog t beaten has

b. Warum glaubst du, dass Otto den Hund t geschlagen hat?
why believe you that Otto the dog t beaten has
'Why do you believe that Otto has beaten the dog?'

In addition, the contrast in (6) indicates that adverbial *was* ('what') cannot be coordinated with another standard *wh*-phrase. This is a fact that one would not expect if adverbial *was* ('what') were a canonical *wh*-interrogative expression.

(6) a. * Wann und was will sich Maria scheiden lassen?
when and what wants REFL Maria divorce let

b. Wann und warum will Maria sich scheiden lassen?
when and why wants Maria REFL divorce let
'When and why wants Maria to divorce?'

Last but not least, adverbial *was* ('what') cannot bear a focus accent. As the contrast in (7a) versus (7b) exemplifies, this is again in contrast to canonical adverbial *wh*-phrases like *warum* ('why') .

(7) a. * Ich möchte wissen, WAS Maria sich scheiden lassen will und
I want to know what Maria REFL divorce let wants and
nicht wann.
not when

b. Ich möchte wissen, WARUM Maria sich scheiden lassen will
I want to know why Maria REFL divorce let wants
und nicht wann.
and not when
'I want to know why Maria wants to divorce and not when.'

In the light of these facts one could wonder whether the adverbial *was*-construction is a *wh*-interrogative construction at all. But evidence for its interrogativity comes from data like (8) and (9).

(8) indicates that adverbial *was*-constructions are not limited to root clauses, but can be combined with question embedding predicates such as *wonder* and *want to know*, which means that the adverbial *was*-construction can be used as an indirect question.

(8) Ich möchte wissen, was Maria den Hund wieder schlägt.
I want to know what Maria the dog again beats
'I want to know why Maria is beating the dog again.'

In addition, example (9) demonstrates that the German equivalents to expressions like *the hell*, *on earth* or *the devil*, whose occurrence is clearly restricted to interrogative phrases, can be added to adverbial *was*.

(9) Was zum Teufel schlägst du schon wieder den Hund?
why the devil beats you REFL again the dog
'Why the devil are you beating the dog again?'

Thus, there seems to be no doubt that adverbial *was* ('what') introduces a *wh*-interrogative clause with a question meaning. On the other hand, it is obvious that this construction at least syntactically does not behave like a standard *wh*-interrogative clause. This raises the question of how we can account for this non-canonical behaviour of the adverbial *was*-construction and which consequences for a HPSGian treatment of *wh*-interrogative clauses in general result from this.

2.2 The adverbial *was*-construction is not restricted to German

Note that the adverbial *was*-construction is by no means an idiosyncratic German construction. For instance Nakao and Obata (2009) discuss accusative *wh*-adjuncts with reason meaning in Japanese. Interestingly enough, the data they provide for Japanese match the German facts. *Nani-o* in example (10) behaves, in grammatical terms, exactly like German *was* ('what') as it is an accusative *wh*-adjunct with reason meaning.

(10) a. Kare-wa nani-o sawai-dei-ru no?
he-TOP what-ACC make-noise-PROG-PRES Q
'Why is he making a noise?'

b. Kare-wa naze sawai-deu-ru no?
he-TOP why make-noise-PROG-PRES Q
'Why is he making a noise?'

In addition, as M. Desmets and A. Rosen (p.c.) pointed out, the same construction type can be observed in French and Czech. Since the adverbial use of interrogative *what* is not confined to German, an adaquate analysis in constraint-based grammar seems to be required. However, having a closer look at existing HPSGian approaches to *wh*-interrogatives, none of them seems to be appropriate to capture the peculiarities of the adverbial *was*-construction. Two major reasons are responsible for this result: Firstly, in all previous accounts it is assumed that basically any *wh*-phrase can be realized in-situ, and secondly there is no device that allows a fronted *wh*-phrase to have access to the information whether an in-situ *wh*-phrase is present or not. Consequently, the grammar overgenerates because there is no way to exclude the ungrammatical examples in (2a) and (4a). The problem arises since in all accounts two structural aspects of interrogative clauses, i.e. (i) the topicalization of a single *wh*-phrase on the one hand, and (ii) the handling of *wh*-phrases in-situ on the other hand, are somehow mixed up by using just one single feature value to keep track of the syntactic *wh*-information. I will briefly elaborate on this issue in the next section.

3 Relevant HPSG approaches to *wh*-interrogatives

It is well-known that the nonlocal feature QUE is usually exploited to represent the *wh*-property. This idea goes back to Pollard and Sags' standard HPSGian analysis of *wh*-interrogatives. QUE, whose value represents a restricted index of type *npro*, is lexically instantiated for all *wh*-words. Subject to the Non-local Feature Principle, the value of the QUE feature percolates in a phrase until bound. In this setup, QUE instantiation and percolation ensures that a *wh*-interrogative clause contains exactly one fronted *wh*-phrase. At the same time, QUE is used to determine the semantic scope of a *wh*-phrase by binding the QUE value at an appropriate constituent. Consequently, the information that a clause contains a *wh*-in-situ phrase is accessible only indirectly.

In their approach to *wh*-scope assignment, Pollard and Yoo (1998) also use the QUE feature to handle *wh*-fronting, pied piping and the licensing of *wh*-in-situ phrases. They, however, suggest that each *wh*-word introduces a quantifier that is represented as a value of QUE, which is a *synsem* feature in their account. In fact, an interrogative operator associated with a *wh*-phrase is stored twice: as value of the QUE feature and as value of the QSTORE feature of the *wh*-phrase. This is depicted in figure 1 showing the partial lexical entry for the *wh*-word *who* according to Pollard and Yoo (1998).

$$\begin{bmatrix} \text{PHON} & \langle \textit{who} \rangle \\ \text{SYNSEM} & \begin{bmatrix} \text{LOCAL} & \begin{bmatrix} \text{CAT} & \text{NP} \\ \text{CONT} & \boxed{2} \begin{bmatrix} \textit{quant} & \\ \text{DET} & \textit{which} \\ \text{RESTIND} & \begin{bmatrix} \textit{npro} & \\ \text{INDEX} & \boxed{1} \\ \text{RESTR} & \left\{ \begin{bmatrix} \text{QUANTS} & \textit{elist} \\ \text{NUCL} \mid \text{INST} & \boxed{1} \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix} \\ \text{QSTORE} & \left\{ \boxed{2} \right\} \end{bmatrix} \\ \text{QUE} & \left\{ \boxed{2} \right\} \end{bmatrix} \end{bmatrix}$$

Figure 1: Partial lexical entry for a *wh*-word according to Pollard and Yoo (1998)

In addition, Pollard and Yoo (1998) implement a Cooper storage mechanism in order to determine the scope of a *wh*-quantifier. They formulate a syntactic licensing constraint on *wh*-retrieval that firstly says that the retrieval of quantifiers introduced by *wh*-in-situ phrases is only allowed if there is a left peripheral *wh*-phrase whose quantifier is simultaneously retrieved, and secondly that any non-empty QUE value of a filler daughter must be retrieved. This constraint thus syntactically cues the point where interrogative meaning is retrieved.

Similar to Pollard and Sags' approach, Pollard and Yoo (1998) cannot account for the ungrammaticality of (2a) and (4a), where *was* ('what') is realized in situ as the information of the realization of a *wh*-in-situ phrase cannot be accessed by the left-peripheral *wh*-filler-phrase. The syntactic licensing constraint on *wh*-retrieval that they formulate only implements the retrieval of quantifiers introduced by *wh*-in-situ phrases in dependence of a simultaneously retrieved left peripheral *wh*-phrase. Nothing is said concerning the relation between a retrieved non-empty QUE value of a filler daughter and a potential *wh*-in-situ phrase. The information that a *wh*-in-situ-phrase might exist could only be derived from the QUANTS-list, but this list cannot be accessed by the fronted *wh*-phrase.

In the construction-based account proposed by Ginzburg and Sag (2000)—an elaborated version of Ginzburg (1992)—the analysis is based on a multi-inheritance hierarchy of sorts with associated sort constraints. Inspired by situation semantics, Ginzburg and Sag (2000) hold the view that questions are basic semantic entities such as individuals and propositions. Grammar objects of sort *question* are distinguished from any other entity in terms of a feature called PARAMS, whose set value must always be non-empty for *wh*-questions. Syntactically, Ginzburg and Sag (2000) basically follow Pollard and Sag (1994) by arguing for a non-local head-driven treatment of wh-interrogatives. The *wh*-property is represented by a set-valued WH feature. *Wh*-words bear an optional WH specification such that the WH value of an interrogative word can either be a singleton set containing a parameter or an empty set as is illustrated in figure 2. This assumption is necessary to

$$\left[\begin{array}{ll} \text{PHON} & \langle \textit{who} \rangle \\ \text{SYNSEM} & \left[\begin{array}{ll} \text{LOCAL} & \left[\begin{array}{ll} \text{CAT} & \text{NP} \\ \text{CONT} & \left[\begin{array}{l} \textit{param} \\ \text{INDEX } \boxed{1} \end{array}\right] \\ \text{STORE} & \left\{ \boxed{2} \left[\begin{array}{l} \textit{param} \\ \text{INDEX } \boxed{1} \end{array}\right] \right\} \end{array}\right] \\ \text{WH} & \{ (\boxed{2}) \} \end{array}\right] \end{array}\right]$$

Figure 2: Partial lexical entry for a *wh*-word according to Ginzburg and Sag (2000)

syntactically handle in-situ *wh*-words without violating the WH constraint saying that all non-initial arguments of a lexeme must be specified as [WH { }]. In addition, the Filler Inclusion Constraint, which requires that the non-head daughter of a clause of sort *wh-interrogative-clause* must be WH-specified, ensures that each interrogative clause is introduced by an expression that is or contains an interrogative *wh*-word. In this setting, the fact that a clause contains a *wh*-phrase in situ is only inferable from a non-singleton PARAMS value of the clause. The non-head daughter of a *wh*-interrogative, however, has no access to this information. Again, the ungrammaticality of (2a) and (4a) cannot be captured. An alternative analysis whereupon the PARAMS value of adverbial *was* ('what') is generally stipulated to be lexically empty is not viable either because such an approach results in a semantic interpretation for the *was*-construction which equates the interpretation of polar questions, which is certainly not eligible. Also, the STORE value cannot be exploited to restrict adverbial *was* ('what') to an initial position and to prevent any other *wh*-phrase from occuring in-situ because firstly an empty STORE at clause-level is a requirement that is valid for all independent clauses including all clauses of sort *wh-interrogative-clause*, and is thus no specific requirement for adverbial *was*-constructions. Secondly, it is not obvious how the STORE value of the clause's head-daughter shall be restricted by the non-head daughter, especially considering the fact that non-*wh*-quantifiers might be regularly stored as well.

van Eynde (2004) enhances Ginzburg and Sags' account by proposing a local functor-driven treatment of the *wh*-property. The gist of his proposal is that all categories are either functors or heads, and functors select their head sisters via a head feature SELECT. In addition, he redefines objects of sort *category* as he introduces a MARKING feature having the values *marked* or *unmarked*. The fundamental architecture of this account is depicted in figure 3.

$$\begin{bmatrix} \textit{cat} & \\ \text{HEAD} & \begin{bmatrix}\textit{part-of-speech} \\ \text{SELECT} \;\; \text{canon-ss} \vee \text{none}\end{bmatrix} \\ \text{MARKING} & \textit{marking} \\ \text{SUBJ} & \textit{list}\left(\textit{synsem}\right) \\ \text{COMPS} & \textit{list}\left(\textit{synsem}\right)\end{bmatrix}$$

Figure 3: Redefinition of objects of type *category* according to van Eynde (2004)

The MARKING value propagates from the functor daughter to the mother in head-functor phrases, or otherwise from the head daughter. For our purposes van Eynde's definition of the WH feature is interesting: He redefines it as a boolean feature having the values *positive* or *negative* and being appropriate for objects of type *marking*. He further stipulates that all words are negatively marked for WH in the lexicon, except for the *wh*-words, which remain lexically underspecified as is shown in figure 4.

$$\begin{bmatrix} \text{PHON} \; \langle \textit{who} \rangle \\ \text{SYNSEM} \,|\, \text{LOCAL} \begin{bmatrix} \text{CAT} & \text{NP}\begin{bmatrix}\text{MARKING} & \begin{bmatrix}\textit{marked} \\ \text{WH} \;\; \alpha\end{bmatrix}\end{bmatrix} \\ \text{CONT} & \textit{parameter}\end{bmatrix}\end{bmatrix}$$

Figure 4: Partial lexical entry for a *wh*-word according to van Eynde (2004)

Since van Eynde adapts Ginzburg and Sags' Filler Inclusion Constraint by requiring that the non-head daughter of a *wh*-interrogative clause must be a sign with a positive WH value, underspecified *wh*-words are compatible with the Filler Inclusion Constraint. If a *wh*-phrase occurs in a left-peripheral position its WH value is positively specified. If a *wh*-phrase is used in situ, its WH value is negatively instantiated. This, however, leads to the result that *wh*-in-situ phrases are in terms of their WH specification not distinguishable from any other non-*wh*-phrase, which means again, that the examples in (2a) and (4a) cannot be excluded, and the grammar overgenerates.

The problem for all existing approaches to *wh*-interrogatives seems to be that they are all based on the assumption that any *wh*-phrase can in principle be fronted or realized in-situ. If we look at the adverbial *was*-construction, this assumption, however, seems to be wrong. Although adverbial *was* ('what') can mark a clause as *wh*-interrogative and functions in this respect as a normal *wh*-phrase, it is at the same time deficient in that it can neither be placed in-situ nor license another *wh*-phrase within the clause. In addition, it cannot be extracted and not be coordinated with an ordinary *wh*-phrase.

To account for these facts, I propose an analysis of the adverbial *was*-construction that is based on the following fundamental assumptions:

- There exist two *wh*-words *was* ('what') in the lexicon: a standard *wh*-pronoun that behaves like a typical argument *wh*-phrase, and a categorically deficient pronoun with adverbial function.
- The peculiarities of the adverbial *was*-construction follow from the deficiency of *was* ('what').
- Adverbial *was* ('what') is distinguished from canonical argument *was* ('what') as it is not analyzed as a filler-phrase, but acts as a functor in the sense of van Eynde (2004), and is thus restricted to a left-peripheral position.
- Two separate syntactic features, both representing *wh*-information, are exploited: one to treat *wh*-fronting and to handle pied piping; the other one to keep track of *wh*-in-situ phrases.

Before I develop this analysis in more detail, I will show next that adverbial *was* ('what') indeed behaves like a deficient pronoun, thereby presuming a theory that divides pronouns in different classes depending on their syntactic weight.

4 Adverbial *was* ('what') as a deficient pronoun

Research on the Germanic pronominal system goes back as far as to Koster (1978). In the course of this research it has been shown that pronouns are not homogeneous at all, but differ distributionally, prosodically, morphologically, syntactically and semantically.

It has been claimed that there exist pronouns with a special syntax in Germanic languages, which makes it necessary to distinguish them from standard pronouns which were called 'strong pronouns'. For instance, the 3rd person neuter pronoun *es* ('it') was taken as a typical example of such a non-canonical pronoun that grammatically differs from typical strong pronouns in German.

Furthermore, the thorough comparison of Germanic pronouns with the set of pronouns in Romance languages has led to the assumption that there are universal pronominal categories. Cardinaletti and Starke (1999) for instance develop an analysis of the languages' pronominal systems in terms of a three way distinction between 'strong', 'weak' and 'clitic pronouns'. They literally point out that the differences between these pronominal categories should be accounted for by a theory of featural deficiency. In addition, Cardinaletti and Starke (1999) claim that there is a ranking in deficiency between these subclasses: clitic pronouns are more deficient with respect to weak pronouns, which are in turn more deficient with respect to strong pronouns. Although the pronominal system of Germanic languages differs from the one of Romance languages in that the morphology of Germanic pronouns is often opaque, there seems to be evidence that a three-way split is indeed justified for Germanic. Haegemann (1999) for instance has demonstrated that the three classes are instantiated in the object pronominal system of West Flemish.

Although Cardinaletti and Starkes' proposal has been critized in several aspects (cf. van Riemsdijk (1999), as far as I know no-one has challenged the principle idea of a tripartite classification of pronouns so far. An even more fine-grained distinction between strong and weak categories has been proposed for instance by Abeillé and Godard (2003). In order to account for French adverbs they introduce a fourth category called 'light'. Since the argumentation here focuses on the fact that adverbial *was* ('what') is deficient or weak in comparison to canonical *wh*-words such as argument *was* ('what'), which generally behaves like a strong element, nothing specific of the proposed analysis here hinges on the question of which of the by now proposed classifications according to the weight of a syntactic cataegory is the more adequate one. The crucial claim made here is that adverbial *was* ('what') is peculiar in that it is not a strong, but a deficient *wh*-word belonging to the class of pronouns.

Evidence for this assumption can be derived from the following properties of adverbial *was* ('what') which correspond to the criteria that Cardinaletti and Starke (1999) and others generally apply to deficient pronouns.

(i) A deficient pronoun must occur at surface structure in a special derived position, which means that it cannot be found in a base position. This clearly applies to adverbial *was* ('what') as the contrast between (3b) and (3c) illustrates. If we compare adverbial *was* ('what') to the *wh*-expression *warum* ('why'), only *warum* ('why') can be positioned in the so-called German Mittelfeld, whereas *was* ('what') is restricted to a clause-initial position.

(ii) The contrast between (3b) and (3c) also supports a second general difference that Cardinaletti and Starke (1999) identified between strong and deficient pronouns as being a matter of distributional asymmetry: Compared to a strong pronoun, a deficient pronoun has an impoverished distribution. If one compares adverbial *was* ('what') with a strong *wh*-word such as *warum* ('why') the same distributional result is achieved.

(iii) A deficient pronoun is incompatible with coordination. The contrast (5a) vs. (5b) clearly demonstrates that this restriction applies to adverbial *was* ('what').

(iv) Contrary to a strong pronoun, a deficient pronoun is incompatible with modification. The following contrasts support the assumption that *was* ('what') behaves similarly.

(11) a. * Was genau schlägst du denn schon wieder den Hund?
what exactly beat you PART PART again the dog

b. Warum genau schlägst du denn schon wieder den Hund?
why exactly beat you PART PART again the dog
'Why exactly are you beating the dog again?'

(v) Deficient elements mostly occur unstressed, which is true for adverbial *was* ('what') if we take into account that *was* ('what') cannot bear a focus accent.

(vi) There is a semantic asymmetry between deficient and strong pronouns. Deficient pronouns are incapable of bearing their own range-restriction. The causal interpretation of adverbial *was* ('what') seems to be possible just because *was* ('what') is semantically underspecified. Therefore it is not surprising that there exists another non-canonical use of *was* ('what') which may occur in exclamative constructions like (12) (cf. d'Avis (2001)).

(12) Was DER seinen Hund schlägt!
what he his dog beats
'How (much) he beats his dog!'

In this case, *was* ('what') specifies a degree instead of a reason as *was* ('what') bears the meaning of *wie sehr* ('how'/'how much').

Taking these facts into account, it suggests itself to assume that adverbial *was* ('what') belongs to the class of deficient pronouns. Such an analysis seems to be superior to a conceivable alternative approach whereupon adverbial *was* ('what') is analyzed as a deviant *wh*-complementizer similar to *how come* in English as one of the reviewers proposed. Let me briefly motivate my view.

First of all, a pronominal status of adverbial *was* ('what') allows to put it in a row with interrogative, relative and indefinite *was* ('what') being homophonous to adverbial *was* ('what') and doubtlessly belonging to the class of pronouns, cf. Gallmann (1997).

Secondly, adverbial *was* ('what') does not behave like typical complementizers in German in two respects. One concerns the position of the finite verb. German as a verb second language possesses complementizers that—apart from few known exceptions—either select verbal phrases with the finite verb in final position or with the finite verb in second position. Adverbial *was* ('what'), however, may occur with both, verb final and verb initial clauses. In the latter case, *was* ('what') even occupies the so-called Vorfeld-position, which is usually not adequate for complementizers. Thus, in this respect *was* ('what') has nothing in common with an ordinary complementizer.

There is a third reason that militates against a complementizer analysis: The meaning of adverbial *was* ('what') as a potential complementizer would depend on its syntactic context. This results from a comparison of the data in (12) and (13).

(13) Was der seinen Hund schlägt?
what he his dog beats
'Why does he beat his dog?'

Was ('what') would function as a causal interrogative complementizer in case of (13), but as a modal exclamative complementizer in case of (12). An analysis that describes the meaning of a complementizer against the meaning of the constituent it combines with, however, is implausible if one considers that the meaning of a complementizer is normally lexically determined.

Last but not least, the fact in (9) is difficult to bring in line with a complementizer analysis. Modifiers like *zum Teufel* ('the devil') only adjoin to a WH-specified lexical item. Complementizers, however, are generally not WH-specified.

Taking these arguments into account it seems to be more fruitful to assume that adverbial *was* ('what') is a clause-initial *wh*-pronoun and not a complementizer as *how come* in English.

In the next section, an analysis of the causal *was*-construction is developed that adequately captures the presented facts.

5 An alternative approach

The fundamental ideas of the proposed analysis are (i) that adverbial *was* ('what') is forced to a clause-initial position and (ii) that two separate mechanisms keep track of the *wh*-property in a clausal phrase structure. Firstly, van Eynde's boolean WH feature is used to ensure that at least one *wh*-phrase is fronted in a *wh*-interrogative clause. And secondly, the QUE feature as defined by Pollard and Yoo (1998) is exploited to license *wh*-in-situ phrases. In the following, I want to explicate this approach in more detail.

To simplify matters I first extent the inventory of head values by differentiating the value *p-nouns* including all pronouns into *strong-p-noun*, *weak-p-noun* and *clitic* as is depicted in figure 5. Accordingly, adverbial *was* ('what') bears a HEAD value *weak-p-noun*. Assuming that the HEAD value of argument *was* ('what') is specified as *strong-p-noun* it is easy to explain why the coordination in example (6a) is ungrammatical. A categorial mismatch between weak and strong pronouns is responsible for the fact that adverbial *was* ('what') cannot be coordinated with a canonical *wh*-interrogative pronoun. Moreover, the aforementioned non-canonical prosodic properties of adverbial *was* ('what'), i.e. that adverbial *was* ('what') neither can be stressed nor focalized follow from the fact that adverbial *was* ('what') is analyzed as a deficient pronoun of type *weak-p-noun*.

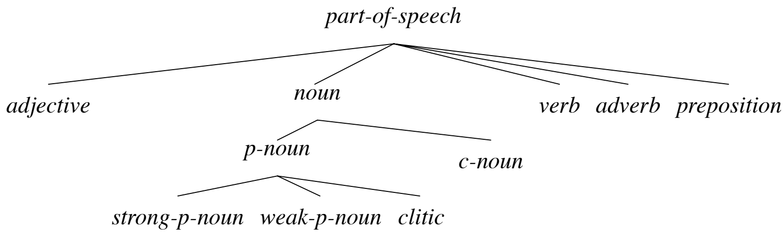


Figure 5: Partition of type *part-of-speech*

I further assume that the adverbial *was*-construction establishes a new interrogative construction type called *wh-functor-interrogative-clause* which inherits from both *interrogative-clause* and *head-functor-phrase* as is given in figure 6.

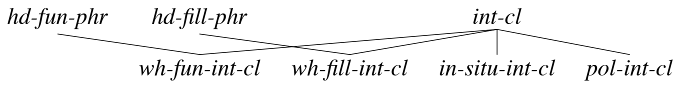


Figure 6: New interrogative construction type for adverbial *was*-construction

Adopting the feature architecture of van Eynde (2004) and in particular his functor treatment according to which functors are defined as signs which select their head sister, the WH feature is associated with the MARKING value as depicted in figure 7. The percolation of the WH value is constrained by van Eynde's

$$\begin{bmatrix} \textit{cat} \\ \text{MARKING} \mid \text{WH} \;\; \textit{wh} \end{bmatrix}$$

Figure 7: WH defined as a feature of type *marking*

Generalized Marking Principle saying that the MARKING value is propagated from the functor daughter if present or from the head daughter otherwise as can be seen in figure 8.

I further assume that a clause of type *wh-fun-int-cl*, which is the clause type used to describe adverbial *was*-constructions, is characterized by a functor daughter

$$\begin{bmatrix} \textit{hd-fun-ph} \\ \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{MARKING} \boxed{1}\ \textit{marking} \\ \text{DTRS} \left\langle \begin{bmatrix} \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \begin{bmatrix} \text{HEAD} \mid \text{SELECT} \boxed{2} \\ \text{MARKING} \boxed{1} \end{bmatrix} \end{bmatrix}, \boxed{3} \right\rangle \\ \text{HEAD-DTR} \boxed{3} \begin{bmatrix} \text{SYNSEM} \ \boxed{2}\ \textit{synsem} \end{bmatrix} \end{bmatrix}$$

Figure 8: Generalized Marking Principle following van Eynde (2004)

which belongs to the class of weak pronouns and whose WH value is positively specified. This restriction is formulated by the constraint given in figure 9.

$$\textit{wh-fun-int-cl} \Rightarrow \begin{bmatrix} \text{DTRS} \left\langle \begin{bmatrix} \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \begin{bmatrix} \text{HEAD} & \textit{weak-p-noun} \\ \text{MARKING} \mid \text{WH} & \textit{positive} \end{bmatrix} \end{bmatrix}, \boxed{1} \right\rangle \\ \text{HEAD-DTR} \boxed{1} \end{bmatrix}$$

Figure 9: Restrictions concerning the functor daughter of the new interrogative construction type for adverbial *was*-construction

Contrary to any other *wh*-word adverbial *was* (‘what’) is lexically treated as a deficient pronoun that bears a positively specified WH value. This follows from the lexical specification for *was* (‘what’), which is depicted in figure 10. Differing

$$\begin{bmatrix} \textit{word} \\ \text{PHON} \left\langle \textit{was} \right\rangle \\ \text{SS} \mid \text{LOC} \mid \text{CAT} \begin{bmatrix} \text{HEAD } \textit{weak-p-noun} \\ \text{MARKING} \mid \text{WH } \textit{positive} \end{bmatrix} \end{bmatrix}$$

Figure 10: Partial lexical entry for adverbial *was* (‘what’), part I

from adverbial *was* (‘what’), canonical *wh*-words such as argument *was* (‘what’) or adverbial *warum* (‘why’) are stipulated to remain lexically unspecified with respect to the WH value.

It follows from the assumptions regarding adverbial *was* (‘what’) that it may introduce a *wh*-interrogative clause since (i) it is compatible with the aforementioned constraint on objects of type *wh-fun-int-cl*, and (ii) it satisfies van Eynde’s Filler Inclusion Constraint (under the tacit assumption that this constraint has been extended to be applicable to clauses of type *wh-fun-int-cl*). The requirement of the Filler Inclusion Constraint that any fronted *wh*-phrase is specified as WH positive is lexically fulfilled in the case of adverbial *was* (‘what’). On the other hand, it is guaranteed that adverbial *was* (‘what’) cannot be placed in situ because in-situ phrases must bear a negatively specified WH value, which is only realizable for *wh*-words whose WH value is lexically unspecified.

The treatment of adverbial *was* (‘what’) as a functor in the sense of van Eynde (2004) allows adverbial *was* (‘what’) to have access to its sister constituent which is the head daughter of a clause of type *wh-fun-int-cl*. Again, this is captured in the lexicon. As figure 11 demonstrates, adverbial *was* ('what') selects a syntactically saturated and negatively WH specified sentential head daughter. Thus, it is impossible that a *wh*-in-situ phrase is realized at the same time. The functor treatment also allows to explain the extraction facts since functors resist extraction.

$$\begin{bmatrix} \textit{word} \\ \text{PHON} \left\langle \textit{was} \right\rangle \\ \text{SS} \mid \text{LOC} \mid \text{CAT} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \textit{weak-p-noun} \\ \text{SELECT} \begin{bmatrix} \textit{synsem} \\ \text{CAT} \begin{bmatrix} \text{HEAD } \textit{verb} \\ \text{SUBJ} \left\langle \right\rangle \\ \text{COMPS} \left\langle \right\rangle \\ \text{MARKING} \mid \text{WH } \textit{negative} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{MARKING} \mid \text{WH } \textit{positive} \end{bmatrix} \end{bmatrix}$$

Figure 11: Partial lexical entry for adverbial *was* ('what'), part II

However, one major problem of the previous accounts has not been solved, yet. If a *wh*-interrogative clause is introduced by adverbial *was* ('what'), so far nothing prevents any other *wh*-phrase to occur in situ, which contradicts the data in (4a). To account for the fact that adverbial *was* ('what')—contrary to argument *was*—does not license a *wh*-in-situ phrase, I suggest to employ the QUE value defined with Pollard and Yoo (1998) as a *synsem* feature. I assume that adverbial *was* ('what') is a functor that differs from any other *wh*-phrase in the requirement that it selects a saturated VP whose QUE value is instantiated by the empty set and thus may not contain a *wh*-element. Adverbial *was* ('what') itself has a filled QUE set which contains, depending on the theoretical setting, either a *wh*-quantifier or a restricted index in case it is stated that *wh*-words intrinsically lack a quantificational force. The partial lexical entry of adverbial *was* ('what') amended with this information is depicted in figure 12 on the next page.

Figure 13 on next page gives an example analysis: *Was* ('what') is analyzed as a deficient pronoun of type *weak-p-noun* lexically marked as WH *positive*. It introduces into the QSTORE a quantifier with a causal meaning. This quantifier is retrieved at the mother, which results in a question meaning of the whole construction. The verbal head daughter is specified as WH negative and has an empty QUE-value due to the selection properties of the functor daughter realized by *was* ('what').

6 Conclusion

I hope to have shown that the adverbial use of the *wh*-expression *was* ('what') establishes a new interrogative construction type which is based on the lexical prop-

$$\begin{bmatrix} \textit{word} \\ \text{PHON} \langle \textit{was} \rangle \\ \text{SS} \mid \text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{weak-p-noun} \\ \text{SELECT} & \begin{bmatrix} \textit{synsem} \\ \text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD } \textit{verb} \\ \text{SUBJ} \langle \rangle \\ \text{COMPS} \langle \rangle \\ \text{MARKING} \mid \text{WH } \textit{negative} \end{bmatrix} \\ \text{QUE } \{\} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{MARKING} \mid \text{WH } \textit{positive} \end{bmatrix} \\ \text{CONT } \boxed{3} \\ \text{QSTORE} & \left\{ \boxed{2} \begin{bmatrix} \text{DET } \textit{which} \\ \text{RESTIND } \boxed{3} & \begin{bmatrix} \text{INDEX } \boxed{1} \\ \text{RESTR} & \left\{ \begin{bmatrix} \text{QUANTS} \langle \rangle \\ \text{NUC} & \begin{bmatrix} \textit{reason} \\ \text{INST } \boxed{1} \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix} \right\} \\ \text{QUE} \{ \boxed{2} \} \end{bmatrix} \end{bmatrix}$$

Figure 12: Partial lexical entry for adverbial *was* ('what'), part III

$$\begin{bmatrix} \text{PHON} \langle \textit{was, schlägst, du, den Hund} \rangle \\ \text{SS} & \begin{bmatrix} \text{LOC} & \begin{bmatrix} \text{CAT} \mid \text{MARKING} \mid \text{WH } \textit{positive} \\ \text{QSTORE} \{\} \\ \text{QUANTS} \langle \{ \boxed{2} \} \rangle \end{bmatrix} \\ \text{QUE } \{\} \end{bmatrix} \end{bmatrix}$$

$$\begin{bmatrix} \text{PHON} \langle \textit{was} \rangle \\ \text{SS} & \begin{bmatrix} \text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{weak-p-noun} \\ \text{SELECT } \boxed{1} \end{bmatrix} \\ \text{MARKG} \mid \text{WH } \textit{positive} \end{bmatrix} \\ \text{QSTORE} \{ \boxed{2}\, \textit{reason} \} \end{bmatrix} \\ \text{QUE } \{ \boxed{2} \} \end{bmatrix} \end{bmatrix} \quad \begin{bmatrix} \text{PHON} \langle \textit{schlägst, du, den Hund} \rangle \\ \text{SS } \boxed{1} & \begin{bmatrix} \text{LOC} & \begin{bmatrix} \text{CAT} \mid \text{MARKG} \mid \text{WH } \textit{negative} \\ \text{QSTORE} \{\} \end{bmatrix} \\ \text{QUE } \{\} \end{bmatrix} \end{bmatrix}$$

Figure 13: Example analysis for example (1)

erties of a categorically deficient pronoun *was* ('what') that bears a causal meaning. By means of empirical facts on prosody, coordination, extraction, and *wh*-in-situ phenomena, I have argued that *was* ('what') behaves ambivalently regarding the *wh*-property: On the one hand, *was* ('what') may introduce a *wh*-interrogative clause, but on the other hand it neither can occur in situ nor can it license ordinary *wh*-in-situ phrases. This behaviour clearly contrasts to canonical *wh*-phrases.

To account for the presented facts, I have proposed an analysis that treats adverbial *was* ('what') as a functor selecting its head daughter. It follows from this analysis that adverbial *was* ('what') is forced to a clause-initial position. I have further argued that a proper analysis of the *was*-construction type necessitates a separation of two pieces of *wh*-information propagating in a *wh*-interrogative clause. I have suggested to exploit the *marking* feature WH for the modeling of *wh*-fronting and pied-piping phenomena, and the *local* feature QUE for the linking of syntactic and semantic information in such a way that it is possible to keep track of *wh*-in-situ phrases.

In this article I focussed on genuine grammatical aspects of the adverbial *was*-construction. Therefore, I had nothing to say about the peculiar pragmatic conditions that are related to the adverbial use of *was* ('what'). In particular, the construction seems to have some special speaker's inferences and might be related to the speaker's illocutionary force. I leave this issue open for future research.

References

Abeillé, Anne and Godard, Danièle. 2003. The Syntactic Flexibility of Adverbs: French Degree Adverbs. In Stefan Müller (ed.), *Proceedings of the HPSG-2003 Conference, Michigan State University, East Lansing*, pages 26–46, CSLI Publications.

Cardinaletti, Anna and Starke, Michal. 1999. The typology of structural deficiency: a case study of the three classes of pronouns. In Henk van Riemsdijk (ed.), *Clitics in the languages of Europe*, pages 145–233, Berlin: Mouton.

d'Avis, Franz-Josef. 2001. *Über w-Exklamativsätze im Deutschen*. Tübingen: Niemeyer.

Gallmann, Peter. 1997. Zur Morphosyntax und Lexik der w-Wörter. Technical Report, Sonderforschungsbereich 340, Tübingen, Stuttgart.

Ginzburg, Jonathan. 1992. *Questions, Queries and Facts: A Semantics and Pragmatics for Interrogatives*. Ph.D.thesis, Stanford University.

Ginzburg, Jonathan and Sag, Ivan. 2000. *Interrogative Investigations*. Stanford: CSLI.

Haegemann, Liliane. 1999. Cardinaletti and Starke: The typology of structural deficiency. Discussion: empirical support from West Flemish. In Henk van Riemsdijk (ed.), *Clitics in the languages of Europe*, pages 257–261, Berlin: Mouton.

Koster, Jan. 1978. *Locality Principles in Syntax*. Dordrecht: Foris.

Nakao, Cizuru and Obata, Miki. 2009. When 'What' Means 'Why': On Accusative wh-adjuncts in Japanese. Technical Report, University of Pennsylvania Working Papers in Linguistics.

Pollard, Carl and Sag, Ivan. 1994. *Head-driven Phrase Structure Grammar*. Stanford/Chicago: CSLI Publications and University of Chicago Press.

Pollard, Carl and Yoo, Eun. 1998. A unified theory of scope for quantifiers and wh-phrases. *Journal of Linguistics* 34, 415–445.

van Eynde, Frank. 2004. Pied piping is a Local Dependency. In Stefan Müller (ed.), *Proceedings of the HPSG-2004 Conference, Center for Computational Linguistics, Katholieke Universiteit Leuven*, pages 313–334, Stanford: CSLI Publications.

van Riemsdijk, Henk (ed.). 1999. *Clitics in the languages of Europe*. Berlin: Mouton.

Negative Concord in Romanian as Polyadic Quantification

Gianina Iordăchioaia and Frank Richter

Universität Stuttgart/Universität Tübingen

Proceedings of the HPSG09 Conference

Georg-August-Universität Göttingen, Germany

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

In this paper we develop an HPSG syntax-semantics of negative concord in Romanian. We show that n-words in Romanian can best be treated as negative quantifiers which may combine by resumption to form polyadic negative quantifiers. Optionality of resumption explains the existence of simple sentential negation readings alongside double negation readings. We solve the well-known problem of defining general semantic composition rules for translations of natural language expressions in a logical language with polyadic quantifiers by integrating our higher-order logic in Lexical Resource Semantics, whose constraint-based composition mechanisms directly support a systematic syntax-semantics for negative concord with polyadic quantification.

1 Introduction

We present an analysis of the syntax and semantics of the core of Romanian Negative Concord (NC) constructions as polyadic quantification in Lexical Resource Semantics (LRS, Richter and Sailer (2004)). Following a proposal by de Swart and Sag (2002) for French, we express the truth conditions associated with Romanian NC constructions by means of negative polyadic quantifiers. Going beyond de Swart and Sag's largely informal treatment of the logical representations for polyadic quantification in HPSG, we extend the logical representation language and modify the interface principles of LRS to accommodate polyadic quantifiers. This way we arrive at a theory of Romanian NC using resumptive polyadic quantifiers. Resumptive polyadic quantifiers are a notorious problem for frameworks which use the lambda calculus in combination with a functional theory of types to define a compositional semantics for natural languages. Our proposal of implementing them with LRS overcomes these fundamental logical limitations, and LRS is powerful enough to specify by standard HPSG devices a precise systematic relationship between a surface-oriented syntax and semantic representations with polyadic quantifiers.

Sentential negation in Romanian is usually expressed by the verbal prefix *nu* (Barbu (2004)). In the absence of other negative elements, *nu* contributes semantic negation (1a). If in addition an n-word such as *niciun* is present (1b), only a negative concord (NC) reading is available, a double negation (DN) interpretation is not. The negation marker (NM) *nu* is obligatory with n-words. In constructions with two n-words, both a NC reading and a DN reading are available (1c).[1]

[†]We would like to thank Janina Radó for proofreading and many suggestions. We also thank Danièle Godard, Doug Arnold and the audience of HPSG09 for stimulating comments and discussion.

[1]The DN reading in (1c) is dependent on a context in which one speaker formulates a negative proposition using the n-constituent *nicio carte* and another speaker denies that proposition by means of the n-constituent *niciun student*. See Iordăchioaia (2009, §3.4.2) for details.

(1) a. Un student **nu** a venit.
a student NM has come
'Some student didn't come.'

b. **Niciun** student *(**nu**) a venit.
no student NM has come
i. 'No student came.' (NC)
ii. # 'No student didn't come.' (DN)

c. **Niciun** student **nu** a citit **nicio** carte.
no student NM has read no book
i. 'No student read any book.' (NC)
ii. 'No student read no book.' (DN)

NC poses an immediate problem for composing the meaning of sentences from the meaning of their parts: Several apparently negative constituents are ultimately interpreted as single sentential negation. "NPI approaches" to NC solve this puzzle by postulating that n-words like the ones in (1b) and (1c) are in fact negative polarity items (NPIs) without inherent semantic negation (Ladusaw (1992)). Such theories, however, cannot account for the DN reading in (1c). (1c) together with (1b) suggests that (a) n-words are exponents of semantic negation, and (b) the negative marker *nu* does not contribute negation in the presence of n-words. As one of its main features, our syntax-semantics interface for Romanian NC acknowledges the lexically negative semantics of n-words and of the NM, and it captures under what circumstances the inherent negativity of the NM can be observed.

The remainder of the paper is structured as follows: First we discuss the data that lead us to conclude that Romanian n-words are indeed negative quantifiers (Section 2). Then we move on to the tools that we need to formulate our theory and extend the logical object language and the principles of LRS in such a way as to have resumptive polyadic quantifiers at our disposal (Section 3). The core of our theory of Romanian NC is presented in Section 4, where we formulate a language-specific principle that captures the properties of simple Romanian NC constructions. In Section 5 we show that our analysis can be extended in a straightforward way to more complex cases which involve scope properties of negative quantifiers in embedded subjunctive clauses. In the final section we briefly summarize the results and speculate about possible future developments.

2 Data

In this section we discuss evidence for the negative semantics of Romanian n-words and for their quantificational behavior. We focus on the properties of n-words in Romanian and on counterevidence for a treatment of Romanian n-words as NPIs. Alternative approaches to NC will not be considered here; a detailed discussion can be found in Iordăchioaia (2009).

NPI approaches to NC rest on two claims: (a) n-words lack negation, and (b) they are semantically licensed by an anti-additive operator (see below for an algebraic characterization of anti-additivity). Ladusaw (1992) argues that the semantic licenser of NPIs may be covert. This proposal has been widely exploited in the minimalist tradition (see, for instance, Zeijlstra (2004)), but is not available in a surface-oriented syntactic framework such as HPSG. Without the option of an empty syntactic operator, the only plausible licenser of n-words in a NC construction like (1b) is the NM. In Romanian the NM is usually obligatory with n-words, which has been interpreted as a consequence of its function as a semantic licenser. Analyses that adopt this view were formulated for Polish NC in Przepiórkowski and Kupść (1999) and Richter and Sailer (1999), and for Romanian in Ionescu (1999). We do not subscribe to this idea and will show instead that although the Romanian NM acts as a licenser for NPIs, it does not behave like a semantic licenser for n-words, and n-words do not need a semantic licenser, as they carry negation themselves.

According to Ladusaw, the semantic licenser of n-words must be at least anti-additive. A negative function f is anti-additive iff for each pair of sets X and Y, $f(X \cup Y) = f(X) \cap f(Y)$. In the absence of n-constituents, the NM *nu* receives an anti-additive interpretation (2):

(2) a. Studenţii **nu** au citit romane *sau* poezii.
students-the NM have read novels or poems
'The students haven't read novels or poems.'

b. = Studenţii **nu** au citit romane *şi* studenţii **nu** au citit poezii.
students-the NM have read novels and students-the NM have read poems
= 'The students haven't read novels and the students haven't read poems.'

If the disjunction that *nu* takes as argument contains n-words, anti-additivity disappears, and the two n-words are interpreted independently under the scope of negation (3):

(3) a. Studenţii **nu** au citit **niciun** roman *sau* **nicio** poezie.
students-the NM have read no novel or no poem
'The students read no novel or no poem.'

b. $\neq$ Studenţii **nu** au citit **niciun** roman *şi* studenţii **nu** au citit **nicio** poezie.
students-the NM have read no novel and students-the NM have read no poem
$\neq$ 'The students read no novel and the students read no poem.'

c. = Studenţii **nu** au citit **niciun** roman *sau* studenţii
students-the NM have read no novel or students-the
nu au citit **nicio** poezie.
NM have read no poem
= 'The students read no novel or the students read no poem.'

If the n-words in (3) are replaced with NPIs, the anti-additivity test succeeds. The contrast between (3) and (4) indicates that *nu* acts as licenser for NPIs but not for n-words.

(4) a. Studenţii **nu** au citit **vreun** roman *sau* **vreo** poezie.
students-the NM have read any novel or anyo poem
'The students didn't read any novel or any poem.'

b. = Studenţii **nu** au citit **vreun** roman *şi* studenţii **nu**
students-the NM have read any novel and students-the NM
au citit **vreo** poezie.
have read any poem
= 'The students didn't read any novel and the students didn't read any poem.'

Evidence for the inherent negativity of n-words comes from fragmentary answers (5a) and past participial constructions (5b), where n-words do not require the presence of the NM and contribute negation alone:

(5) a. A: Who was at the door?
B: **Nimeni**.
nobody

b. articol **de nimeni** citat
article by nobody cited
'article which hasn't been cited by anybody'

In these contexts n-words exhibit anti-additivity (6), and they can also license NPIs. The NPI *vreo* can be licensed by the anti-additive n-word *nimeni* but not by the universal quantifier *toată* (7).

(6) a. A: Who was at the door?
B: **Nimeni** cunoscut *sau* important. = **Nimeni** cunoscut *şi*
nobody known or important nobody known and
nimeni important.
nobody important

b. articol [de **nimeni** citat *sau* lăudat] = articol [de **nimeni** citat
article by nobody cited or praised article by nobody cited
şi de **nimeni** lăudat]
and by nobody praised

'article which hasn't been cited or praised by anybody'

(7) articol [de **nimeni**/*de toată lumea citat la *vreo* conferinţă]
article by nobody/by all people cited at any conference
'article which hasn't been cited by anybody at any conference'

The data in (6) and (7) clearly show that n-words carry negative semantics, which is hard to reconcile with the idea that they need a semantic licenser. Besides their negative content, n-words display scope properties that are similar to those of bona fide quantifiers and contrast with those of NPIs. We observe that n-words can build NC with a NM across a subjunctive clause boundary (8a), but not across a 'that' complementizer (8b). This behavior is paralleled by universal quantifiers, which can take wide scope over an operator in the matrix clause from an embedded subjunctive clause (9a), but not from an embedded 'that'-clause (9b).

(8) a. Ion **nu** a încercat să citească **nicio** carte.
John NM has tried SJ read no book
'John didn't try to read any book.'

b. Ion **nu** a zis că a citit *vreo*/***nicio** carte.
John NM has said that has read any/no book

(9) a. *Un* student a încercat să citească **fiecare** carte.
a student has tried SJ read every book
'Some student tried to read every book.'
i. $\exists > \forall$; ii. $\forall > \exists$

b. *Un* student a zis că a citit **fiecare** carte.
a student has said that has read every book
'Some student said that s/he read every book.'
i. $\exists > \forall$; ii. $\# \forall > \exists$

In addition, adjunct clauses and relative clauses block NC formation (10) and wide scope of embedded universal quantifiers (11), but not NPI licensing (10):

(10) a. **Nu** am dezvăluit secrete [care să-l fi expus pe
NM have revealed secrets that SJ-CL be exposed PE
***niciun**/*vreun* coleg].
no/any colleague
'I didn't reveal secrets that exposed any colleague.'

b. **Nu** am spus asta [pentru că mi-o ceruse ***niciun**/*vreun*
NM have said this because CL-CL asked no/any
prieten].
friend
'I didn't say that because any friend had asked me to.'

(11) a. *Un* student a dezvăluit secrete [care l-au expus pe
a student has revealed secrets that CL-have exposed PE
fiecare coleg].
every colleague
'Some student revealed secrets that exposed every colleague.'
i. $\exists > \forall$; ii. # $\forall > \exists$

b. *Un* student a spus asta [pentru că i-o ceruse **fiecare**
a student has said this because CL-CL asked every
prieten].
friend
'Some student said that because every friend had asked him to.'
i. $\exists > \forall$; ii. # $\forall > \exists$

The negative semantics and the quantificational properties of n-words explain the possibility of a DN reading with two n-words in (1c). The DN reading is the interpretation we expect with two negative quantifiers. In this respect there is no difference between the semantic status of n-words in Romanian and in DN languages like standard English or German, where DN is the only interpretation for two co-occurring n-constituents. What remains to be explained is the availability of the NC reading in (1c).

Following de Swart and Sag (2002), we analyze determiner n-words and negative NP constituents as quantifiers of Lindström type $\langle 1, 1\rangle$ and $\langle 1\rangle$, respectively (see Lindström (1966)). They may combine by resumption to form a polyadic quantifier of type $\langle 1^n, n\rangle$ or $\langle n\rangle$ (van Benthem (1989), Keenan and Westerståhl (1997), Peters and Westerståhl (2006)) and thus give rise to an NC interpretation. The negative marker *nu* is analyzed as a negative quantifier of type $\langle 0\rangle$ that is absorbed under resumption with other negative polyadic quantifiers. The relevant technical details will be sketched in our LRS implementation of polyadic quantification and resumption below.

3 LRS with Polyadic Quantifiers

For our analysis we need a higher-order logical language with negative polyadic quantifiers. Here we briefly outline its crucial properties and indicate how to integrate it with LRS.

We assume a simple type theory with types e and t. Functional types are formed in the usual way. The syntax of the logical language provides function application, lambda abstraction, equality and negative polyadic quantifiers. By standard results this is enough to express the usual logical connectives and monadic quantifiers. In reference to the simple type theory, we call our family of languages Ty1. Var and $Const$ are a countably infinite supply of variables and constants of each type:

Definition 1 Ty1 Terms*: Ty1 is the smallest set such that:*
$Var \subset Ty1$, $Const \subset Ty1$,
for each $\tau, \tau' \in Type$*, for each* $\alpha_{\tau\tau'}, \beta_\tau \in Ty1$*:*

$(\alpha_{\tau\tau'}\beta_\tau)_{\tau'} \in Ty1$,

for each $\tau, \tau' \in Type$*, for each* $i \in \mathbb{N}^+$*, for each* $v_{i,\tau} \in Var$*, for each* $\alpha_{\tau'} \in Ty1$*:*

$(\lambda v_{i,\tau}.\alpha_{\tau'})_{(\tau\tau')} \in Ty1$,

for each $\tau \in Type$*, and for each* $\alpha_\tau, \beta_\tau \in Ty1$*:*

$(\alpha_\tau = \beta_\tau)_t \in Ty1$,

for each $\tau \in Type$*, for each* $n \in \mathbb{N}^0$*, for each* $i_1, i_2, ..., i_n \in \mathbb{N}^+$*, for each* $v_{i_1,\tau}, v_{i_2,\tau}, ..., v_{i_n,\tau} \in Var$*, for each* $\alpha_{t1}, \alpha_{t2}, ..., \alpha_{tn}, \beta_t \in Ty1$*:*

$(NO(v_{i_1,\tau}, ..., v_{i_n,\tau})(\alpha_{t1}, ...\alpha_{tn})(\beta_t))_t \in Ty1$.

The standard constructs receive their usual interpretation. Here we only state the interpretation of negative polyadic quantifiers:

Definition 2 The Semantics of Ty1 Terms
(clause for negative polyadic quantifiers only)
For each model M *and for each variable assignment* $a \in Ass$*, for each* $\tau \in Type$*, for each* $n \in \mathbb{N}^0$*, for each* $i_1, i_2, ..., i_n \in \mathbb{N}^+$*, for each* $v_{i_1,\tau}, v_{i_2,\tau}, ..., v_{i_n,\tau} \in Var$*, for each* $\alpha_{t1}, \alpha_{t2}, ..., \alpha_{tn}, \beta_t \in Ty1$*:*

$[\![NO(v_{i_1,\tau}, ..., v_{i_n,\tau})(\alpha_{t1}, ..., \alpha_{tn})(\beta_t)]\!]^{M,a}$ *= 1 iff*
for every $d_{i_1}, d_{i_2}, ..., d_{i_n} \in D_{E,\tau}$,

$[\![\alpha_{t1}]\!]^{M,a[v_{i_1,\tau}/d_{i_1}]} = 0$ *or* $[\![\alpha_{t2}]\!]^{M,a[v_{i_2,\tau}/d_{i_2}]} = 0$ *or* ...

or $[\![\alpha_{tn}]\!]^{M,a[v_{i_n,\tau}/d_{i_n}]} = 0$ *or* $[\![\beta_t]\!]^{M,a[(v_{i_1},...,v_{i_n})/(d_{i_1},...,d_{i_n})]} = 0$.

(12) shows the truth conditions that we obtain for the translation of the Romanian counterparts of *John didn't come* (12a) and *No teacher didn't give no book to no student*, where all NPs are n-constituents and form a ternary negative quantifier by resumption (12b):

(12) a. For $n = 0$, $[\![NO()()(come'(j))]\!]^{M,a} = 1$ iff $[\![come'(j)]\!]^{M,a} = 0$

b. For $n = 3$, $v_{i_1} = x, v_{i_2} = y, v_{i_3} = z$, $\alpha_{t1} = teacher'(x)$, $\alpha_{t2} = book'(y)$, $\alpha_{t3} = student'(z)$ and $\beta_t = give'(x, y, z)$,
$[\![NO(x, y, z)(teacher'(x), book'(y), student'(z))$
$(give'(x, y, z))]\!]^{M,a} = 1$ iff for every $d_1, d_2, d_3 \in D_{E,e}$,

$$[\![teacher'(x)]\!]^{M,a[x/d_1]} = 0 \text{ or } [\![book'(y)]\!]^{M,a[y/d_2]} = 0 \text{ or}$$
$$[\![student'(z)]\!]^{M,a[z/d_3]} = 0 \text{ or}$$
$$[\![give'(x,y,z)]\!]^{M,a[(x,y,z)/(d_1,d_2,d_3)]} = 0$$

Minor adjustments suffice to integrate these logical representations in LRS. In the signature, the appropriateness of *gen-quantifier* of Richter and Kallmeyer (2009) is generalized to lists of variables (instead of single variables), and the restrictor of quantifiers now contains a list of expressions:

```
me  TYPE  type
    gen-quantifier  VAR  list
                    RESTR  list
                    SCOPE  me
```

A new statement in the theory of well-formed logical expressions (13) restricts polyadic generalized quantifiers to the form given in DEFINITION 1. The four relations mentioned in (13) are defined in such a way that they guarantee that $\boxed{1}$ is a list of variables, all variables have the same type $\boxed{3}$, the expressions in the list of restrictors $\boxed{2}$ are of type t, and there are exactly as many restrictor expressions as variables:

(13) $$\textit{gen-quantifier} \rightarrow \begin{bmatrix} \text{TYPE } \textit{truth} \\ \text{VAR } \boxed{1} \\ \text{RESTR } \boxed{2} \\ \text{SCOPE} | \text{ TYPE } \textit{truth} \end{bmatrix}$$
$$\wedge \texttt{variable-list}(\boxed{1}) \wedge \texttt{same-type-list}(\boxed{3},\boxed{1})$$
$$\wedge \texttt{truth-list}(\boxed{2}) \wedge \texttt{same-length}(\boxed{1}, \boxed{2})$$

We follow the usual notational conventions in LRS and often write descriptions of expressions of the semantic representation language as (partial) logical expressions. For describing polyadic quantifiers we use the notation $Q(\vec{v}, \vec{\phi}, \psi)$. Here $\vec{v}$ and $\vec{\phi}$ are shorthand for a (possibly empty) list of variables and a (possibly empty) list of expressions; ψ is a single expression. In the analysis of Romanian below we will assume that there is an appropriate subsort of *gen-quantifier* in our grammar which is interpreted as negative polyadic quantifier. In our notation this family of quantifiers will be denoted by $no(\vec{v}, \vec{\phi}, \psi)$.

The clause of the SEMANTICS PRINCIPLE governing the combination of quantificational determiners with nominal heads has to be adjusted to polyadic quantifiers. The relevant clause is shown in (14). Except for the generalization from monadic quantifiers to polyadic quantifiers, it is identical to the corresponding clause in (Richter and Kallmeyer, 2009, p. 65).

(14) THE SEMANTICS PRINCIPLE, Clause 1
If the non-head is a quantifier, then its INCONT value is of the form $Q(\vec{v}, \vec{\phi}, \psi)$, the INCONT value of the head is a component of a member[2]

[2] The symbol "$\triangleleft_{\in}$" is the infix notation of the new relation `subterm-of-member`, a generalized subterm relation.

of the list $\vec{\phi}$, and the INCONT value of the non-head daughter is identical to the EXCONT value of the head daughter:

$$\left[\text{DTRS}|\ \text{SPR-DTR}|\text{SS}|\ \text{LOC}\begin{bmatrix}\text{CAT}|\ \text{HEAD} & \textit{det}\\ \text{CONT}\ |\text{MAIN} & \textit{gen-quantifier}\end{bmatrix}\right] \longrightarrow$$

$$\left(\left[\text{DTRS}\begin{bmatrix}\text{H-DTR}\ |\text{LF} & \begin{bmatrix}\text{EXCONT} & \boxed{1}\\ \text{INCONT} & \boxed{2}\end{bmatrix}\\ \text{SPR-DTR}|\text{LF} & \left[\text{INCONT}\ \ \boxed{1}\begin{bmatrix}\textit{gen-quantifier}\\ \text{RESTR}\ \boxed{3}\end{bmatrix}\right]\end{bmatrix}\right] \wedge \boxed{2} \triangleleft_{\in} \boxed{3}\right)$$

Resumption will be implemented in LRS as identity of quantifiers contributed by lexical elements. For that reason no special technical apparatus for the resumption operation has to be introduced in preparation of our analysis of negative concord in Romanian in the next section.

With the integration of polyadic quantifiers and the modified clause of the SEMANTICS PRINCIPLE we have completed the adjustments in LRS needed to formulate our theory of NC. Before we turn to the analysis in the next section, we briefly review three standard LRS principles that will play a role in our examples. These are the LRS PROJECTION PRINCIPLE, the INCONT PRINCIPLE and the EXCONT PRINCIPLE. The LRS PROJECTION PRINCIPLE governs the relationship of the attribute values of EXCONT, INCONT and PARTS at phrases relative to their syntactic daughters. It is responsible for EXCONT and INCONT identity along syntactic head projections, and for the inheritance of the elements of PARTS lists by phrases from their daughters:

(15) LRS PROJECTION PRINCIPLE (Richter and Kallmeyer, 2009, pp. 47–48)
In each *phrase,*
1. the EXCONT values of the head and the mother are identical,
2. the INCONT values of the head and the mother are identical,
3. the PARTS value contains all and only the elements of the PARTS values of the daughters.

The INCONT PRINCIPLE and the EXCONT PRINCIPLE constrain the admissible values of the INCONT and the EXCONT attribute in syntactic structures. The INCONT PRINCIPLE is the simpler one of them. It guarantees two things: First, the internal content of a sign (the part of its semantics that is outscoped by any operator the sign combines with along its syntactic projection) is always semantically contributed by the sign, i.e. it is a member of its PARTS list. And second, the internal content is in the external content of a sign. In a first approximation (which is precise enough for our purposes) this means that the internal content contributes its semantics within the maximal syntactic projection of a sign.

(16) The INCONT PRINCIPLE (Richter and Kallmeyer, 2009, p. 47)
In each *lrs,* the INCONT value is an element of the PARTS list and a component of the EXCONT value.

The EXCONT PRINCIPLE is slightly more complex. Its first clause requires that the external content of a non-head daughter be semantically contributed from within the non-head-daughter. The second clause is a closure principle and says that the semantic representation of an utterance comprises all and only those pieces of semantic representations that are contributed by the lexical items in the utterance.

(17) The EXCONT PRINCIPLE (Richter and Kallmeyer, 2009, p. 47)
Clause 1:
In every phrase, the EXCONT value of the non-head daughter is an element of the non-head daughter's PARTS list.
Clause 2:
In every utterance, every subexpression of the EXCONT value of the utterance is an element of its PARTS list, and every element of the utterance's PARTS list is a subexpression of the EXCONT value.

The effects of these principles will be relevant for the examples in the next two sections.

4 The Analysis of Romanian NC

We will proceed in two steps. In Section 4.1 we lay out the analysis of sentential negation with the verbal prefix *nu* using a lexical rule. In Section 4.2 we turn to NC in simple sentences.

4.1 Sentential Negation

The analysis of simple negated sentences without n-constituents like (1a) follows immediately from the lexical analysis of verbs with the NM prefix *nu*. The affixal nature of *nu* is extensively argued for in Barbu (2004). Following assumptions similar to ours in Ionescu (1999) and the parallel analysis of the Polish negative marker in Przepiórkowski and Kupść (1997), we formulate the lexical rule in (18) that relates each verb form of the appropriate kind to a corresponding negated form.

(18) THE NM LEXICAL RULE

$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \boxed{4} \\ \text{SS}|\text{LOC}|\text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{VFORM} & \textit{fin} \vee \textit{inf} \\ \text{NEG} & - \end{bmatrix} \end{bmatrix} \\ \text{LF} & \begin{bmatrix} \text{EXCONT} & \boxed{0} \\ \text{INCONT} & \boxed{1} \\ \text{PARTS} & \boxed{2} \end{bmatrix} \end{bmatrix} \longmapsto \left(\begin{bmatrix} \text{PHON} & Neg(\boxed{4}) \\ \text{SS}|\text{LOC}|\text{CAT}|\text{HEAD} & \begin{bmatrix} \text{NEG} & + \end{bmatrix} \\ \text{LF}|\text{PARTS} & \boxed{2} \oplus \left\langle \boxed{3}\, no(\vec{u}, \vec{\gamma}, \delta) \right\rangle \end{bmatrix} \wedge \boxed{1} \triangleleft \delta \wedge \boxed{3} \triangleleft \boxed{0} \right)$$

The NM attaches to finite and infinitival verb forms as indicated by the VFORM value in (18). The boolean NEG feature value ensures that the NM is attached to a verb only once. All verb forms in the lexicon are specified as [NEG –] and may have a [NEG +] counterpart only if they undergo the lexical rule. The function Neg in the PHON value description of the output is responsible for the correct phonological forms with the verbal prefix. It permits reduction of *nu* to *n–* depending on the first phoneme in the input's verb form.

The semantic counterpart to the prefix *nu* in the phonological form is a negative quantifier on the verb's PARTS list, marked by the tag $\boxed{3}$ in the lexical rule. The interpretation of the verb form as negated is a consequence of the requirement that the internal content of the verb $\boxed{1}$ be a subterm of the nuclear scope δ of this quantifier ($\boxed{1} \triangleleft \delta$ in the output description of the lexical rule). The negative quantifier $\boxed{3}$ is also a subterm of the external content $\boxed{0}$ of the verb ($\boxed{3} \triangleleft \boxed{0}$). This condition will become important in the analysis of embedded clauses in Section 5 and will be responsible for the inability of the negation on an embedded verb form to outscope a matrix verb. As we will see later, negative quantifiers contributed by n-words in argument position will, under certain conditions, have the option of taking wide scope from embedded clauses.

The negative verb form *nu a venit* in our sentence (1a) is licensed by the NM LEXICAL RULE and shown below:

(19) *nu a venit* ('NM has come', licensed by the NM LEXICAL RULE)

$$\begin{bmatrix} \textit{word} \\ \mathrm{PHON} \; \langle \texttt{nu, a, venit} \rangle \\ \mathrm{SS} \mid \mathrm{LOC} \begin{bmatrix} \mathrm{CAT} \begin{bmatrix} \mathrm{HEAD} \mid \mathrm{NEG} \; + \\ \mathrm{VAL} \mid \mathrm{SUBJ} \; \langle \mathrm{NP}_{\boxed{1a}} \rangle \end{bmatrix} \\ \mathrm{CONT} \begin{bmatrix} \mathrm{INDEX} \mid \mathrm{VAR} \; \textit{no-var} \\ \mathrm{MAIN} \; \boxed{3a}\, come' \end{bmatrix} \end{bmatrix} \\ \mathrm{LF} \begin{bmatrix} \mathrm{EXCONT} \; \boxed{0} \\ \mathrm{INCONT} \; \boxed{3}\, come'(\boxed{1a}) \\ \mathrm{PARTS} \; \langle \boxed{3}, \boxed{3a}, \boxed{7}\, no(\vec{u}, \vec{\gamma}, \delta) \rangle \end{bmatrix} \end{bmatrix} \wedge \boxed{3} \triangleleft \boxed{0} \wedge \boxed{3} \triangleleft \delta \wedge \boxed{7} \triangleleft \boxed{0}$$

With standard LRS mechanisms in combination with a language-specific constraint that excludes the existential quantifier originating from *un student* from occurring in the immediate scope of negation, we obtain $some(x, student'(x), no((), (), come'(x)))$ as the truth condition for (1a). The variable and restrictor lists of the negative quantifier are empty (Lindström type $\langle 0 \rangle$) because the negative verb does not introduce a variable, and the sentence does not provide a restrictor.

4.2 NC Constructions

Determiner n-words contribute negative quantifiers of underspecified Lindström type $\langle 1^n, n \rangle$. In their LRS representation they lexically contribute exactly one new variable. The (relevant part of the) lexical entry of the determiner *niciun* exemplifies this pattern (20a). Unlike the negated verb in (19), *niciun* introduces a variable (x), and the negative quantifier $no(\vec{v}, \vec{\alpha}, \beta)$ binds x ($x \in \vec{v}$). In addition, the variable is a subterm of the nuclear scope ($x \triangleleft \beta$) and a subterm of a member in the restrictor list of the quantifier ($x \triangleleft_{\in} \vec{\alpha}$). These conditions guarantee the existence of a restrictor and prevent empty quantification.

(20) a.

$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \langle \texttt{niciun} \rangle \\ \text{SS} \mid \text{LOC} & \begin{bmatrix} \text{CAT} \mid \text{HEAD} & \begin{bmatrix} \textit{det} & \\ \text{SPEC} & \text{N}_{\boxed{1a}} \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \text{INDEX} \mid \text{VAR} & \boxed{1a}\, x \\ \text{MAIN} & \boxed{1}\, no(\vec{v}, \vec{\alpha}, \beta) \end{bmatrix} \end{bmatrix} \\ \text{LF} & \begin{bmatrix} \textit{lrs} & \\ \text{EXC} & \textit{me} \\ \text{INC} & \boxed{1}\, no(\vec{v}, \vec{\alpha}, \beta) \\ \text{PARTS} & \langle \boxed{1}, \boxed{1a}\, x \rangle \end{bmatrix} \end{bmatrix} \wedge x \in \vec{v} \wedge x \triangleleft_{\in} \vec{\alpha} \wedge x \triangleleft \beta$$

b.

$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \langle \texttt{student} \rangle \\ \text{SS}\ \boxed{7} \mid \text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{noun} \\ \text{VAL} \mid \text{SPR} & \left\langle \text{DetP}_{\boxed{1a}} \right\rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \text{INDEX} \mid \text{VAR} & \boxed{1a} \\ \text{MAIN} & \boxed{2a}\, student' \end{bmatrix} \end{bmatrix} \\ \text{LF} & \begin{bmatrix} \textit{lrs} & \\ \text{EXC} & \textit{gen-quantifier} \\ \text{INC} & \boxed{2}\, student'(\boxed{1a}) \\ \text{PARTS} & \langle \boxed{2}, \boxed{2a} \rangle \end{bmatrix} \end{bmatrix}$$

With the lexical entries of the determiner and the noun we have all necessary ingredients to investigate simple NC constructions with one n-word like sentence (1b). The relevant parts of the structure are shown in FIGURE 1.

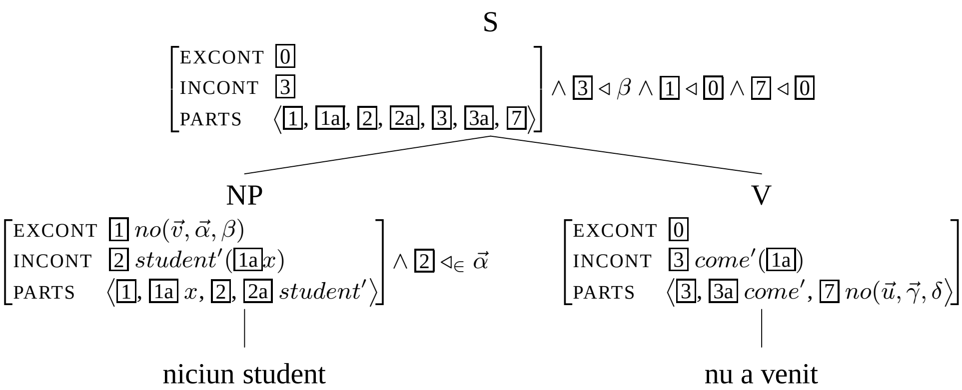


Figure 1: LRS analysis of (1b) *Niciun student nu a venit*

According to the LRS PROJECTION PRINCIPLE, the NP inherits the INCONT value $\boxed{2}$ of its nominal head. Due to the first clause of the SEMANTICS PRINCIPLE the internal content must be a subterm of a member of the restrictor list of the quantifier ($\boxed{2} \triangleleft_{\in} \vec{\alpha}$). The EXCONT value is identified with the INCONT value $\boxed{1}$ of the determiner due to the interaction of the first clause of the EXCONT PRINCIPLE with the other restrictions on the EXCONT of the NP. At the S node of the sentence two more restrictions become relevant. All lexically introduced pieces of semantic representation must be realized in the EXCONT of the sentence, including the EXCONT of the NP and the negative polyadic quantifier from the PARTS list of the verb ($\boxed{1} \triangleleft \boxed{0}$, $\boxed{7} \triangleleft \boxed{0}$). Moreover, the standard clause of the LRS SEMANTICS PRINCIPLE for combining NP-quantifiers in argument position with verbal projections requires that the polyadic quantifier of the NP take scope over the verb ($\boxed{3} \triangleleft \beta$).

All these restrictions together license three distinct expressions in the EXCONT of the sentence. Only one of them, shown in (21a), corresponds to the linguistic facts, the other two result from possible scope interactions of the negative quantifier of the verb and the NP-quantifier. The NC reading (21a) obtains if the two negative quantifiers get identified, meaning that $\boxed{1} = \boxed{7}$, $\vec{v} = \vec{u} = x$, $\vec{\alpha} = \vec{\gamma} = student'(x)$, and $\beta = \delta = come'(x)$.

(21) a. $no(x, student'(x), come'(x))$ $\boxed{0} = \boxed{1} = \boxed{7}$

b. $no(x, student'(x), no((), (), come'(x)))$ $\boxed{0} = \boxed{1}$; $\boxed{3} = \delta$; $\beta = \boxed{7}$

c. $no((), (), no(x, student'(x), come'(x)))$ $\boxed{0} = \boxed{7}$; $\boxed{3} = \beta$; $\delta = \boxed{1}$

(21b) and (21c) are impossible DN readings of (1b) and have to be excluded by the theory of Romanian NC. At the same time we have to take care that an n-word in a sentence obligatorily triggers the NM on the finite verb. We achieve both goals in one step by adapting the NEG CRITERION of Richter and Sailer (2004) to Romanian and the polyadic quantifier approach.

(22) THE NEG CRITERION for Romanian
If a negative quantifier of type higher than $\langle 0 \rangle$ outscopes a finite verb within the verb's external content, then the PARTS list of the verb must contain a negative quantifier of type higher than $\langle 0 \rangle$.

$$\forall \boxed{0}\, \forall \boxed{1}\, \forall \boxed{2} \left(\begin{bmatrix} \textit{word} \\ \text{SS}\,|\,\text{LOC} \begin{bmatrix} \text{CAT}\,|\,\text{HEAD} \begin{bmatrix} \textit{verb} \\ \text{VFORM} \; \textit{fin} \end{bmatrix} \\ \text{CONT}\,|\,\text{MAIN} \; \boxed{1} \end{bmatrix} \\ \text{LF}\,|\,\text{EXCONT} \; \boxed{0} \end{bmatrix} \wedge \boxed{2}\, no(\vec{v}, \vec{\alpha}, \beta) \triangleleft \boxed{0} \wedge \vec{v} \neq () \wedge \boxed{1} \triangleleft \beta \\ \rightarrow \exists \boxed{3}\, \exists \boxed{4} \left(\boxed{3}\, no(\vec{u}, \vec{\gamma}, \delta) \wedge \vec{u} \neq () \wedge \left[\text{LF}\,|\,\text{PARTS} \; \boxed{4} \right] \wedge \boxed{3} \in \boxed{4} \right) \right)$$

Intuitively, the NEG CRITERION says that the presence of an n-word in a sentence requires the presence of a (possibly different) n-word that undergoes resumption with the NM on the verb. More precisely, the NEG CRITERION is sensitive to the presence of a negative quantifier of a type higher than $\langle 0 \rangle$ in the EXCONT of a finite verb (contributed by at least one n-word). In that constellation a negative quantifier must also be on the PARTS list of the verb. Since those verbs that are licensed by lexical entries do not carry negative quantifiers in their PARTS lists, this means that only verbs licensed by the NM LEXICAL RULE are eligible. But since the quantifier contributed by a negative verb originally has an empty variable list, it would be of the excluded type $\langle 0 \rangle$ if it were not identified with a quantifier contributed by an n-word. It is due to the fact that the NEG CRITERION requires a quantifier of a type higher than $\langle 0 \rangle$ on the verb's PARTS list that identification with a quantifier from at least one n-word is necessary.

If we apply this reasoning to our example in FIGURE 1 we see that the negative quantifier contributed by the n-word and the negative quantifier on the PARTS list of the verb must be identical. We obtain an obligatory NC reading, and the other two readings in (21) are correctly ruled out.

In sentences with more than one n-word such as (1c), the negative quantifier contributed by the verb must undergo resumption with at least one of the two quantifiers contributed by the n-words for the reasons just described. If one n-word does not undergo resumption with the NM and the other n-word, we obtain the DN reading in (23a). However, there is also the possibility that all the negative quantifier contributions in the sentence are identified. The number of variables contributed by the individual n-words determines the type of the resumptive quantifier. For (1c) with two n-words, each contributing one variable, the second available alternative is resumption of all three negative quantifiers, which leads to a quantifier of type $\left\langle 1^2, 2 \right\rangle$ for the NC reading, shown in (23b).

(23) a. $no(x, student'(x), no(y, book'(y), read'(x, y)))$ (DN)

b. $no((x, y), (student'(x), book'(y)), read'(x, y))$ (NC)

5 N-words in Embedded Subjunctive Clauses

To complete our analysis, we investigate the function of the NM in NC constructions and show that our theory can be extended to account for locality conditions on the scope of negative quantifiers in NC constructions in complex sentences.

5.1 The NM as a Scope Marker

We argued that the NM cannot be a semantic licenser of n-words, as it does not maintain anti-additivity in the relevant contexts (3). We also saw that in NC constructions the negation contributed by the NM must always undergo resumption with at least one n-word, as decreed by the NEG CRITERION for Romanian (22). But if the NM is neither a semantic licenser, nor a real negation contributor in NC, what is its role in these constructions and why is it obligatory with n-words?

We think that an answer to these questions can be found in complex sentences like (24) where an n-word is contained in an argument phrase in an embedded subjunctive clause. In this kind of construction the negative quantifier may take wide scope over the matrix verb (24a) or narrow scope within the subjunctive clause (24b). Parallel observations hold for English n-words embedded in infinitival clauses (25). But unlike in the ambiguous English construction, in Romanian the scope of the quantifier is resolved by the (obligatory) NM: The scope of the negative quantifier is associated with the verb that carries the NM ((24a) vs. (24b)). We see that the NM functions as a *syntactic* licenser for n-words; the NM marks the sentential scope of the negative quantifier (cf. also Ionescu (1999, 2004)).

(24) a. Ion **nu** i-a cerut Mariei [să citească **nicio** carte].
John NM CL-has asked Mary SJ read no book
'There is no book that John asked Mary to read.'

b. Ion i-a cerut Mariei [să **nu** citească **nicio** carte].
John CL-has asked Mary SJ NM read no book
'John asked Mary not to read any book.'

(25) I will force you to marry **no one**. (Klima (1964, p. 285))

a. 'I wo**n't** force you to marry anyone.'

b. 'I would force you **not** to marry anyone.'

Assume that we augment the type theory of our semantic representation language by a type s for worlds and adjust the truth conditions of natural language expressions to Ty2 in the usual way. Moreover, assume for the moment that the EXCONT of matrix and embedded clause are distinct. With these modifications our theory captures (24a) and (24b).

In both sentences, independent LRS principles for quantifiers in argument position dictate that the negative quantifier associated with *nicio carte* must outscope the verb in the embedded clause. Let us look at (24a). Suppose *nicio carte* takes scope in the embedded clause. Then the NEG CRITERION is violated since the non-negated verb cannot have a negative quantifier on its PARTS list. Suppose it takes scope in the matrix clause. Then the NEG CRITERION is satisfied by resumption of the negative quantifier from *nicio carte* with the quantifier of the negated verb. We obtain the truth conditions $no(y, book'(y), ask'(john', mary', read'(mary', y)))$. The converse holds in (24b). The embedded verb has a negative marker and a negative quantifier on its PARTS, which means that *nicio carte* can take scope within the verb's EXCONT by resumption ($ask'(john', mary', no(y, book'(y), read'(mary', y))))$). It cannot take scope in the matrix clause, because the matrix verb lacks a negative quantifier on its PARTS list.

5.2 Complex Sentences with Two NMs

The situation becomes even more complex when both the matrix and the embedded verb in a complex sentence carry a NM:

(26) Ion **nu** i-a cerut Mariei [să **nu** citească **nicio** carte].
John NM CL-has asked Mary SJ NM read no book

a. ‘There is no book John asked Mary not to read.’

b. ‘John didn’t ask Mary not to read any book.’

The sentence (26) has two readings as indicated in the two translations. The negative quantifier *nicio carte* may enter in NC with the matrix verb (26a) or with the embedded verb (26b). In either case, the other verb contributes a type $\langle 0 \rangle$ negative quantifier to the interpretation. This means that one negation outscopes the other.

In preparation of our analysis of (26), we start with the simpler case of a complex sentence without n-word but with NM at the matrix verb and the embedded verb (27). The relevant parts of its analysis tree are shown in FIGURE 2.

(27) Ion **nu** i-a cerut Mariei [să **nu** citească *Nostalgia*].
John NM CL-has asked Mary SJ NM read nostalgia-the
‘John didn’t ask Mary not to read *The Nostalgia*.’

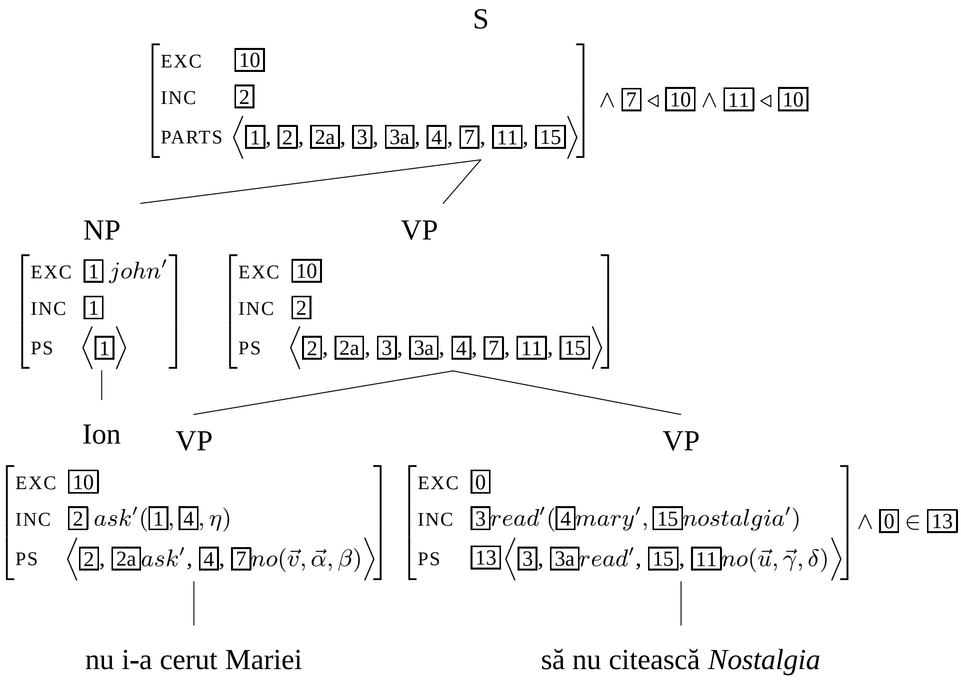


Figure 2: LRS analysis of (27) *Ion nu i-a cerut Mariei să nu citească* Nostalgia

The EXCONT of the non-head daughter VP on the right, which is the embedded subjunctive clause, must be an element of the PARTS list of that VP (EXCONT PRINCIPLE). The smallest piece of semantic representation which is eligible without violating any other LRS principles is the INCONT value $\boxed{3}$. The largest piece of semantic representation that the EXCONT $\boxed{0}$ of the embedded subjunctive clause can be identified with is the negative quantifier $\boxed{11}$, which is contributed by the verb *nu citească* and is licensed by the NM LEXICAL RULE. Since the lexical rule guarantees that this negative quantifier is a subterm of the external content of the verb (see (18)), we must conclude that $\boxed{0}$ equals $\boxed{11}$.

It may be suprising that nothing said so far prevents the negative quantifier of the embedded verb in FIGURE 2 from taking scope in the matrix sentence. The reason is that nothing forces the quantifier $\boxed{11}$ to take *immediate* scope over the predicate $\boxed{3}$, the matrix predicate may intervene. As a consequence, $\boxed{11}$ may be identified with the matrix negation or trigger DN within the matrix clause. Neither of the resulting semantic representations expresses possible truth conditions for the sentence in (27). As our analysis stands, a NM at an embedded verb could even outscope an affirmative matrix verb, giving the sentence in (28) the reading in (28b):

(28)

Ion	i-a	cerut	Mariei	[să	**nu**	citească	*Nostalgia*].
John	CL-has	asked	Mary	SJ	NM	read	nostalgia-the

a. ‘John asked Mary not to read *The Nostalgia*.’

b. # ‘John didn’t ask Mary to read *The Nostalgia*.’

A new clause of the SEMANTICS PRINCIPLE prevents this undesired effect and ensures that the external content of the complement clause of a propositional attitude verb remains within the scope of the matrix verb:

(29) THE SEMANTICS PRINCIPLE, Clause 2
If the head-daughter of a phrase has a MAIN value with a propositional argument η and the non-head-daughter is a propositional complement, then the EXCONT value of the complement must be a subterm of η.

In our example in FIGURE 2 the new clause of the SEMANTICS PRINCIPLE makes the EXCONT of the subjunctive clause $\boxed{0}$ a subterm of the scope η of the verb ask'. The negative quantifier $\boxed{11}$ contributed by the NM on the embedded verb is now a subterm of η and the only reading we obtain for (27) is the one in which both verbs are negated (30), as desired.

(30) $no((),(),ask'(john',mary',no((),(),read'(mary',nostalgia'))))$

Everything is now in place for the analysis of the two readings of the ambiguous sentence (26). A description of the tree structure is given in FIGURE 3. The only difference from FIGURE 2 is the negative quantifier in the embedded VP which takes the position of the proper name *Nostalgia*. For reasons of space, information carried by identical tags as in FIGURE 2 is not repeated in FIGURE 3.

There are three negative quantifiers whose scope interaction must be determined. The restriction $\boxed{0} \in \boxed{13}$ (known from the previous example) leaves two possibilities: $\boxed{0}$ could be identical with $\boxed{6}$ or with $\boxed{11}$. If $\boxed{0} = \boxed{6}$ we are in the situation

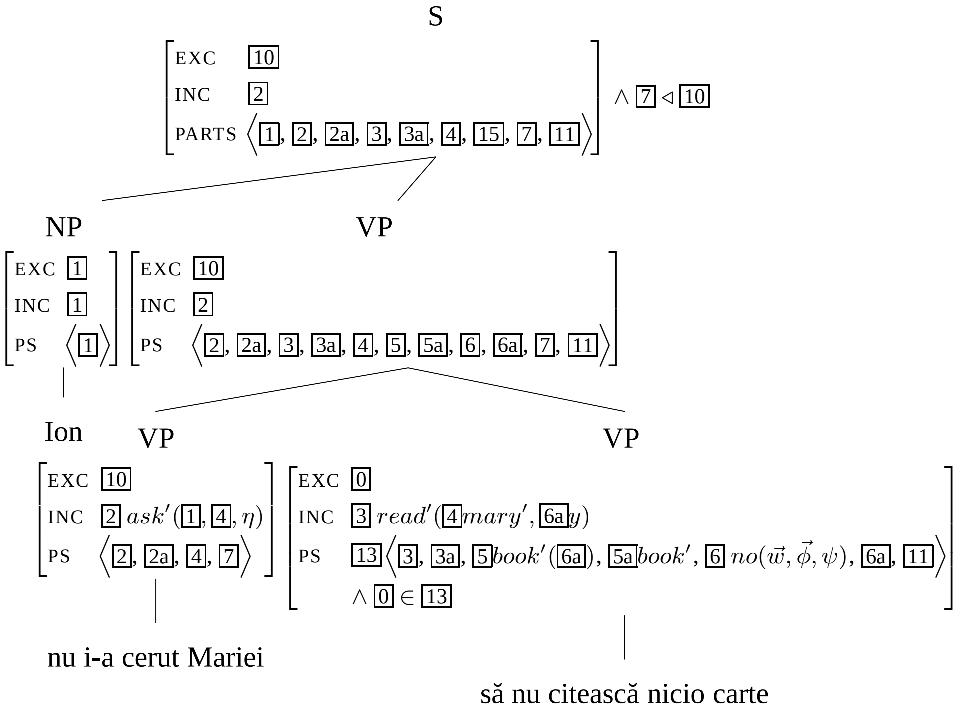


Figure 3: LRS analysis of (26) *Ion nu i-a cerut Mariei să nu citească nicio carte*

in which the negative quantifier $\boxed{6}$ of *niciun student* is interpreted in the embedded clause: Being identical with $\boxed{0}$ it is a subterm of η and cannot take scope in the matrix clause. On top of this, the NEG CRITERION forces resumption between $\boxed{6}$ and $\boxed{11}$, we obtain a NC reading in the subjunctive clause and the interpretation (31a) for (26). If $\boxed{0} = \boxed{11}$ the negative quantifier $\boxed{6}$ can take scope in the matrix clause where it undergoes resumption with $\boxed{7}$ to obey the NEG CRITERION. The result is a NC reading in the matrix clause and the interpretation (31b) for (26):

(31) a. $no((), (), ask'(john', mary', no(y, book'(y), read'(mary', y))))$
b. $no(y, book'(y), ask'(john', mary', no((), (), read'(mary', y))))$

In this section we showed that our theory of NC in Romanian contains all basic ingredients to account for the properties of negative quantifiers and NC in complex sentences. The analysis is still incomplete in at least two respects: We did not properly integrate our theory of polyadic quantifiers with two-sorted type theory; and we did not carefully consider the full range of data that is relevant for a theory of NC in complex sentences. While the logical extension should be straightforward, the empirical questions are challenging. What are the speakers' intuitions about the scope of negative quantifiers in complex sentences with two or more n-words? An unconstrained theory predicts scope interactions that native speakers most likely will not perceive given the usual difficulties with multiple negations. It would be important to find out which readings are available and preferred, and which grammatical or processing constraints are at play.

6 Conclusion

The present analysis of NC in Romanian applies the approach that was pioneered by an analysis of French in de Swart and Sag (2002). Our theory considerably extends de Swart and Sag's proposal by explicitly integrating a higher-order logic with polyadic quantification in HPSG. We expect that the formulation of the polyadic quantifier approach to NC in LRS will make it possible to unify this line of research with the typological approach to NC in Polish, French and German presented in Richter and Sailer (2006). Last but not least, adding polyadic quantification to LRS opens the door to exploring a whole range of new semantic phenomena in HPSG such as cumulative and *same/different* (unreducible) polyadic quantifiers (Keenan (1992), Keenan and Westerståhl (1997)). Since our constraint-based syntax-semantics interface supports the integration of polyadic quantifiers, HPSG theories can take full advantage of them. This brings within reach an explicit specification of the syntax and semantics of constructions that require unreducible polyadic quantifiers for an adequate rendering of their truth conditions and have, for that reason, turned out to be problematic in other grammar frameworks.

References

Barbu, Ana-Maria. 2004. The negation NU: lexical or affixal item. In Emil Ionescu (ed.), *Understanding Romanian Negation. Syntactic and Semantic Approaches in a Declarative Perspective*, pages 68–82, Bucharest University Press.

Ionescu, Emil. 1999. A Quantification-based Approach to Negative Concord in Romanian. In Geert-Jan M. Kruijff and Richard T. Oehrle (eds.), *Proceedings of Formal Grammar 1999*, pages 25–35, Utrecht.

Ionescu, Emil. 2004. The Semantic Status of Romanian N-words in Negative Concord Constructions. In Emil Ionescu (ed.), *Understanding Romanian Negation. Syntactic and Semantic Approaches in a Declarative Perspective*, pages 83–118, Bucharest University Press.

Iordăchioaia, Gianina. 2009. *Negative Concord with Negative Quantifiers: A Polyadic Quantifier Approach to Romanian Negative Concord*. Ph. D. thesis, submitted, University of Tübingen.

Keenan, Edward L. 1992. Beyond the Frege Boundary. *Linguistics and Philosophy* 15, 199–221.

Keenan, Edward L. and Westerståhl, Dag. 1997. Generalized Quantifiers in Linguistics and Logic. In Johan van Benthem and Alice ter Meulen (eds.), *Handbook of Language and Logic*, pages 837–893, Amsterdam: Elsevier Science.

Klima, Edward. 1964. Negation in English. In Jerry A. Fodor and Jerrold J. Katz (eds.), *The Structure of Language*, pages 246–323, Englewood Cliffs: Prentice Hall.

Ladusaw, William A. 1992. Expressing negation. In *Proceedings of SALT 2*, pages 237–259, Columbus: The Ohio State University.

Lindström, Per. 1966. First Order Predicate Logic with Generalized Quantifiers. *Theoria* 32, 186–195.

Peters, Stanley and Westerståhl, Dag. 2006. *Quantifiers in Language and Logic*. Oxford and New York: Oxford University Press.

Przepiórkowski, Adam and Kupść, Anna. 1997. Negative Concord in Polish. Technical Report, Institute of Computer Science, Polish Academy of Sciences.

Przepiórkowski, Adam and Kupść, Anna. 1999. Eventuality Negation and Negative Concord in Polish and Italian. In Robert Borsley and Adam Przepiórkowski (eds.), *Slavic in HPSG*, pages 211–246, Stanford: CSLI Publications.

Richter, Frank and Kallmeyer, Laura. 2009. Feature Logic-based Semantic Composition: A Comparison between LRS and LTAG. In Anders Søgaard and Petter Haugereid (eds.), *Typed Feature Structure Grammars*, pages 32–92, Frankfurt a.M.: Peter Lang.

Richter, Frank and Sailer, Manfred. 1999. LF Conditions on Expressions of Ty2: An HPSG Analysis of Negative Concord in Polish. In Robert Borsley and Adam Przepiórkowski (eds.), *Slavic in HPSG*, pages 247–282, Stanford: CSLI Publications.

Richter, Frank and Sailer, Manfred. 2004. Basic Concepts of Lexical Resource Semantics. In *ESSLLI 2003 – Course Material I*, volume 5 of *Collegium Logicum*, Kurt Gödel Society Wien.

Richter, Frank and Sailer, Manfred. 2006. Modeling Typological Markedness in Semantics: The Case of Negative Concord. In Stefan Müller (ed.), *Proceedings of the 13th International Conference on Head-Driven Phrase Structure Grammar*, pages 305–325, CSLI Publications.

de Swart, Henriëtte and Sag, Ivan A. 2002. Negation and Negative Concord in Romance. *Linguistics and Philosophy* 25, 373–417.

van Benthem, Johan. 1989. Polyadic Quantifiers. *Linguistics and Philosophy* 12, 437–464.

Zeijlstra, Hedde H. 2004. *Sentential Negation and Negative Concord*. Utrecht: LOT Publications.

How Hard a Problem Would This Be to Solve?

Paul Kay (University of California, Berkeley)
and
Ivan A. Sag (Stanford University)

Proceedings of the HPSG09 Conference

Department of Linguistics, Stanford University

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

This paper analyzes the interrelation of two understudied phenomena of English: discontinuous modifier phenomenon (**so** *willing to help out* **that they called early**; **more** *ready for what was coming* **than I was**) and the complex pre-determination phenomenon (*this delicious a lasagna*; *How hard a problem* (*was it*)*?*). Despite their independence, they frequently occur intertwined, as in **too** *heavy a trunk* **(for me) to lift** and **so** *lovely a melody* **that some people cried**. This paper presents a declarative analysis of these and related facts that avoids syntactic movement in favor of monotonic constraint satisfaction. It demonstrates how an explicit, sign-based, constructional approach to grammatical structure captures linguistic generalizations, while at the same time accounting for idiosyncratic facts in this seemingly complex grammatical domain.

1 Introduction

Two understudied phenomena of English are intimately intertwined but, insofar as they are studied at all, are not usually related. The discontinuous dependent phenomenon (DD) illustrated in (1) and the complex pre-determination (CPD) phenomenon illustrated in (2)[1] are independent. That is, each of these phenomena may occur independently of the other:

(1) a. [[**so** willing to help out] **that they called early**]

b. [[[**too** far] behind on points] **to quit**]

c. [[[**more** ready] for what was coming] **than I was**]

d. [[**as** prepared for the worst] **as anyone**]

e. [[the **same** courage in the face of adversity] **as yours**]

(2) a. [[**this** delicious] a lasagna]. . .

b. [[**that** friendly] a policeman]. . .

c. [[**How** hard] a problem] (was it)?

d. [**What** a fiasco] (it was)!

[†]For their helpful comments and/or discussion regarding the ideas presented here, we would like to thank Charles Fillmore, Dan Flickinger, Laura Michaelis, Chris Potts, Stefan Müller, Peter Sells and Frank Van Eynde.

[1]CPD is also known as the "Big Mess" Construction. See Berman 1974, Arnold and Sadler 1992, and Van Eynde 2007.

The oddity (the "non-core" property) of DD examples like those in (1) is that they appear to call for a discontinuous constituent analysis. The oddity of CPD examples like those in (2) is that they present an adjective modifying an NP (or DP), rather than a nominal (a common noun phrase or "$\overline{\mathrm{N}}$") – specifically an NP determined by the singular indefinite article *a*.

Although, as we have seen in (1) and (2), DD and CPD may appear independently, they frequently occur intertwined as in (3):

(3) a. [[[**too** heavy] a trunk] **(for me) to lift**]

b. [[[**so** lovely] a melody] **that some people cried**]

c. [[[**more** sincere] an apology] **than her critics acknowledged**]

d. [[[**as** good] a singer] **as many professionals**]

Unsurprisingly, the initial lexical licenser determines the three-way distributional distinction displayed in (1), (2) and (3).

Licensers of DD but not CPD include those comparative governors listed in (4):[2]

(4) *same...as, similar...to, equal...to/with, identical... to/with,* ADJ-*er...than, rather...than, ...else than, ...enough that, ...other than*

Complement-selecting adjectives, verbs, and nouns also participate in DD, as we will see. Licensers of CPD but not DD include:

(5) *this, that, how*

And licensers of both DD and CPD are listed, exhaustively we believe, in (6):[3]

(6) *so, too, more, less, as, such*

It is notable that comparative licensers are split between those that do not [(4)] and those that do [(6)] license CPD. There are licensers of CPD but not DD, DD but not CPD, and both DD and CPD.

More than one DD can occur in a clause, as exemplified in (7).

(7) a. **so** much $\underline{\textit{more}}$ satisfied $\underline{\textit{than the last time}}$ **that he couldn't stop smiling**

b. [[[**too** many $\underline{\textit{fewer}}$] supporters] $\underline{\textit{than her opponent}}$ **(for her) to rely on appeals to her base**]

[2]See Huddleston and Pullum 2002: 1104.

[3]It should be noted that *such* is different from the other adjective specifiers in (6). In particular, *such*, like exclamative *what*, functions essentially as the portmanteau of a specifier and an adjective.

c. [[[[**enough** *bigger*] an audience] *than last time*] **to require standing room only**]

In examples such as (7) the multiple DDs form nested dependencies. The corresponding crossed dependencies in (8) are impossible:

(8) a. ***so** much *more* satisfied **that he couldn't stop smiling** *than the last time*

b. ***too** many *fewer* supporters **(for her) to rely on appeals to her base** *than her opponent*

c. ***enough** *bigger* an audience **to require standing room only** *than last time*

Other DDs may, however, participate with arguments or modifiers in either nested [(9b,d)] or crossed [(9a,c)] dependencies:

(9) a. Kim was [[[**more** *willing*] **than Pat is**] *to wash the dishes*].

b. Kim [[[is [**more** *willing*] now] *to wash the dishes*] **than Pat is**].

c. I [[[sent out [**more** *books*] yesterday] **than ever before**] *that I really liked*].

d. I [[[sent out [**more** *books*] yesterday] *that I really liked*] **than ever before**].

In general,

(10) **All DD licensers except** *so, too,* **and** *enough* **can participate in crossed dependencies with arguments and other dependents.**

We will need to formulate the lexical entries for the licensers and, critically, the relevant phrasal constructions, in such a way as to account for all the above facts, plus some more to be mentioned.

2 Previous Proposals

There are no fully worked out analyses of DD in the syntactic literature, though there are discussions of various aspects of DD. Perhaps the most detailed of these proposal is due to Chae (1992), who extends the GPSG analysis of gap-binding by allowing a word like *too* to transmit its gap-binding potential to a higher node, e.g. to the adjective phrase *too hot* in examples like (11):

(11) This is [[**too** hot] [**to touch __**]]$_{\mathrm{AP}}$.

Binding of the gap takes place when a nonempty SLASH specification and its appropriate licensing specification are both passed up to the same point in the tree, i.e. the AP labelled in (11).

Flickinger and Nerbonne (1992) analyze examples like (12), proposing to allow SUBCAT information to be inherited from multiple daughters in structures like (12):

(12) An [[**easy** man] [**to please** _]]$_{\overline{\mathrm{N}}}$ ...

On their proposal, an $\overline{\mathrm{N}}$ like *easy man* inherits its subcateogrization potential from both *easy* and *man* and hence can select *to please* _ as a complement.

The EXTRA feature was first proposed by Pollard in unpublished work and appears briefly in Pollard and Sag's (1994, p. 368) sketch of extraposition in comparative phrases. Subsequent analyses using the EXTRA feature to analyze various extraposition phenomena in English and German include Keller 1995, Van Eynde 1996, Bouma 1996, Kim and Sag 2005 and Crysmann to appear.

Kiss (2005; see also Wittenburg 1987) treats German relative clause extraposition as an anaphoric dependency, rather than a syntactic one, introducing a feature ANCHORS to pass up a set of indices from NPs within a given phrase, each of which can be associated with an extraposed relative clause at a higher level of structure. See Müller 2004 and Crysmann to appear for assessments of the various alternative approaches.

CPD has been discussed by many researchers in the transformational literature, culminating perhaps in the work of Kennedy and Merchant (2000), who provide a useful review and a comprehensive proposal that even addresses complex predeterminers with *of* (e.g. *how much of a difference*), which we cannot discuss here. However, their proposal is stated in terms of complex structures, a rich array of empty categories, and movement operations whose control they are unable to specify. In particular, as they note (cf. their footnote 28), their analysis seems to require appeal to an unformulated constraint on phonetic form in order to account for the most basic facts of CPD, i.e. the contrasts given in (13) below.

The most successful analysis of CPD to date, in our view, is that of Van Eynde (2007).[4] A key aspect of this analysis, which we follow here in the main, is the replacement of Pollard and Sag's (1994) features MOD and SPEC by the single feature SELECT (SEL). The SEL analysis allows Pollard and Sag's SPR feature to be eliminated, as well.

None of the proposals just mentioned provides a treatment of the interaction of DD and CPD. It turns out, however, that this interaction will follow straightforwardly from the analysis we propose here.

3 Analysis

In this paper, we will employ **Sign-Based Construction Grammar** (SBCG), a version of HPSG that blends in key elements of Berkeley Construction Grammar, of the sort developed in such works as Fillmore et al. 1988, Michaelis and Lambrecht 1996, Fillmore 1999, Kay and Fillmore 1999, and Kay 2002. For a more detailed exposition of SBCG than can be presented here, the reader is referred to Sag in press, 2010, and other papers in Boas and Sag 2010.

[4]This is an outgrowth of earlier work by Van Eynde (1998), which in turn builds directly on Allegranza 1998. See also Van Eynde 2006 and Allegranza 2007.

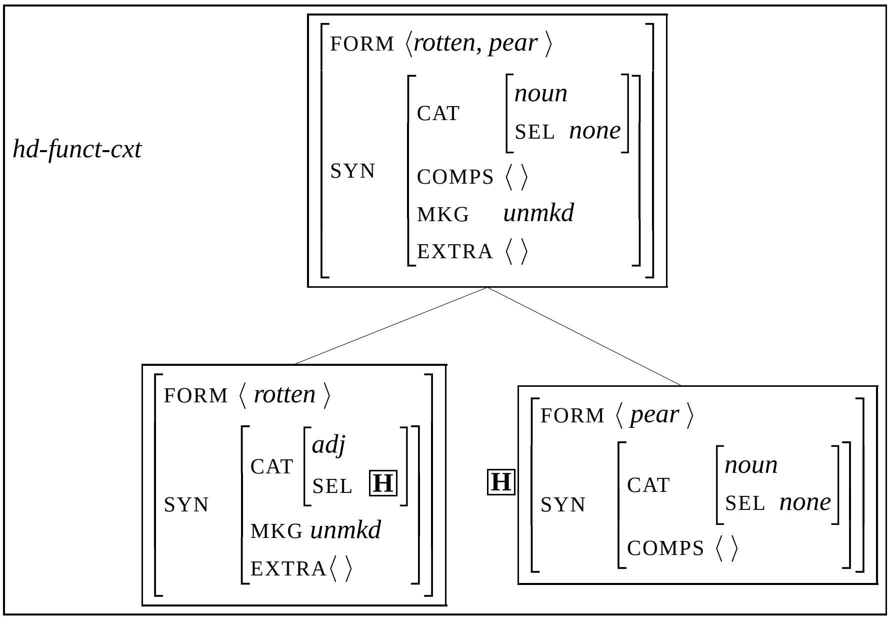


Figure 1: A Head-Functor Construct

In the introduction, we sketched a few of the more salient distributional facts about DDs. We begin the more analytical discussion with CPD structures, as illustrated in (2) and (3). As already noted, the interesting property of these structures is that they contain adjective phrases modifying determined NPs, rather than the usual adjectival modification of undetermined common nominal expressions (CNPs), as illustrated in (13):

(13) a. a [rotten pear] (cf. *rotten a pear)

b. a [mere bagatelle] (cf. *mere a bagatelle)

c. the [old book]

d. her [seven [lonely nights]]

The SBCG representation of the bracketed expression in (13a), a feature structure of type *head-functor-construct*, is given in Figure 1.[5] Beginning with the first daughter (specified as [FORM ⟨*rotten*⟩]) we note that the SYN value has three attributes: CAT, MKG and EXTRA. As indicated, the CAT(EGORY) value is a feature structure of type *adj(ective)*. This feature structure includes a specification for the

[5] We use familiar HPSG notation for our grammatical descriptions. Resolved feature structure models, by contrast, are presented as boxed attribute-value matrices. Boxed tree structures indicate fully resolved feature structures of (some subtype of) the type *construct*. These are functions from the domain {MTR, DTRS}, where MTR (MOTHER) is *sign*-valued and the value of DTRS (DAUGHTERS) is a list of *sign*s.

feature SEL, whose value is represented by the tag **H**, indicating that this value has been equated with the value of another feature in the same diagram. This analysis provides a unified treatment of modifiers, specifiers, determiners and other "markers" in terms of lexically varying specifications for the SEL feature, which in turn correspond to the varying possibilities for (in this construction) the second daughter. The MKG (MARKING) value of the first daughter, *unmkd* (*unmarked*), reflects the fact that adjectives are so specified lexically. And following Van Eynde (2007), the mother's MKG value is identified with that of the functor daughter.[6]

The EXTRA feature plays a central role in the present discussion. It is a nonlocal, list-valued feature that provides the mechanism for a wide range of extrapositions (in line with the arguments offered by Keller, Van Eynde, and Bouma), including those illustrated in (14):[7]

(14) a. **It** seems **that your hair is burning**.
(extraposition from subject)

b. They regret **it** very much **that we could not hire Mosconi**.
(extraposition from object)

c. I am **unwilling** when sober **to sign any such petition**.
(extraposition of VP complement)

d. He **lowered** the nitro bottle gently **onto the floor**.
(extraposition of PP complement)

e. **An article** appeared yesterday **about the situation is Kazakhstan**.
(extraposition of PP modifier)

f. **A man** walked in **who was wearing striped suspenders**.
(extraposition of relative clause)

The EXTRA feature thus works much like SLASH (GAP): A lexical entry or lexical construction requires an item on the EXTRA list of a sign. When this sign serves as the daughter of some phrasal construct, its non-empty EXTRA specification becomes part of the mother's EXTRA list and this continues until a higher structure (a *head-extra-construct*) realizes the item as a constituent sign whose mother's EXTRA list is free of the now realized ("extraposed") item. We will see how this works in detail below. For the moment we note that in a *hd-func-cxt* like *rotten pear*, the mother inherits the EXTRA value from the non-head (functor) daughter.

The second daughter ([FORM ⟨*pear*⟩]) is the head daughter, as indicated by the boxed **H** preceding the outer brackets. Its CAT value, as indicated, is a feature structure of type *noun* and its COMPS value is the empty list. The mother sign

[6]Note that the features LOCAL, NONLOCAL, and HEAD are not just being supressed in our displays. They have in fact been eliminated from the grammar.

[7]We will not attempt to establish this broad claim in the present paper, but we intend the EXTRA feature and the constructions that mention it eventually to cover all the data in (14).

$$
\textit{hd-func-cxt} \Rightarrow \begin{bmatrix} \textit{hd-cxt} & \\ \textsc{mtr} & \begin{bmatrix} \textsc{syn} & \begin{bmatrix} \textsc{comps} & L_1 \\ \textsc{mkg} & Y \\ \textsc{extra} & L_2 \end{bmatrix} \end{bmatrix} \\ \textsc{dtrs} & \left\langle \begin{bmatrix} \textsc{syn} & \begin{bmatrix} \textsc{cat} & [\textsc{sel}\ H\,] \\ \textsc{mkg} & Y \\ \textsc{extra} & L_2 \end{bmatrix} \end{bmatrix}, H{:}\begin{bmatrix} \textsc{syn} & [\textsc{comps}\ L_1] \end{bmatrix} \right\rangle \\ \textsc{hd-dtr} & H \end{bmatrix}
$$

Figure 2: Head-Functor Construction

([FORM ⟨*rotten*, *pear*⟩]) of this construct inherits its CAT and COMPS specifications from the head daughter and its MKG and EXTRA values from the functor (non-head) daughter. The construction that licenses this construct is the Head-Functor Construction, shown in Figure 2.[8,9]

This construction specifies the inheritance by the mother of the MKG and EXTRA values from the functor daughter that we observed in the *rotten pear* construct in Figure 1. It also specifies the inheritance by the mother of the COMPS value from the head daughter. The identification of the mother and head-daughter's CAT values is of course absent from (14), since head-functor constructs are a subtype of *headed-construct* (*hd-cxt*), which in turn is constrained by the Head Feature Principle, which guarantees that (in any headed construct) the head daughter's CAT value is identical to the CAT value of its mother. The Head Functor Construction thus licenses adjectivally modified nominals and determined noun phrases, among other local structures.

We now turn our attention to the CPD phenomenon we illustrated in (2)–(3) above. We cannot use the Head-Functor Construction to license CPD noun phrases like [[*so big*] [*a mess*]], because (1) ordinary adjectives, like *big* or *rotten*, select only undetermined nominals, as illustrated in (13a,b), and (2) since SEL is a CAT feature, the Head-Functor Construction would incorrectly require that the mother's SEL value be the same as that of the head daughter.

Van Eynde (2007) has proposed a constructional HPSG solution at the level of the NP. That is, to license a noun phrase like [[*so big*] [*a mess*]] Van Eynde proposes a construction whose mother is a noun phrase and whose first daughter is an adjective phrase marked "degree", which necessitates that it contain a degree

[8] Space limitations preclude the discussion of semantics in this paper. We have in mind an MRS-style semantics (Copestake et al. 2005), though nothing hinges on this choice.

[9] Van Eynde (2006, 2007) couches his proposal in terms of phrasal types, using the framework of Ginzburg and Sag (2000). For convenience, we refer to his phrasal type constraints as SBCG constructions. The reader should also be aware that Van Eynde posits multiple subtypes of his head-functor phrasal type, a complication that considerations of space require us to ignore here.

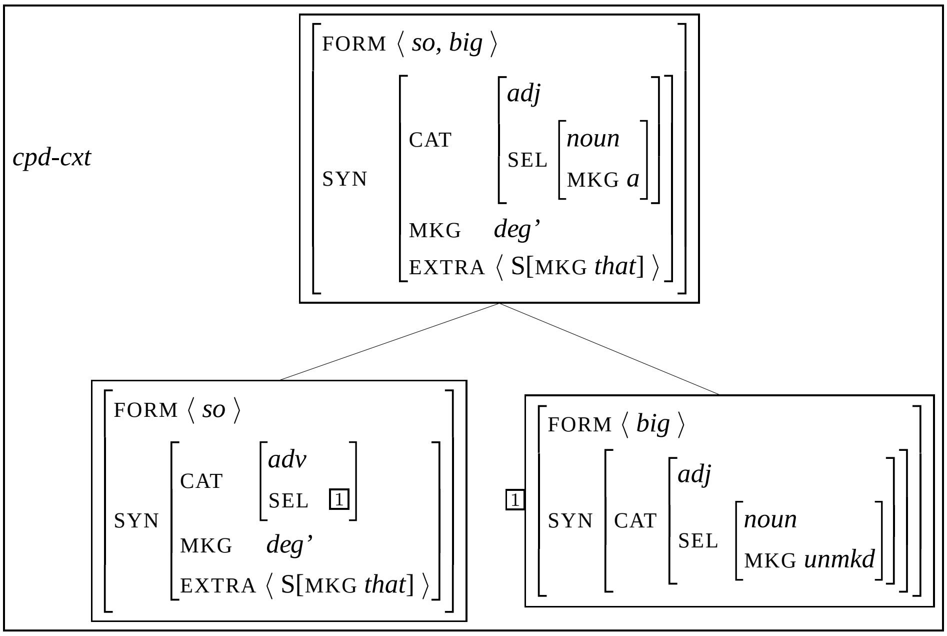


Figure 3: A Complex Pre-Determiner Construct

modifier from the list given in (6), excluding *such* (which is lexically specified to select a singular, indefinite NP). In Van Eynde's (2007) "Big Mess" construction, which is distinct from his Head-Functor construction, the adjectival daughter does **not** select the nominal head; rather the Big Mess construction specifies merely that the indices of the two daughters are identified.

We present here a related analysis that operates inside the adjective phrase, rather than at the NP level. This choice encodes a different intuition, namely that the special property of the CPD phenomenon is the apparent divergence of the selectional potential of an AP from that of its lexical head. On this view, *big* selects an undetermined nominal, but *so big* selects a singular, indefinite NP. The selectional process is the same as in normal adjectival modification: once the special AP *so big* is constructed to select an NP rather than a nominal (CNP) expression, the AP and the NP are combined by the familiar Head-Functor Construction. The need for a special construction arises only in building the AP.[10]

[10]Our account, unlike Van Eynde's, provides a uniform treatment of Big Mess APs (*so big*) and lexical expressions, e.g. *what*, *such*, and *many*, which may appear in pre-determiner position (*what/such/many a fool!*). That is, *what*, *such*, and *many* can bear exactly the same SEL value as the phrases licensed by the CPD Construction. Although these words select bare plurals (*Such fools!*), which Big Mess APs do not, all these facts could presumably be accommodated in a lexicon with multiple constraint inheritance. However, there is considerable lexical idiosyncrasy in this domain, as Van Eynde observes, and the additional generalization captured by our approach is arguably unimpressive in the light of it. We are not aware of further data that would distinguish our analysis from an appropriate extension of Van Eynde's on empirical grounds.

$$\textit{cpd-cxt} \Rightarrow \begin{bmatrix} \textsc{mtr} & \begin{bmatrix} \textsc{syn} & \begin{bmatrix} \textsc{cat} & \begin{bmatrix} \textit{adj} & \\ \textsc{sel} & \begin{bmatrix} \textsc{syn} & \begin{bmatrix} \textsc{cat} & \textit{noun} \\ \textsc{mkg} & \textit{a} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \textsc{mkg} & \textit{deg'} \\ \textsc{extra} & L_1 \end{bmatrix} \end{bmatrix} \\ \textsc{dtrs} & \left\langle \begin{bmatrix} \textsc{syn} & \begin{bmatrix} \textsc{cat} & [\textsc{sel}\ \ X] \\ \textsc{mkg} & \textit{deg'} \\ \textsc{extra} & L_1 \end{bmatrix} \end{bmatrix}, X{:} \begin{bmatrix} \textsc{syn} & \begin{bmatrix} \textsc{cat} & \begin{bmatrix} \textit{adj} & \\ \textsc{sel} & \begin{bmatrix} \textsc{syn} & \begin{bmatrix} \textsc{cat} & \textit{noun} \\ \textsc{mkg} & \textit{unmkd} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

Figure 4: Complex Pre-Determiner Construction

The CPD construct *so big* is shown in Figure 3. Starting with the first daughter ([FORM ⟨*so*⟩]), we note that its category is adverb and that it selects its right sister, indicated by the tag $\boxed{1}$. This constituent is specified as [MKG *deg'*], which is a lexical property of all and only the lexical items listed in (6), other than *such*. The EXTRA list contains a single item, which is a *that*-marked clause. The second daughter ([FORM ⟨*big*⟩]) is of category adjective and selects an unmarked nominal head. The mother of this construct ([FORM ⟨*so, big*⟩]) inherits its MKG and EXTRA values from the first daughter, as in a *hd-func-cxt*. Another similarity with a *hd-func-cxt* is the identification of the type of mother's CAT value (*adj*) with that of the second daughter. But here the parallelism with the Head-Functor Construction breaks down; we note that the second daughter is not the head daughter and the SEL values of the mother and second daughter differ. In particular, since the second daughter reflects the selection restriction of the lexical item *big*, *viz.* [MKG *unmkd*], it must be an undetermined nominal. By contrast, the mother's SEL value is a nominal sign specified as [MKG *a*], i.e. an NP determined by the article *a*.

The CPD Construction is sketched in Figure 4. A construct licensed by this construction is not a headed construct, as we have just seen. Although the category **type** of the mother (*adj*) matches that of the second daughter, the SEL values do not match: the mother selects an NP specified as [MKG *a*], but the second daughter selects a common noun, an NP specified as [MKG *unmkd*]. As in the construct it licenses that we have just considered (Figure 3), the MKG and EXTRA values of the first daughter and the mother are identified. The first daughter is specified as [MKG *deg'*], identifying it as one of the lexical licensers of the CPD phenomenon.

A noun phrase like *so big a mess* is licensed as follows. The AP *so big* is put together by the CPD construction, as we saw in Figures 3 and 4. The NP *a mess* is assembled by the familiar Head-Functor Construction [Figure 2 above]. The AP *so big* is licensed by the CPD construction in Figure 4, which guarantees that it has the properties sketched in (15):

(15)
$$\boxed{\begin{bmatrix} \text{FORM} & \langle \textit{so, big} \rangle \\ \text{SYN} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{SEL} & \begin{bmatrix} \textit{noun} & \\ \text{MKG} & \textit{a} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}}$$

Therefore, the Head-Functor Construction is appropriate to combine *so big* and *a mess* via the former's selection of the latter, with the resulting construct shown in (16):

(16) *hd-func-cxt*
$$\boxed{\begin{bmatrix} \text{FORM} & \langle \textit{so, big, a, mess} \rangle \\ \ldots & \end{bmatrix}}$$
$$\boxed{\begin{bmatrix} \text{FORM} & \langle \textit{so, big} \rangle \\ \ldots & \end{bmatrix}} \quad \boxed{\begin{bmatrix} \text{FORM} & \langle \textit{a, mess} \rangle \\ \ldots & \end{bmatrix}}$$

And the mother of the construct in (16) has the properties shown in (17):[11]

(17)
$$\boxed{\begin{bmatrix} \text{FORM} & \langle \textit{so, big, a, mess} \rangle \\ \text{SYN} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \textit{noun} & \\ \text{SEL} & \textit{none} \end{bmatrix} \\ \text{SUBJ} & \langle\,\rangle \\ \text{COMPS} & \langle\,\rangle \\ \text{MKG} & \textit{deg'} \\ \text{EXTRA} & \langle \text{S}[\textit{that}] \rangle \end{bmatrix} \end{bmatrix}}$$

Having put together constructs like *so big a mess*, we now need to account for an extraposed *that*-clause, extraposed in the sense that while it is introduced by *so*, it is only realized following *mess*. Moreover, it need not immediately follow *mess*, as shown in (18):

(18) [[[**so** big a mess] resulted from the meeting of the committee on the seventeenth of August] **that it took hours to clean it up**].

[11]Following Müller's (2009) account of predicative NPs, which creates them via a unary ("pumping") construction from nonpredicative NPs, we have a straightforward account of predicative uses, e.g. examples like *She is so big a fan that she bought season tickets*, *Kim is too honest a guy to do that*, etc.

The mechanism for realizing extraposed elements and the positions in which they can be realized will occupy much of our attention for the remainder of this paper.

We noted that in both the Head-Functor Construction and the CPD construction the mother inherits its EXTRA value from the first daughter. The lexical entry for *so* is the source of the eventually extraposed *that*-clause, as shown in (19):

(19) $$\begin{bmatrix} \text{FORM} & \langle \textit{so} \rangle \\ \text{SYN} & \begin{bmatrix} \text{CAT} & [\text{SEL}\ [\text{SYN}\ [\text{EXTRA}\ \ L_1]]] \\ \text{EXTRA} & L_1 \oplus \langle \text{S}[\textit{that}] \rangle \end{bmatrix} \end{bmatrix}$$

The lexical entry for *so* stipulates that its EXTRA list includes a *that*-clause appended to ($\oplus$) the EXTRA list of the element that *so* selects. That is, *so* says in effect "My extra list consists of the EXTRA list of the element I select followed by a *that*-clause." Various constructions, including the CPD constructions, specify the EXTRA value of the mother in terms of the EXTRA values of the daughters, in the case of the constructions we have seen so far – and also the Subject-Predicate Construction, presented below – the mother's EXTRA value is identified with the EXTRA value of the first daughter. Often the EXTRA list of the selected element will be empty, as in the case of *big*. The result is that when *so* and *big* are combined, the EXTRA value of the mother (*so big*) is just the singleton list containing S[*that*]. The EXTRA values of both *a* and *mess* are the empty list, so the EXTRA value of *a mess* is the empty list. Hence, the EXTRA value of *so big a mess* will consist of the single item S[*that*], which originated on the EXTRA list of the lexical entry for *so*, got "passed up" to *so big* by the CPD Construction and then again to *so big a mess* by the Head-Functor Construction.

How do extraposed elements get off the EXTRA list and realized in the sentence? The extraposition analysis we are proposing follows previous GPSG/HPSG treatments of nonlocal dependencies.[12] At the site of introduction, lexical or constructional constraints ensure that the unrealized element corresponds to an element of the SLASH (or GAP) – or, in this case, EXTRA – list of the minimal phrase containing the gap. General principles then require that this feature specification be inherited by the mothers of successively larger constructs – these phrases form the middle of the filler-gap dependency. Certain constructions then license the presence of these "slashed" phrases, typically introducing a new phrase (the filler) that is identified with the SLASH value of its sister phrase (at the top of the filler-gap dependency). The construction realizing extraposed elements, the Head-Extraposition Construction,[13] is given in (20):

[12]See Gazdar 1981, Pollard and Sag 1994, Bouma et al. 2001, and Levine and Hukari 2006.

[13]See Pollard and Sag 1994, Keller 1995, Van Eynde 1996, Bouma 1996, Kim and Sag 2005 and Crysmann to appear.

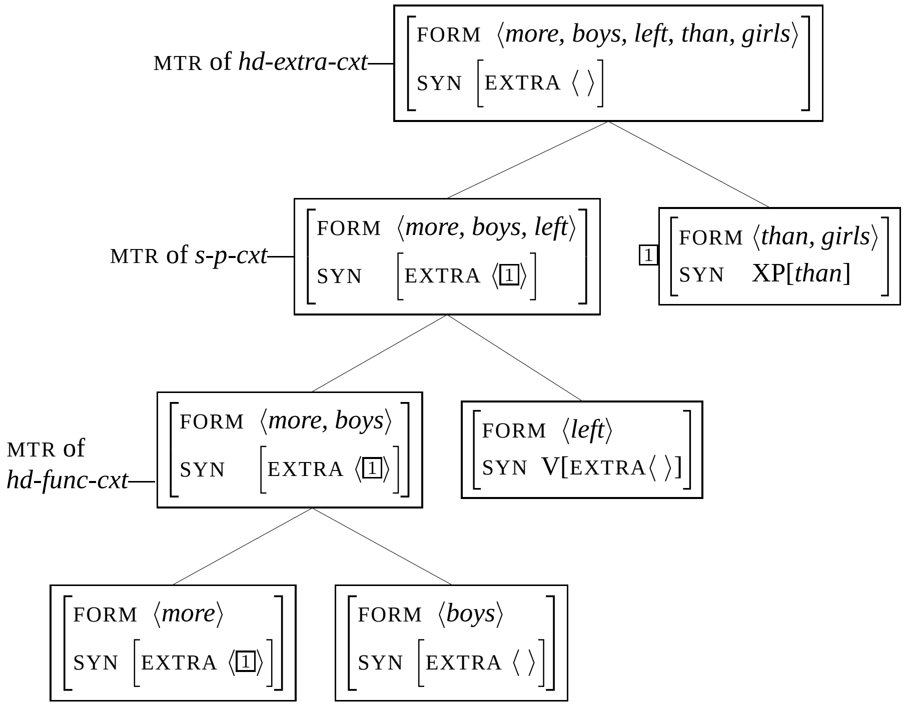


Figure 5: A Head-Extraposition Derivation

(20) **Head-Extraposition Construction:**

$$\textit{hd-extra-cxt} \Rightarrow \begin{bmatrix} \text{MTR} & \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{COMPS} & L_1 \\ \text{EXTRA} & L_2 \end{bmatrix} \end{bmatrix} \\ \text{DTRS} & \left\langle H : \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{COMPS} & L_1 \\ \text{EXTRA} & \langle X \rangle \oplus L_2 \end{bmatrix} \end{bmatrix}, X \right\rangle \end{bmatrix}$$

The Head-Extraposition Construction in (20) realizes the initial element of the EXTRA list of the head (first) daughter as the second daughter. The EXTRA list of the mother is the EXTRA list of the head daughter minus the element realized as the second daughter. This means that the order of elements on a non-singleton EXTRA list corresponds to the linear order of those elements in a binary-branching head-extraposition derivation.

The combination of the three lexical and constructional processes is exemplified in Figure 5. Starting at the lower left, we see that *more*, in combining with *boys*, records on its EXTRA list the requirement for a *than*-phrase, represented by the tag $\boxed{1}$, adding this element to the empty EXTRA list of its selected sister *boys*. The Head-Functor Construction identifies the EXTRA list of its functor daughter *more* with that of the mother of the construct it licenses (*more boys*). When *more boys* and *left* combine in accordance with the Subject-Predicate Construction, the EXTRA list of the first (non-head) daughter *more boys* also becomes the EXTRA list of the mother *more boys left* (because the EXTRA list of the head daughter must be empty) – see below. The construct combining *more boys left* and *than girls* is licensed by the Head-Extraposition Construction [(20)], which realizes the sole member of the head daughter's EXTRA list (the XP[*than*]) as the second daughter *than girls* of the highest construct in Figure 5. The EXTRA list of this construct's mother is the empty list.

Extraposed elements obey certain ordering restrictions, as we saw in examples (7)–(9) above. In order to specify where extraposed elements can be realized we need to consider further constructions. First, we note that some extraposed complements, either arising within the VP or extraposed from the subject, can be permuted with arguments of predicates and also with other extraposed elements, such as relative clauses:

(21) a. Kim was **more** *willing* **than Pat** *to wash the dishes*.

b. Kim was **more** *willing* *to wash the dishes* **than Pat**.

c. I sent out **more** *books* yesterday *that I really liked* **than ever before**.

d. I sent out **more** *books* yesterday **than ever before** *that I really liked*.

(22) a. **More** *books* arrived *that I actually liked* **than I expected**.

b. **More** *books* arrived **than I expected** *that I actually LIKED*.

As noted earlier, not all extraposed elements have this property. In particular, as summarized in (10) above, complements of *too*, *so* and *enough* do not permute with arguments or other extraposed dependents, as shown again by the examples in (23):

(23) a. The boys are **so** *proud* now *of their achievements* **that they've become unbearable**.

b. *The boys are **so** *proud* now **that they've become unbearable** *of their achievements*.

c. Nichelle is **so** much *taller* now *than Beavis* **that people think she's in middle school**.

d. *Nichelle is **so** much *taller* now **that people think she's in middle school** *than Beavis*.

Two things need to be explained about the data of (21)–(23): (1) the fact just mentioned, that comparative complements permute while *so*, *too* and *enough* complements don't, and (2) the prior fact that some extraposed complements permute with elements that are patently extraposed. We account for the latter fact, the crossed dependencies in (21a) and (22b) – by postulating two unary lexical constructions. The first "moves" arguments from the COMPS list to the EXTRA list; the second allows nouns to be constructed that have a relative clause on their EXTRA list.[14] An initial sketch of these constructions is given in (24) and (25):[15]

(24) **Complement Extraposition Construction:**

$$
\textit{comp-extra-cxt} \Rightarrow \left[\begin{array}{ll}
\text{MTR} & \left[\begin{array}{ll} \multicolumn{2}{l}{\textit{word}} \\ \text{SYN} & \left[\begin{array}{ll} \text{SUBJ} & \langle \text{NP} \rangle \\ \text{COMPS} & L_1 \\ \text{EXTRA} & L_2 \oplus \langle\, X \,\rangle \end{array}\right] \end{array}\right] \\
\text{DTRS} & \left\langle \left[\begin{array}{ll} \multicolumn{2}{l}{\textit{word}} \\ \text{SYN} & \left[\begin{array}{ll} \text{SUBJ} & \langle \text{NP} \rangle \\ \text{COMPS} & L_1 \bigcirc \langle\, X \,\rangle \\ \text{EXTRA} & L_2 \end{array}\right] \end{array}\right] \right\rangle
\end{array}\right]
$$

(25) **Nominal Modifier Extraposition Construction:**

$$
\textit{nm-extra-cxt} \Rightarrow \left[\begin{array}{ll}
\text{MTR} & \left[\begin{array}{ll} \multicolumn{2}{l}{\textit{word}} \\ \text{FORM} & \langle\, Y \,\rangle \\ \text{SYN} & \left[\begin{array}{ll} \text{CAT} & \textit{noun} \\ \text{COMPS} & L_1 \\ \text{EXTRA} & L_2 \oplus \langle\, X[\text{SEL}\ \ Z\,] \,\rangle \end{array}\right] \end{array}\right] \\
\text{DTRS} & \left\langle Z : \left[\begin{array}{ll} \multicolumn{2}{l}{\textit{word}} \\ \text{FORM} & \langle\, Y \,\rangle \\ \text{SYN} & \left[\begin{array}{ll} \text{CAT} & \textit{noun} \\ \text{COMPS} & L_1 \\ \text{EXTRA} & L_2 \end{array}\right] \end{array}\right] \right\rangle
\end{array}\right]
$$

The Complement Extraposition Construction "pumps" a daughter (intuitively, one that is a "predicator") with an item *X* (anywhere) on its COMPS list to a mother

[14] A relative clause otherwise functions as a nominal modifier selecting the nominal it modifies via SEL; see Sag submitted.

[15] In (24), $\bigcirc$ denotes the "shuffle" relation, as opposed to the append relation ($\oplus$) used in (25) and in (19) and (20) above. See Reape 1994.

$$
\textit{comp-extra-cxt} \quad
\begin{array}{c}
\left[\begin{array}{ll}
\text{FORM} & \langle \textit{proud} \rangle \\
\text{SYN} & \left[\begin{array}{ll} \text{COMPS} & \langle\ \rangle \\ \text{EXTRA} & \langle \text{PP}[\textit{of}] \rangle \end{array}\right]
\end{array}\right] \\
| \\
\left[\begin{array}{ll}
\text{FORM} & \langle \textit{proud} \rangle \\
\text{SYN} & \left[\begin{array}{ll} \text{COMPS} & \langle \text{PP}[\textit{of}] \rangle \\ \text{EXTRA} & \langle\ \rangle \end{array}\right]
\end{array}\right]
\end{array}
$$

Figure 6: A Complement Extraposition Construct

predicator where X appears as the last element of the EXTRA list and is absent from the COMPS list. As the final element on the EXTRA list, X is the last element on the list to be realized by the Head Extraposition Construction [(20) above] and hence appears in the sentence after any other elements realized from this list.[16] Multiple extraposition dependencies typically arise when one of these extraposition dependencies interacts with one of the extraposition dependencies induced lexically (by *so*, *more*, etc.). A *comp-extra-cxt* (a post-lexical construct in the terminology of Sag 2010) is illustrated in Figure 6, where the daughter's COMPS list contains a PP[*of*], and its EXTRA list is empty. The mother's COMPS list is empty – the PP[*of*] appears on the EXTRA list.

Let us now return to the fact that, unlike other extraposed modifer complements (such as *than*- or *as*-phrases), *so*, *to* and *enough* complements never participate in crossed dependencies. We account for this via the lexical entries shown in (26):

(26) a.
$$
\left[\begin{array}{ll}
\text{FORM} & \langle \textit{so} \rangle \\
\text{SYN} & \left[\begin{array}{ll} \text{CAT} & \left[\text{SEL}\ \ [\text{SYN}\ [\text{EXTRA}\ L_1]]\right] \\ \text{EXTRA} & L_1 \oplus \langle \text{S}[\textit{that}] \rangle \end{array}\right]
\end{array}\right]
$$

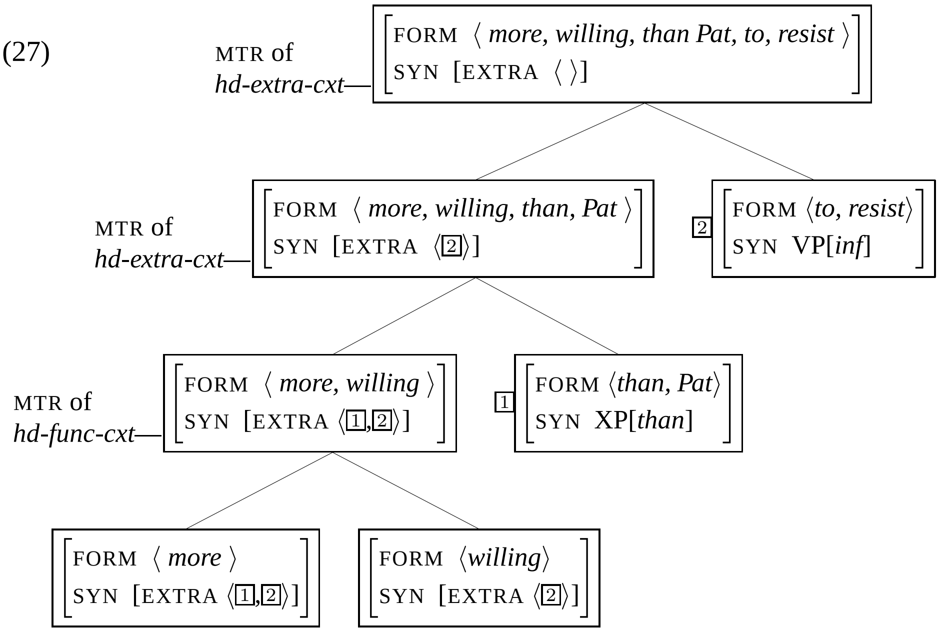
b.
$$
\left[\begin{array}{ll}
\text{FORM} & \langle \textit{more} \rangle \\
\text{SYN} & \left[\begin{array}{ll} \text{CAT} & \left[\text{SEL}\ \ [\text{SYN}\ [\text{EXTRA}\ L_1]]\right] \\ \text{EXTRA} & L_1 \bigcirc \langle \text{XP}[\textit{than}] \rangle \end{array}\right]
\end{array}\right]
$$

We have already seen that *so* adds its S[*that*] complement at the right end of the EXTRA list, ensuring that it will be realized highest (hence latest, rightmost) in the structure of any element realized from the same list. Note that the entry for *more* is

[16]Because the Head-Extraposition Construction is binary, only one extraposed element is introduced at each level of structure. Hence, multiple extrapositions involve a nested, left-branching derivational structure.

the same, with the important difference that the XP[*than*] complement is added not at the end, but at an arbitrary position within the selected element's EXTRA list (as specified by the use of ◯, rather than ⊕). This arrangement allows complements of comparative modifiers to be realized either earlier (hence lower, to the left) or later (hence higher, to the right) of other elements realized from their list – except for *so/too/enough* complements, as illustrated in (27):

(27)



We have seen that *so/too/enough* complements must follow comparative complements if they reside on the same EXTRA list. However, if the comparative element is within the subject NP and the *so/too/enough* licensor is within the VP of a subject-predicate clause, then it is in fact required that the *so/too/enough* dependent linearly precede the *than*-phrase (extraposition is bounded by the VP):

(28) a. **More** girls were <u>*so*</u> happy <u>*that they cheered*</u> **than boys**.

b. ***More** girls were <u>*so*</u> happy **than boys** <u>*that they cheered*</u>.

We account for this by formulating the Subject-Predicate Construction as shown in Figure 7. A construct licensed by the Subject-Predicate Construction is a headed construct with a mother and two daughters. The mother's syntax specifies it to be non-inverted and finite, with empty SUBJ and COMPS lists and, crucially in the present context, an EXTRA list that is identified with that of the first (subject) daughter. The subject daughter satisfies the subject valence requirement (*Y*) of the head VP daughter, The EXTRA list of the latter must be empty, ensuring that any

$$
s\text{-}p\text{-}cxt \Rightarrow \left[\begin{array}{ll}
\multicolumn{2}{l}{\textit{hd-cxt}} \\
\text{MTR} & \left[\begin{array}{ll} \text{SYN} & \left[\begin{array}{ll} \text{CAT} & \left[\begin{array}{ll} \text{VFORM} & \textit{fin} \\ \text{INV} & - \end{array}\right] \\ \text{SUBJ} & \langle\ \rangle \\ \text{EXTRA} & L \end{array}\right] \end{array}\right] \\
\text{DTRS} & \left\langle Y\text{: } [\text{EXTRA}\ \ L\ ]\ ,\ H\text{: } \left[\begin{array}{ll} \text{SYN} & \left[\begin{array}{ll} \text{SUBJ} & \langle\, Y \,\rangle \\ \text{COMPS} & \langle\ \rangle \\ \text{EXTRA} & \langle\ \rangle \end{array}\right] \end{array}\right] \right\rangle
\end{array}\right]
$$

Figure 7: The Subject-Predicate Construction

extraposed elements that arise within the VP of a subject-predicate construct are realized within that VP.

Finally, we note that it is not just subject-predicate clauses that inherit the extraposition potential of their first daughter. This is also true of filler-gap constructions:

(29) a. [[[How many **more** talents] did she have] **than the other candidate**]?

b. [[[Which **candidate**] did he support] **who had signed the legislation**]?

c. [[[How many **soups**] he had sampled] **that he didn't like**]!

d. [[[**So** eager] was he to see the comet] **that he stayed up all night**].

4 Conclusion

In this paper, we have seen that the complex pre-determination ("Big Mess") phenomenon and the discontinuous dependency phenomenon are independent – either may occur in a sentence without the other. Nevertheless we find them frequently intertwined because there are seven lexical entries (*so*, *too*, *more*, *less*, *as*, *such*, and *how*) that contain features which play key roles in both constructions. The CPD phenomenon requires a special construction (in our analysis or the alternative suggested in Van Eynde 2007); the DD phenomenon follows from the properties of certain lexical licensors and the grammatical mechanisms that govern exptaposition in general. The details of the distribution of DD complements derive from the interaction of (1) a general construction for realizing elements of the EXTRA list, (2) specifications on phrasal constructions determining the contents of the mother's EXTRA list as a function of the EXTRA lists of the daughters, and (3) various lexical specifications for relevant lexical licensors. We believe that the general approach we have adopted here has provided a vehicle for the precise representation of these phenomena in a way that has allowed us to abstract the significant generalizations they present, to elucidated their interactions with other aspects of grammar, and to thereby explicate the interaction of the idiosyncratic, the general, and the gray area in between.

References

Allegranza, Valerio. 1998. Determiners as functors: NP structure in Italian. In Sergio Balari and Luca Dini (eds.), *Romance in Head-driven Phrase Structure Grammar*, volume 75 of *CSLI Lecture Notes*, pages 55–108, Stanford: CSLI Publications.

Allegranza, Valerio. 2007. *The Signs of Determination*. Frankfurt am Main: Peter Lang.

Arnold, Doug and Sadler, Louise. 1992. Non-Modifying Adjectives in HPSG. In *Working Papers in Language Processing*, volume 35, Dept. of Languages and Linguistics, U. Essex.

Berman, Arlene. 1974. *Adjectives and adjective complement constructions in English*. Ph. D.thesis, Harvard University.

Boas, Hans and Sag, Ivan A. (eds.). 2010. *Sign-Based Construction Grammar*. Stanford: CSLI Publications.

Bouma, Gosse. 1996. Extraposition as a Nonlocal Dependency. In *Proceedings of Formal Grammar 96*, pages 1–14, Prague.

Bouma, Gosse, Malouf, Rob and Sag, Ivan A. 2001. Satisfying Constraints on Extraction and Adjunction. *Natural Language and Linguistic Theory* 1(19), 1–65.

Chae, Hee-Rahk. 1992. *Lexically Triggered Unbounded Discontinuities in English: An Indexed Phrase Structure Grammar Approach*. Ph. D.thesis, The Ohio State University.

Copestake, Ann, Flickinger, Dan, Pollard, Carl J. and Sag, Ivan A. 2005. Minimal Recursion Semantics: an Introduction. *Research on Language and Computation* 3(4), 281–332.

Crysmann, Berthold. to appear. On the Locality of Complement Clause and Relative Clause Extraposition. In Gert Webelhuth, Manfred Sailer and Heike Walker (eds.), *Rightward Movement in a Comparative Perspective*, Amsterdam: John Benjamins.

Fillmore, Charles J. 1999. Inversion and constructional inheritance. In Gert Webelhuth, Jean-Pierre Koenig and Andreas Kathol (eds.), *Lexical and Constructional Aspects of Linguistic Explanation*, Studies in Constraint-Based Lexicalism, Chapter 21, pages 113–128, Stanford: CSLI Publications.

Fillmore, Charles J., Kay, Paul and O'Connor, Mary C. 1988. Regularity and Idiomaticity in Grammatical Constructions: The Case of *Let Alone*. *Language* 64, 501–538.

Flickinger, Daniel P. and Nerbonne, John. 1992. Inheritance and Complementation: A Case Study of *Easy* Adjectives and Related Nouns. *Computational Linguis-*

tics 19(3), 269–309, walter Dalemans and Gerald Gazdar, eds., *Inheritance and Natural Language Processing*, special issue.

Gazdar, Gerald. 1981. Unbounded Dependencies and Coordinate Structure. *Linguistic Inquiry* 12, 155–84.

Huddleston, Rodney D. and Pullum, Geoffrey K. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Kay, Paul. 2002. English Subjectless Tagged Sentences. *Language* 78.3, 453–481.

Kay, Paul and Fillmore, Charles. 1999. Grammatical Constructions and Linguistic Generalizations: The *What's X Doing Y?* Construction. *Language* 75.1, 1–33.

Keller, Frank. 1995. Towards an Account of Extraposition in HPSG. In *Proceedings of the 7th Conference of the EACL*, pages 301–306, Dublin, student session.

Kennedy, Christopher and Merchant, Jason. 2000. Attributive comparative deletion. *Natural Language and Linguistic Theory* 18, 89–146.

Kim, Jong-Bok and Sag, Ivan A. 2005. Variations in English Object Extraposition. In *The Proceedings of the 41st Annual Meeting of the Chicago Linguistic Society*, pages 192–212, Chicago: CLS.

Kiss, Tibor. 2005. Semantic Constraints on Relative Clause Extraposition. *Natural Language and Linguistic Theory* 23(2), 281–334.

Levine, Robert D. and Hukari, Thomas. 2006. *The Unity of Unbounded Dependency Constructions*. CSLI Lecture Notes, No. 166, Stanford University: CSLI Publications.

Michaelis, Laura A. and Lambrecht, Knud. 1996. Toward a Construction-Based Model of Language Function: The Case of Nominal Extraposition. *Language* 72, 215–247.

Müller, Stefan. 2004. Continuous or Discontinuous Constituents? A Comparison between Syntactic Analyses for Constituent Order and Their Processing Systems. *Research on Language and Computation* 2(2), 209–257, Special Issue on Linguistic Theory and Grammar Implementation.

Müller, Stefan. 2009. On Predication. In Stefan Müller (ed.), *The Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar*, Stanford: CSLI Publications.

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

Reape, Mike. 1994. Domain Union and Word Order Variation in German. In John Nerbonne, Klaus Netter and Carl J. Pollard (eds.), *German in Head-Driven Phrase Structure Grammar*, CSLI Lecture Notes, No. 46, pages 151–197, Stanford University: CSLI Publications.

Sag, Ivan A. 2010. Sign-Based Construction Grammar: An informal synopsis. In Hans Boas and Ivan A. Sag (eds.), *Sign-Based Construction Grammar*, CSLI.

Sag, Ivan A. in press. Feature Geometry and Predictions of Locality. In Greville Corbett and Anna Kibort (eds.), *Features: Perspectives on a Key Notion in Linguistics*, Oxford: Clarendon Press.

Sag, Ivan A. submitted. English Filler-Gap Constructions.

Van Eynde, Frank. 1996. An HPSG Treatment of *it*-Extraposition without Lexical Rules. In G. Durieux, W. Daelemans and S. Gillis (eds.), *CLIN VI. Papers from the Sixth CLIN Meeting*, pages 231–248, University of Antwerp, Center for Dutch Language and Speech.

Van Eynde, Frank. 1998. The Immediate Dominance Schemata of HPSG. A deconstruction and a reconstruction. In P.A. Coppen, H. van Halteren and L. Teunissen (eds.), *Computational Linguistics in the Netherlands 1997. Selected Papers from the Eighth CLIN Meeting*, pages 119–133, Amsterdam/Atlanta: Rodopi.

Van Eynde, Frank. 2006. NP-Internal Agreement and the Structure of the Noun Phrase. *Journal of Linguistics* 42(1), 139–186.

Van Eynde, Frank. 2007. The Big Mess Construction. In Stefan Müller (ed.), *The Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar, Stanford University*, pages 415–433, Stanford: CSLI Publications.

Wittenburg, Kent. 1987. Extraposition from NP as Anaphora. In Geoffrey Huck and Almerindo Ojeda (eds.), *Discontinuous Constituency*, Syntax and Semantics, No. 20, pages 428–445, New York: Academic Press.

An Alternative to the HPSG Raising Principle on the Description-Level

David Lahm

Universität Tübingen

Proceedings of the HPSG09 Conference

Göttingen University, Göttingen, Germany

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

I reconsider the HPSG RAISING PRINCIPLE which is introduced in Pollard and Sag (1994) to constrain the way in which lexical entries describe the SUBCAT lists of the words they license. On the basis of whether a complement is assigned a semantic role in a lexical entry or not, this entry may not or must describe this complement as structure-shared with the unrealised subject of some other (non-subject) complement. The formal status of this principle is still unclear, as it is formulated as a 'meta principle' that does not talk about linguistic objects directly but rather about the lexical entries that license them. I show that, although its meaning cannot be expressed faithfully by the usual kind of constraints employed in HPSG, the RAISING PRINCIPLE can nevertheless be replaced by two such constraints which make largely the same predictions. Most importantly, these constraints interact with the output values of description-level lexical rules in the style of Meurers (2001) in a way that makes predictions available that Pollard and Sag (1994) intended the RAISING PRINCIPLE to make but that it cannot possibly make if description-level lexical rules are employed.

1 Introduction

In chapter 3 of Pollard and Sag (1994) the RAISING PRINCIPLE (RP) is introduced. This principle's initial motivation was to enforce under certain circumstances what I shall call the *raising configuration* on the SUBCAT lists of words:[1] If a subject is raised, it is identical to the *synsem* object belonging to the subject or object it is raised to. This is illustrated by the SUBCAT lists in (1).

(1) a. $\left\langle \boxed{1}, \text{VP}\left[\text{SUBCAT } \langle \boxed{1} \rangle\right] \right\rangle$

b. $\left\langle \text{NP}, \boxed{1}, \text{VP}\left[\text{SUBCAT } \langle \boxed{1} \rangle\right] \right\rangle$

The list described in (1a) corresponds to a subject-to-object raising verb like *seem*. The raising configuration holds on the list since the subject (i.e. the first element of the list) is token-identical to the subject of the unsaturated VP. Similarly, the raising configuration holds on the list described in (1b). This list might be that of a subject-to-object raising verb like *believe* or *expect*.

1.1 Description-Level and Meta-Level

In Pollard and Sag (1994), linguistic regularities were expressed on two different levels, which I shall call (following Meurers (2001)) the description-level and the

[†]I thank Frank Richter, Janina Radó, Manfred Sailer, Gerald Penn and Ivan Sag for valuable discussions and advice; I thank the three anonymous reviewers of an earlier draft of this paper for helpful comments.

[1]In the following I shall sometimes be talking about raising verbs instead of raising words, usually because certain properties of verbs are an issue. This should not distract from the fact that the RP as well as the principles introduced in this paper talk about words in general.

meta-level. The descriptions that are formulated on the description-level talk about linguistic objects directly and constrain them to certain shapes. On the meta-level, lexical rules are formulated that derive new lexical entries (which are descriptions of words) from existing ones, thus allowing for what has been called 'horizontal generalisations' in the lexicon. Apart from the lexical rules, constraints on the shape of not linguistic objects but instead the descriptions constraining them can be imagined, and the RAISING PRINCIPLE was intended as just this kind of meta-level constraint.

While a precise formal explication of the meaning of principles formulated on the description-level has been provided in the form of (R)SRL,[2] no such formal rendering has so far been given for meta-level rules and principles. Hence no precise account of the RPs meaning exists so far, and the details of its effects may thus always be subject to some amount of speculation. Furthermore, since the principle was intended to constrain not only basic lexical entries but likewise those generated as the outputs of lexical rules, it is incompatible with a description-level approach to lexical rules in the style of Meurers (2001) for principled reasons, since in this approach *word* objects are related to other *word* objects, while the originally envisioned meta-level formalisation of lexical rules should have related lexical entries to lexical entries. If use is made of a description-level formalisation of lexical rules, the RAISING PRINCIPLE will no more be able to constrain the output values of the rules, since these will be in a domain different from the one the principle talks about (linguistic objects vs. lexical entries).

Since neither a satisfactory formulation of meta-level lexical rules nor of the RP have yet been given, I consider the meta-level approach a dead end for the time being. It follows that, for the purpose of full formalisation, description-level lexical rules are called for. It is then an obvious question whether some replacement of the RP in terms of descriptions can be given as well.[3] To my knowledge, no such alternatives have been attempted to give, probably because the RP as it stands talks about the way words are described by lexical entries, which clearly is something that cannot be done in the same manner by a description that is formulated on the same level as the lexical entries themselves. From this it follows that the effects of the alternative that will be offered here cannot be quite the same as those of the RP (under any interpretation it might possibly be given).

1.2 The issue

In this paper, I offer a replacement for the RP formulated on the description-level. The replacement is intended to have at least the effects that the RP was positively intended to have. In section 2, I briefly review the idea of the original RP and take a look at the features of this principle that make it so problematic to express on the description level. In section 3, taking my departure from the theory of

[2] cf. Richter (2004).

[3] Descriptions are considered expressions of (R)SRL or the AVM notation for (R)SRL, as defined in Richter (2004). I consider nothing else a true description.

English presented in the first eight chapters of Pollard and Sag (1994), excluding the CONTROL THEORY, I offer two new principles as an alternative to the RP, namely the ARGUMENT CONSISTENCY PRINCIPLE (ACP) and the COHERENCE OF VALENCE PRINCIPLE. In section 4 I argue that, if description-level lexical rules are employed, these two principles achieve what the RP was intended to achieve with respect to predicting the shapes of the output values of lexical rules. I show that the argument about the Subject Extraction Lexical Rule (SELR) output for the null relativiser that Pollard and Sag (1994, p. 219, fn. 4.) make, as well as the one about the impossibility of Null Complement Anaphora with raising verbs and raising adjectives, in contrast to its possibility with equi verbs and adjectives (cf Pollard and Sag (1994, p. 140-142), can be reconstructed using the new principles. Apart from this one, hardly any attempts at precise arguments employing the RP are known to me, which is probably due to its dubious formal status.[4] By giving a formalised description-level alternative to the RP while preserving the effects intended by Pollard and Sag (1994), I hope to make a more precise discussion of the meaning, effects and necessity of constraints ensuring the raising configuration possible.

2 The RAISING PRINCIPLE

In this section I shall take a brief look at the RP of Pollard and Sag (1994).

(2) (META-LEVEL) RAISING PRINCIPLE *Appendix Version*

> Let E be a lexical entry in which the (description of the) SUBCAT list L contains (a description corresponding to) a member X (of L) that is not explicitly described in E as an expletive. Then in (the description of) the CONTENT value, X is (described as) assigned no semantic role if and only if L (is described as if it) contains a nonsubject whose own SUBCAT value is <X>.

The crucial aspect of this formulation is the reference that is made to the way things are described by lexical entries, in particular the phrases *not explicitly described in E as an expletive* and *(described as) assigned no semantic role*. I consider each of these in turn.

- *not explicitly described in E as an expletive*

[4] A further is about the output values of a lexical rule accounting for prepositional passives in Tseng (2007). It is interesting to note that Tseng (2007), unlike Pollard and Sag (1994), does not assume that the RP can be used to predict the shape of output values of lexical rules. For Tseng, the rules themselves have to be formulated so as to comply with the principle.

If a member of a SUBCAT list of a *word* object *w* is not explicitly described as expletive in the lexical entry that licenses *w* (and, of course, if it is neither explicitly described as referential), then, in *w*, this member of the list might actually be expletive and just as well might not. All that is required is that the relevant complements of *word* objects licensed by the entry need not necessarily be expletive, while it is still possible for them to actually be expletive. This fact is made use of in the RP to account for the fact that (for example) raising verbs behave identically in cases where the verbal complement's subject is expletive and in cases where it is not.

- *(described as) assigned no semantic role*

  I distinguish between *role assignment* and *role filling*. *Role assignment* takes place in the domain of lexical entries (which the RP constrains). A *synsem* object is assigned a semantic role in a *word* object's CONTENT value if a token identity of (part of) the semantics of the *synsem* object (usually a member of the *word*'s SUBCAT list) and a semantic object filling a role in the *word*'s CONTENT value is specified in the entry that licenses the *word*.[5] *Role filling* denotes the case where (part of) the semantics of a *synsem* object actually is the value of some role attribute (like RUNNER, KNOWER, SOA-ARG etc.) evaluated on a *psoa* and similar cases (which are explained in detail below). A *synsem* object may fill a role without being assigned it. This is so because the token identity required for role filling might neither be ruled out by the word's lexical entry nor enforced by it. In this case role filling in a word licensed by the entry is possible, but not necessary. In (3), e.g., the subject fills the roles DISMEMBERER and DISMEMBEREE, but is only assigned the role DISMEMBEREE.

  (3) James dismembered himself

  The notion of role filling can and will be explicated precisely on the description level, while role assignment can not.

Any lexical entry that does not explicitly describe some member of the SUBCAT lists of the *word*s it describes as expletive has to assign this SUBCAT list member a semantic role just in case it descibes it also as the subject of some unsaturated complement on the same list. So in any given *word* object *w* any complement that is not assigned a role in the CONTENT value of *w* by the lexical entry that licenses *w* must be the subject of some unsaturated complement of *w*, provided the complement is not explicitly described as expletive in the entry that licenses *w*.

[5]Note that my explication of the meaning of role assignment does not seem to fully agree with the meaning intended in Pollard and Sag (1994). Given my explication, the phrase *(described as) assigned no semantic role* should be replaced by *assigned a semantic role by (or in) E*. But this shift in meaning does no harm here and it is convenient to condense the relevant distinction in two distinct expressions.

Since this structure sharing is not enforced by directly constraining the linguistic objects at issue, but indirectly by requirig the lexical entries to ensure it, it will hold no matter whether the actually shared *synsem* in some *word* the entry licenses is referential or expletive.

Assume (4) to be a partial lexical entry for the verb *seem*.

(4)
$$\begin{bmatrix} \textit{word} & \\ \text{SYNSEM}|\text{LOCAL} & \begin{bmatrix} \text{CATEGORY}|\text{SUBCAT} & \left\langle \boxed{1}, \text{VP}\left[\text{INF}, \text{SUBCAT}\langle\boxed{1}\rangle\right]:\boxed{2} \right\rangle \\ \text{CONTENT} & \begin{bmatrix} \textit{seem} \\ \text{SOA-ARG} \ \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The principle then accounts for the following data.

(5) a. It is great to like her
b. He seems to like her
c. It seems to be great to like her
d. There seems to be good reason to like her
e. *There seems to be great to like her

(5a) is licensed since the lexical entry for the relevant form of *be* explicitly requires its subject to be expletive *it*. So the RP does not impose any further constraints on the entry. In constrast to that, the entry of *seem* does not impose any constraints on its subject. As a consequence of that, there are concrete instances of *seem* (i.e. *word* objects) that do have a referential subject (as in (5b)) and others that have an expletive subject (as in (5c), (5d)). Since the subject of *seem* is not assigned any semantic role in the lexical entry, the lexical entry is constrained by the RP to describe it as identical to that of the word's VP complement (i.e. the co-tagging in (4) which is done using the tag $\boxed{1}$ is actually enforced by the principle). This identity must then obtain in all *word* objects the entry licenses, no matter whether the subject is expletive as in (5c), (5d) or referential as in (5b). So the desired effect is achieved and the subject of *seem* also is the subject of the embedded VP. Obviously, (5e) will be ruled ungrammatical since *be great to like her* requires expletive *it* as its subject.

Although most of these considerations seem to be quite clear intuitively, neither the meaning of the RP nor that of meta-level lexical rules has been pinned down formally so far. So in the remainder of this paper I shall investigate whether and at what cost the effects the principle was intended to have in Pollard and Sag (1994) can be achieved by constraints formulated on the description-level.

3 The Description-Level Approach

3.1 Auxilliary Relations

Before introducing and discussing the description-level replacement for the RP that I shall offer, I need to introduce two auxillary relations: `raised-on` and `role-filler`.

3.1.1 Raised-on

The relation `raised-on` is a relation between *synsem* objects and (intuitively, SUBCAT) lists. It obtains between a *synsem* and a *list* object if the *synsem* object is a member of the list and also is the single element of some non-initial (i.e. non-subject) *synsem* element of that list. If for some list there exists a *synsem* object that is raised on the list, I also say that the raising configuration obtains on the list. Formally, the relation is defined as follows;

(6) $\forall s \forall l$
$(\texttt{raised-on}(s, l) \ \leftrightarrow$

$\exists \boxed{1} \exists \boxed{2}$
$({}^{l}\left[\text{REST } \boxed{1}\right] \wedge$
$\texttt{member}(s,\ l) \wedge$
$\texttt{member}(\boxed{2}\left[\text{LOCAL}|\text{CATEGORY}|\text{SUBCAT } \langle s \rangle\right], \boxed{1})))$

3.1.2 Role-filler

To define the `role-filler` relation, I first need to introduce a new attribute called LEXICAL-SEMANTIC-CONTRIBUTION (LSC). LSC is appropriate to the sort *content* and takes as its value either an object of sort *psoa* or of sort *none*. If present, the psoa which is its value will always be the ‘characteristic’ psoa of the word, the one that the word adds to the overall meaning of its phrase. For example, in the case of an adjective like *blue*, it will be the *blue*-psoa introduced by the word.[6] The attribute is needed for the treatment of semantically vacuous words, as will become clear later.

The relation `role-filler`, a relation between *synsem* objects, covers three distinct ways in which the CONTENT value of one such object can fill a role in that of the other. These are:

(i) The CONTENT values of both objects are identical.

[6]I assume that the LSC values of head-daughter and mother are always identical. The formulation of a principle to enforce this is of course trivial. The attribute might be independently motivated, cf. e.g. De Kuthy (2000, p. 101).

This covers the base form complement of infinitival *to*, which simply identifies its own CONTENT value with that of its base form complement. It also covers case marking prepositions, which do the same. If identity of CONTENT values were not regarded as an instance of role filling, the base form complement of *to* and the prepositional objects of case marking prepositions would be required to be raised.[7]

(ii) The CONTENT value of the first argument is a *psoa* that is also an element of the $\mathrm{CONTENT}|\mathrm{RESTR}$ list of the second.

This covers the clausal complement of the null relativiser, the CONTENT value of which appears on the relativisers RESTR list and must not be required to be raised.

(iii) The CONTENT or $\mathrm{CONTENT}|\mathrm{INDEX}$ value of the first argument is the value of some attribute $\rho \in \mathcal{R}$ evaluated on the $\mathrm{LOCAL}|\mathrm{CONTENT}|\mathrm{LSC}|$-$\mathrm{NUCLEUS}$ value of the second argument, where $\mathcal{R}$ is the set of those and only those attributes that denote functions on *psoas* corresponding to semantic roles (the obvious case).

The formal definition of the relation is given in (7)

(7) $$\forall x \forall y$$
$$(\texttt{role-filler}(x, y) \leftrightarrow$$

$$\exists \boxed{1}$$
$$(({}^{x}\left[\text{LOCAL}|\text{CONTENT} \begin{bmatrix} \textit{nom-obj} \\ \text{INDEX } \boxed{1} \end{bmatrix}\right] \vee (\neg \boxed{1}[\textit{nom-obj}] \wedge {}^{x}[\text{LOCAL}|\text{CONTENT } \boxed{1}])) \wedge$$
$$({}^{y}[\text{LOCAL}|\text{CONTENT } \boxed{1}] \vee$$
$$\exists \boxed{2}({}^{y}[\text{LOCAL}|\text{CONTENT}|\text{RESTR } \boxed{2}] \wedge \texttt{member}(\boxed{1}[\textit{psoa}], \boxed{2})) \vee$$
$$\bigvee_{\rho \in \mathcal{R}} {}^{y}[\text{LOCAL}|\text{CONTENT}|\text{LSC}|\text{NUCLEUS}|\rho \ \boxed{1}])))$$

3.2 The Principles

In this section I shall investigate the possibilities of gaining the effects that the RP was supposed to have while using only constraints that can be formulated on the description level. I shall start by considering why two naive ‘translations’ of the meta-principle to the description-level will not work. While one of these appears to be truly inadequate, the other will be kept, but a second principle will be added to it.

The problem that has to be faced when trying to construe an alternative to the RP on the description-level is that description-level constraints can never make ref-

[7]A similar notion on the metalevel is implicitly employed in Pollard and Sag (1994). There, too, (at least) the three cases covered here will have to be considered as instances of *role assignment*.

erence to the way things are described in lexical entries. Hence no formulation corresponding to the phrase *not explicitly described in E as an expletive* as employed in the RP will be available if an alternative to the principle is to be formulated on the description level. This leads to the question how the fact is to be accounted for that raising verbs show identical behaviour in cases where the subject to be raised is expletive and in cases where it is not; that is, why expletive raising is generally possible.

The RP achieves this by constraining lexical entries to specify certain token identities, but descriptions can only constrain linguistic objects to have certain shapes, where in any (by definition totally well-typed and sort-resolved) *word* object, a given complement of the *word* is either expletive or is not. So, if one simply tried to reformulate the principle as closely as possible as a constraint on *word*-objects, the condition *not explicitly described in E as an expletive* could only be (i) dropped or (ii) taken over as *not an expletive*. (The condition *not assigned a semantic role by E* will be translated as *not filling a semantic role*. This treatment will lead to problems discussed in section 3.4.)

(i) Is not an option. It would result in a principle somewhat like (8).

(8) For every *word* object, every *synsem* object on the *word*'s SUBCAT list is raised on that list if it does not fill a role in the *word*'s CONTENT value.[8]

This is of course no viable solution. Since no expletive ever fills a role it would amount to requiring every expletive to be raised. So expletives on the SUBCAT lists of words that are lexically specified to select for expletives would also be required to be raised, which clearly is not a welcome effect. Lexically specified selection is where selection for expletives originates, and it is not hard to see that introducing such a principle would in effect require every sentence containing an expletive to be infinite.

(ii) Would result in something like the principle (9).

(9) The ARGUMENT CONSISTENCY PRINCIPLE (*informal version*)

For every *word* object, every *synsem* object on the *word*'s SUBCAT list that is neither expletive nor does fill a role in the *word*'s CONTENT value is raised on that list.

[8]In the following principles, the *if and only if* formulation present in the RP will be replaced by a unidirectional implication. In the RP, the condition assures that complements can only be described as raised if they are not assigned a role. That means that the lexical entry of a subject or object control verb may not specify this verb's subject or object, respectively, to be identical to that of the VP complement because it assigns it a role. Yet in a given concrete instance of the word, this identity may hold. cf. Pollard and Sag (1994, p. 140, fn. 40), Przepiórkoswki and Rosen (2004). If the biconditional were retained in the description-level analysis, this would exclude this identity in all instances of the word, which is a completely undesired effect.

This principle ensures that the semantics of every non-expletive argument that happens to be on a *word*'s SUBCAT list must either be used in the semantics of this *word* directly or raised from some other complement of the *word*. Following a suggestion by Frank Richter and Janina Radó, I shall henceforth call it the ARGUMENT CONSISTENCY PRINCIPLE (ACP).

This principle alone will not have the desired effect. This is so because, in some sense, nothing is said by it about expletive complements. While referential complements that do not fill roles are required to be raised, expletives are not. This leads to the licensing of such sentences as in (10).

(10) a. *There seems to like her
b. *There seems to be obvious that you like her

(10a) could mean, e.g., *you seem to like her*. Since expletive subjects would not be required to be raised, the VP's subject could correspond to *you* while *seem* could have some expletive (*there*, in this case) as its subject. (10b) is similar; in this case both subjects are expletive, but the expletives are different although they should be the same.

In essence, the question seems to be how to distinguish complements that are lexically specified to be expletive from those that are not. On the meta-level, this can be done by simply looking at the lexical entries. On the description level, it can not.

The analysis offered here attacks the problem from a different angle. Assume that the RP is dropped and the ACP is incorporated into the theory and consider the sentences in (10). Both of them have in common that the subject of the embedded VP is in some way 'lost': it does not get realised by direct combination with the VP according to the head-subject-schema, but neither is it realised anywhere else. Realisation of the subject 'somewhere else' is just what would have happened had the subject been raised. In the sentences in (10), it would have been realised as the subject of the raising verb. In cases with subject-to-object raising as the object.

If the ACP were accompanied by some principle that prevented subjects of embedded VPs to be lost in the manner displayed in (10), and if this principle further required every expletive subject of such a VP to be raised, the problem with the sentences in (10) would be solved. These considerations lead to the formulation of the following principle, which I call the COHERENCE OF VALENCE PRINCIPLE (CVP).[9]

[9]It was pointed out to me by Manfred Sailer (personal communication) that the CVP as stated in this paper may be in conflict with analyses of so-called 'arbitrary' PRO as in (1).

(1) It is easy to like her

In (1), the unrealised subject of the infinitival VP is not realised in any of the ways the CVP requires such subjects to be realised in. However, the question whether this is a fatal problem cannot be answered without any detailed analysis of conditions under which 'arbitrary' PRO can occur. This paper is not the right place to give such an analysis.

(11) THE COHERENCE OF VALENCE PRINCIPLE (*preliminary informal version*)

For every *word* object whose SUBCAT list contains a non-subject complement whose SUBCAT list has the form <X>, one of the following must be the case:

- The LOCAL value of X is a member of the *word*'s SLASH set
- The LOCAL|CONTENT|INDEX value of X is identical to that of another member of the *word's* SUBCAT list
- X is raised on the SUBCAT list of the *word*, which is always the case for expletives

The first of the three bullets is needed to license the output values of the SUBJECT EXTRACTION LEXICAL RULE. The second covers the case of equi verbs, where the unrealised subject of the embedded VP only needs to share its index with that of its controller. The third allows for the realisation of the VP's subject via raising and enforces raising if the subject is expletive.

It is easily seen that by the CVP, the examples in (10) are ruled ungrammatical. In the case of (10a), the embedded VP's subject will be required to share at least its index with some element of the matrix verb's SUBCAT list. Only the matrix verb's subject is available for sharing the index with. But if its index is shared with that of the VP's subject it will neither be expletive nor fill a role and hence be required by (9) to be raised, which is the desired result. In the case of (10b), raising is enforced directly by the CVP, since the subject of the embedded VP is expletive.

3.3 Subject-to-Object Raising

Subject-to-object raising verbs give rise to a further problem. Consider (12).

(12) a. *John believes there to like her
b. John believes himself to like her

The analysis presented so far licenses (12a) with the semantics of (12b). To see this, assume that the subject of the VP *to like Kim* in (12a) has the same index as the *synsem* object corresponding to *John*. Then the sentence complies to the CVP: the unrealised subject of the VP has been realised as required by the CVP by virtue of having the same index as the subject of the matrix verb. Nothing requires the expletive to be raised, since nothing requires its index to be referential, which the CVP does for the subject of (10a), thus ruling this sentence ungrammatical. In (10a), only the subject is available for sharing its index with that of the unrealised subject in accordance with the CVP. In (12a) the index should be required to be shared with the accusative object of the matrix verb. But by the current formulation

For similar reasons, *tough* constructions might be problematic, too.

of the CVP, it may share it with the matrix verb's subject just as well. If this is the case, nothing prevents expletive *there* from occuring as the accusative object, which, as an expletive, is not required to be raised by (9).

It appears possible to solve this problem by requiring unrealised VP subjects to be realised as the most oblique complement of the word selecting for the VP that is still less oblique than the VP itself.[10] In the case of *seem*, this would be the subject, in the case of *believe*, the accusative object. This leads to the revised version of the CVP shown in (13).

(13) THE COHERENCE OF VALENCE PRINCIPLE (*final informal version*)

For every *word* object whose SUBCAT list contains a non-subject complement Y whose SUBCAT list has the form $<$X$>$, one of the following must be the case:

- The LOCAL value of X is a member of the *word*'s SLASH set
- The most oblique member of the described *word* object's SUBCAT list that is still less oblique than Y (call it Z) is such that
  - The LOCAL|CONTENT|INDEX value of X is identical to that of Z or
  - X is identical to Z, which is always the case if X is an expletive

3.4 Subject-to-Object Raising and Reflexives

It has already been remarked above that not only the phrase *not explicitly ... as an expletive* but also the notion of *role assignment* cannot be faithfully expressed on the description level. This leads to undesired consequences in the case of subject-to-object raising verbs with reflexive accusative objects: (14a) is possible with the SUBCAT list (14b), where $\boxed{1} \neq \boxed{2}$. That is, the reflexive is not required to be raised as it should.

(14) a. John believes himself to like her

b. $\left\langle \text{NP}_i, \boxed{1}_i\text{:}\textit{reflexive}, \text{VP}\left[\text{SUBCAT } \langle\boxed{2}\rangle\right]\right\rangle$

This is the case because the indices of subject and accusative object are (and must be, by PRINCIPLE I of the BINDING THEORY) identical. So, since the subject fills the role BELIEVER, the object also fills this role. It is not possible to distinguish on the description-level between role filling by lexical assignment and externally enforced role filling, in this case role filling enforced by the BINDING THEORY. To force reflexives like the one in (14a) to be raised, it is inevitable to be able to tell which member of a SUBCAT list it is that fills a certain role; that means it must be impossible for distinct SUBCAT list members to fill the same role.

[10] If the *to*-phrase in sentences like *he seems to me to like her* is to be analysed as a complement of *seem*, this would probably be inconsistent with the present suggestion. I however doubt that an analysis of the *to*-phrase as an complement is mandatory.

To achieve this, I introduce a sort of 'intermediary index' which must be unique for every member of a SUBCAT list. The attribute INDEX, evaluated on an intermediary index, will yield a 'standard' index. It must then be these standard indices that the BINDING THEORY and most other principles talk about, while intermediary indices are only relevant for more exotically flavoured purposes.

To make this proposal precise, I introduce the sort (species) *intind* (for *intermediary index*) as a direct subsort of *object* in the way shown in (15).

(15) *intind*
    INDEX *index*

The only attribute appropriate to *intind* is INDEX, and to evaluate it on an object of this sort yields an object of sort *index*. Evaluating INDEX[11] on an object of sort *nom-obj* will now yield an object of sort *intind*. The attributes that represent roles on *psoa*s must now likewise yield objects of sort *intind* when evaluated on such an object.

The benefit of *intind*s is that, while objects of sort *index* can be required to be indentical for different members of the same SUBCAT list by virtue of the (appropriately reformulated) BINDING THEORY, their *intind*s do not have to be indentical. To say more, they may not even be identical, which is enforced by the following constraint.

(16) UNIQUENESS CONDITION on intemediary indices

$$\left[\textit{category}\right] \rightarrow$$

$$\left[\text{SUBCAT } \boxed{1}\right] \wedge$$
$$\forall x \forall y$$
$$(\texttt{member}(x,\boxed{1}) \ \wedge \ \texttt{member}(y,\boxed{1}) \rightarrow$$
$$(x\text{LOCAL}|\text{CONTENT}|\text{INDEX}= y\text{LOCAL}|\text{CONTENT}|\text{INDEX}\rightarrow x = y))$$

The constraint states that for any two members of a SUBCAT list, having the same LOCAL|CONTENT|INDEX value means to be identical.[12]

Since role attributes evaluated on *psoa*s yield *intind*, it is now the case that no two distinct *synsem* objects on the SUBCAT list of any *word* can ever fill the same role in the CONTENT value of this *word*. So in the case of (14), although the 'standard' indices of subject and accusative object, which are embedded in the *intind*s under the attribute INDEX, are identical, the *intinds* are not. Hence the subject fills the role BELIEVER, but the reflexive object does not and will be required to be raised.

[11]Which could now be renamed INTIND, but it need not.

[12]Note that still the same *synsem* object could occur twice on the same list. If necessary, this possibility might be ruled out by a further constraint, requiring that any *synsem* object occur at most once on any SUBCAT list.

3.5 Formal Statement of the Principles

The ARGUMENT CONSISTENCY PRINCIPLE as well as the COHERENCE OF VALENCE PRINCIPLE can be stated formally in (the AVM notation for) RSRL. To do this is the sole purpose of this section.

(17) The ARGUMENT CONSISTENCY PRINCIPLE (*formal version*)

$$
\begin{array}{l}
\left[\textit{word}\right] \rightarrow \\[1ex]
\left(\left[\textsc{synsem}\ \boxed{1}\left[\textsc{local}|\textsc{category}|\textsc{subcat}\ \boxed{2}\right]\right] \wedge\right. \\[1ex]
\forall x \\
(\mathtt{member}(x, \boxed{2}) \rightarrow \\
((\neg\mathtt{role\text{-}filler}(x, \boxed{1})\ \wedge \neg\, {}^{x}\!\left[\textsc{local}|\textsc{content}|\textsc{index}|\textsc{index}\ \textit{expletive}\right] \rightarrow \\
\mathtt{raised\text{-}on}(x, \boxed{2})))))
\end{array}
$$

The principle says that for every object on a *word* object's SUBCAT list, if it does not fill a role and is not expletive, then it is raised on the SUBCAT list.

(18) The COHERENCE OF VALENCE PRINCIPLE (formal version)

$$
\begin{array}{l}
\left[\textit{word}\right] \rightarrow \\[1ex]
\left(\left[\textsc{synsem}\ \begin{bmatrix}\textsc{local}|\textsc{category}|\textsc{subcat} & \boxed{1}\left[\textsc{rest}\ \boxed{2}\right] \\ \textsc{nonloc}|\textsc{slash} & \boxed{3}\end{bmatrix}\right] \wedge\right. \\[1ex]
\forall x \forall y \\
(\mathtt{member}({}^{y}\!\left[\textsc{local}|\textsc{category}|\textsc{subcat}\ \left\langle {}^{x}\!\left[\textsc{local}\ \boxed{4}\right]\right\rangle\right], \boxed{2})\ \rightarrow \\[1ex]
\mathtt{member}(\boxed{4}, \boxed{3}) \vee \\
\exists \boxed{6} \\
(\mathtt{to\text{-}the\text{-}left}(\boxed{6}, y, \boxed{1})\ \wedge \\
\forall z(\mathtt{to\text{-}the\text{-}left}(z, y, \boxed{1})\ \wedge z \neq \boxed{6} \rightarrow \mathtt{to\text{-}the\text{-}right}(\boxed{6}, z, \boxed{1})) \wedge \\
(x = \boxed{6} \vee \\
({}^{x}\!\left[\textsc{local}|\textsc{content}|\textsc{index}|\textsc{index}\ \boxed{5}\right] \wedge \boxed{6}\!\left[\textsc{local}|\textsc{content}|\textsc{index}|\textsc{index}\ \boxed{5}\right])) \wedge \\
({}^{x}\!\left[\textsc{local}|\textsc{content}|\textsc{index}|\textsc{index}\ \textit{expletive}\right] \rightarrow x = \boxed{6}))))
\end{array}
$$

The relation `to-the-right`, defined in Richter (2004, ch. 4.2), holds of its argument triple if the last argument is a list and the first argument occurs to the right of the second argument on that list. The relation `to-the-left` is defined with respect to the relation `to-the-right`.

(19) The relation `to-the-left`

$$
\begin{array}{l}
\forall x \forall y \forall z \\
(\mathtt{to\text{-}the\text{-}left}(x, y, z) \leftrightarrow \neg\mathtt{to\text{-}the\text{-}right}(x, y, z) \wedge x \neq y)
\end{array}
$$

An object x is to the left of an object y on a list z iff x is not to the right of and not identical with y.

The principle requires that for any *word* object whose SUBCAT list contains an unsaturated complement, this complement's subject's LOCAL value must either be in the described *word*'s SLASH set, or there is some element on the described *word*'s SUBCAT list that is the most oblique complement of the word that is still less oblique than the unsaturated complement.[13] With this object, then, the unsaturated complement's subject is either identical or index-shared.

4 Interaction with Description-Level Lexical Rules

The RP was not only used in Pollard and Sag (1994) to constrain the content of basic lexical entries, but likewise to constrain the content of output values of lexical rules. This made sense intuitively, since lexical rules were conceived of as rules generating new lexical entries from existing ones. With the RP understood as constraining all lexical entries, the output values of lexical rules had to be constrained as well.

Since no satisfactory formalisation of the meta-level rules employed in Pollard and Sag (1994) has yet been established, the description-level formalisation given by Meurers (2001) seems the only possibility so far to have a formal account of lexical rules at all. But if description-level lexical rules are employed, lexical rules are understood as descriptions of *lex-rule* objects. An application of a rule can be understood as an object licensed by the rule. Input to and output of an application of a rule are represented as components of the corresponding *lex-rule* object, that is, as *word* objects that can be reached from the *lex-rule* object via the attributes IN and OUT, respectively. It is a crucial aspect of this formalisation that no new lexical entries for the rules' outputs are generated at all. Hence the RP, even if it were precisely formalised, could never constrain the outputs of these rules.

From this it follows that, even if the RP as a meta-level statement were to be retained and a satisfactory formalisation of it were given, then either lexical rule outputs could no more be constrained by it or finding a satisfactory formalisation of meta-level lexical rules would again be an issue. In this section, I assume that description-level lexical rules are used and investigate how much of the predictions made about the output values of lexical rules using the RP in Pollard and Sag (1994) can be reconstructed employing these rules and the principles introduced above. Only two such arguments about output values of lexical rules are know to me, both stemming from Pollard and Sag (1994). I consider both of them in turn.

4.1 The Null Relativiser

The most explicitly formulated and also most interesting argument known to me in which the RP is employed in to reason about output values of a lexical rule

[13] Recall that the order on SUBCAT lists reflects the obliqueness hierarchy.

concerns the SUBJECT EXTRACTION LEXICAL RULE (SELR) output for the null relativiser. In this section, I briefly summarise the central aspects of the relativiser and the lexical rule and review this argument. I then show that the structure sharing that the RP should have guaranteed in the objects licensed by the rule's output for the relativiser is also enforced by the new principles in the output values of a corresponding description-level rule. Thus the effects the RP was intended to have are preserved.

A null relativiser is employed in the treatment of relative clauses suggested in Pollard and Sag (1994). Given this phonetically empty word, the relative clause (20a) will have the structure shown in (20), with the relativiser represented as **e**.

(20) a. <man> to whom you gave your pocket
b. <man> [to whom [**e** [you gave your pocket]]].

The lexical entry for **e** requires the LOCAL value of the relative phrase to be a member of the SLASH set of the clause which the relativiser takes as its complement. The relativiser binds this SLASH set member. Furthermore it requires the relativiser to select for a subject[14] with just the LOCAL value of the relative phrase. Together with the identification of the index which is described by the entry as the single member of the subject's INHER|REL set with the relativiser's CONTENT|INDEX value and the INDEX of the modified N', (20a) is licensed with the syntax indicated in (20b) and appropriate semantics.

It is for the analysis of sentences like (21a) that the SUBJECT EXTRACTION LEXICAL RULE (SELR) comes into play. The desired analysis of this sentence is shown in (21).

(21) a. <man> who gave his pocket to you
b. <man> [who [**e** [gave his pocket to you]]]

Without the SELR, an analysis as indicated in (21b) would be blocked by the TRACE PRINCIPLE for English, which forbids the extraction of subjects to rule sentences like (22) ungrammatical.

(22) *Who did Bob tell you that will visit us?

Yet the extraction of subjects seems possible in cases like (24), where the subject of the embedded clause *who will visit us* is fronted.

(23) Who did Bob tell you will visit us?

Pollard and Sag (1994) account for these sentences by introducing the SELR as shown in (24).

[14] Recall that in the theory presupposed here, *subject of x* simply means *initial member of the SUBCAT list of x*.

(24) $$\left[\text{SS}\mid\text{LOC}\mid\text{CAT}\mid\text{SUBCAT } \left\langle \text{Y}, \ldots, \text{S}[\textit{unmarked}], \ldots \right\rangle\right] \mapsto$$

$$\left[\text{SS}\begin{bmatrix}\text{LOC}\mid\text{CAT}\mid\text{SUBCAT} & \left\langle \text{Y}, \ldots, \text{VP}\left[\text{SS}\begin{bmatrix}\text{LOC}\mid\text{CAT}\mid\text{SC} & \left\langle[\text{LOC }\boxed{1}]\right\rangle \\ \text{NONLOC}\mid\text{INH}\mid\text{SL} & \{\}\end{bmatrix}\right], \ldots\right\rangle \\ \text{NONLOC}\mid\text{INHER}\mid\text{SLASH} & \{\boxed{1}\}\end{bmatrix}\right]$$

This lexical rule guarantees that for each word that subcategorises for an unmarked non-subject sentential complement there is another word, subcategorising for a VP instead of the sentence and having the LOCAL value of this VP's subject as the single member of its INHER|SLASH set.[15] (23) is then explained as resulting from an application of the SELR to *tell*. It is thus analysed as a HEAD-FILLER-CONSTRUCTION with *did Bob tell you will visit us* as the head and *Who* as the filler.

The SELR is employed to derive the null relativiser for (21) from that employed in (20), and here it is that the RP comes into play. Consider (25a).[16] This description of a *synsem* object fits the input schema of the SELR and so can serve as a part of its input (assuming for the moment meta-level lexical rules), yielding an output about as shown in (25b), if the effects of the RP are ignored.

(25) a. $$\begin{bmatrix}\text{LC} & \begin{bmatrix}\text{CAT} & \begin{bmatrix}\text{HD} & \begin{bmatrix}\textit{rltvzr} & \\ \text{MOD} & \text{N'}\left[\text{NL}\mid\text{TO-BIND}\mid\text{REL }\{\boxed{1}\}\right]:\begin{bmatrix}\text{INDEX} & \boxed{1}\\ \text{RESTR} & \boxed{2}\end{bmatrix}\end{bmatrix} \\ \text{SC} & \left\langle \begin{bmatrix}\text{LC} & \boxed{3}\\ \text{NL}\mid\text{INHER}\mid\text{REL} & \boxed{1}\end{bmatrix}, \text{S}\left[\text{INHER}\mid\text{SL }\{\boxed{3}\}, \textit{fin, unmarked}\right]:\boxed{4}\right\rangle \\ \text{CNT} & \begin{bmatrix}\textit{npro} & \\ \text{INDEX} & \boxed{1}\\ \text{RESTR} & \{\boxed{4}\}\cup\boxed{2}\end{bmatrix}\end{bmatrix}\end{bmatrix} \\ \text{NL}\mid\text{T-B}\mid\text{SL } \{\boxed{3}\} & \end{bmatrix}$$

b. $$\begin{bmatrix}\text{LC} & \begin{bmatrix}\text{CAT} & \begin{bmatrix}\text{HD} & \begin{bmatrix}\textit{rltvzr} & \\ \text{MOD} & \text{N'}\left[\text{NL}\mid\text{TO-BIND}\mid\text{REL }\{\boxed{1}\}\right]:\begin{bmatrix}\text{INDEX} & \boxed{1}\\ \text{RESTR} & \boxed{2}\end{bmatrix}\end{bmatrix} \\ \text{SC} & \left\langle \boxed{5}[\text{INHER}\mid\text{REL }\boxed{1}], \text{VP}\begin{bmatrix}\text{SC} & \left\langle\boxed{6}[\text{LC }\boxed{7}]\right\rangle, \textit{fin}\\ \text{INHER}\mid\text{SLASH} & \{\}\end{bmatrix}:\boxed{4}\right\rangle \\ \text{CNT} & \begin{bmatrix}\textit{npro} & \\ \text{INDEX} & \boxed{1}\\ \text{RESTR} & \{\boxed{4}\}\cup\boxed{2}\end{bmatrix}\end{bmatrix}\end{bmatrix} \\ \text{NL} & \begin{bmatrix}\text{INHER}\mid\text{SL} & \{\boxed{7}\}\\ \text{T-B}\mid\text{SL} & \{\boxed{3}\}\end{bmatrix}\end{bmatrix}$$

[15] Note that this is the first case considered in the CVP.

[16] I need to abbreviate some of the attribute names to make the AVM fit the page. I also sometimes abbreviate paths by leaving attributes out where this should not lead to any confusion.

Pollard and Sag (1994) now argue that, to comply with the RP, (25b) in addition needs to specify that $\boxed{5} = \boxed{6}$, which should also imply $\boxed{3} = \boxed{7}$.

The argument runs as follows.

> To see this, consider the following facts: (1) the SELR input specifies structure sharing between the $\mathrm{TO-BIND|SLASH}$ value and the first complement's LOCAL value; (2) SELR specifies stucture sharing between the output's $\mathrm{INHER|SLASH}$ value and the LOCAL value of its VP complement's SUBCAT value; and (3) the Raising Principle requires that there be structure sharing between the LOCAL value of the VP complement's SUBCAT value and the LOCAL value of the first complement. (In identifying *synsem* objects, the Raising Principle of course identifies the LOCAL values within the *synsem* objects.)
>
> Pollard and Sag (1994, p. 219, fn. 4)

The central point is of course (3): because the subject is not assigned a role by the entry in (25a), it must be raised.[17] It can only be raised from the VP complement, so must be raised from there.[18] So, according to this argument, the subject of the null relativiser in (25b) and the subject of its VP complement must be the same. This does not and may not result from the SELR itself; the form of *tell* occuring in (23) must be licensed by the output of SELR. There, *Bob* is the subject of tell, but the extracted subject corresponds to *who*, So they clearly cannot be identical. Differently, in (21b), where the subject has been extracted from the verbal complement of the relativiser, identity of the relativiser's subject to that of the VP is crucial, as shown in (26).

(26) a. $<$man$>$ *who$_i$ **e** $\_\_{}_j$ gave his pocket to you
     b. $<$man$>$ who$_i$ **e** $\_\_{}_i$ gave his pocket to you

Clearly, (26a) does not make any sense (given, of course, $i \neq j$), so (26b) must be enforced. This is done by the RP.

The ACP will have similar effects. If description-level lexical rules are employed, the descriptions in (25) must be seen as descriptions of the INPUT value

[17] Note that it is necessary to consider it as an instance of *role assignment* when the CONTENT value is described in a lexical entry as a member of the described *word* object's $\mathrm{CONTENT|RESTR}$ list. Otherwise the second complement of each relativiser would not be assigned a role and hence required to be described as raised by the RP. This fact is reflected in the second clause of the definition of *role filling*, given above.

[18] As was already noted in Pollard and Sag (1994, p. 216, fn. 3), this argument applies to the rule's input as well, but since the verbal complement there is a saturated sentence, the subject cannot be raised. So (25a) actually cannot license anything at all. The analysis presented here suffers from the same problem. I assume that some modification of the ACP along the lines indicated in Pollard and Sag (1994) is in order. This would amount to requiring any non-expletive non-role-filler to be either raised or 'SLASH raised', i.e. have its LOCAL value on the SLASH set of another member of the SUBCAT list it is on, rather than be on the complement's SUBCAT list itself.

and OUTPUT value of some *lex-rule* object. All co-taggings indicate structure sharing as in any other description.[19] Both described objects must of course obey all constraints on *word* objects. Since the relativiser does not contribute any semantics of its own, it has no characteristic *psoa*. So its CONTENT|LSC value is *none*. Hence none of its complements can be a role filler according to the last clause of the definition of *role filling*. The VP complement is a role filler according to the second clause of the definition, but the subject is not. So the last possibility for the relativiser's subject to fill a role would be identity of its CONTENT value to that of the relativiser, according to clause (i) of the definition. This is not possible either.[20]

4.2 Null Complement Anaphora

Null complement anaphora is a complement drop phenomenon; the *to*-infinitival complement of a raising verb can be dropped if it can be inferred from the context. But this is only possible with equi verbs, not with raising verbs.[21] The contrast is exemplified in (27).

(27) a. John tries to understand this article and Janet also tries
b. *John seems to understand this article and Janet also seems

Pollard and Sag (1994, p. 141) assume that NCA is "a lexical process that removes an infinitival complement [...] from the SUBCAT list of verbs or adjectives". They offer no detailed analysis and also remain silent about the exact nature of this lexical process, but it seems safe to assume that an analysis based on lexical rules was intended. Such a rule might look like the one given in (28).

(28) $$\begin{bmatrix} \text{SS}|\text{LOC}|\text{CAT} \begin{bmatrix} \text{HEAD} & \textit{verb} \vee \textit{adj} \\ \text{SUBCAT} & \boxed{1}\left\langle \text{NP}_{\boxed{2}_{ref}} | \textit{list} \right\rangle \oplus \left\langle \text{VP}\left[\text{SUBCAT}\ \left\langle \text{NP}_{\boxed{2}} \right\rangle, \textit{inf}\right] \right\rangle \end{bmatrix} \end{bmatrix} \mapsto \\ \left[\text{SS}|\text{LOC}|\text{CAT}|\text{SUBCAT}\ \boxed{1}\right]$$

The rule takes as input a lexical entry describing verbs and adjectives subcategorising for an infinitival complement, this being the last element on their SUBCAT list, and returns a lexical entry where this element has been removed from the list. Considering the application of that rule to the lexical entry of *seem*, the relevant parts of which are given in (29a), the result would be the entry in (29b)

[19]In the meta-level approach they were usually ambiguous between identity of object and identity of description.

[20]If cases exist in which this identity can hold, these cases are pathological. Their existence could hence not be used to criticise the present theory but only to criticise itself.

[21]It should be noted that the arguments made here about NCA cannot be generalised so as to explain also the phenomenon of VP deletion as in *She can do it and Jack also can* or *She wants to go and Jack also wants to*, as noted by Pollard and Sag (1994, p. 142, fn. 43). I shall not further concern myself with this problem.

(29) a. $$\left[\textsc{ss}|\textsc{loc}\ \begin{bmatrix}\textsc{cat} & \begin{bmatrix}\textsc{head} & \textit{verb}\\ \textsc{subcat} & \left\langle \boxed{1}\mathrm{NP}[\textit{nom}],\ \mathrm{VP}[\textsc{subcat}\ \boxed{1},\ \textit{inf}]{:}\boxed{2}\right\rangle\end{bmatrix}\\ \textsc{cont} & \begin{bmatrix}\textit{seem}\\ \textsc{psoa-arg}\ \boxed{2}\end{bmatrix}\end{bmatrix}\right]$$

b. $$\left[\textsc{ss}|\textsc{loc}\ \begin{bmatrix}\textsc{cat} & \begin{bmatrix}\textsc{head} & \textit{verb}\\ \textsc{subcat} & \left\langle \boxed{1}\mathrm{NP}[\textit{nom}]\right\rangle\end{bmatrix}\\ \textsc{cont} & \begin{bmatrix}\textit{seem}\\ \textsc{psoa-arg}\ \boxed{2}\end{bmatrix}\end{bmatrix}\right]$$

But (29b) clearly violates the raising principle: The subject is not assigned a semantic role in the content, so it should be raised, which it is not and cannot be. Thus the RP rules out (29b) as a possible lexical entry and therewith (27b). Furthermore, as desired, equi verbs can undergo NCA, the relevant lexical entries for the example *try* being shown in (30).

(30) a. $$\left[\textsc{ss}|\textsc{loc}\ \begin{bmatrix}\textsc{cat} & \begin{bmatrix}\textsc{head} & \textit{verb}\\ \textsc{subcat} & \left\langle \mathrm{NP}[\textit{nom}]_{\boxed{1}},\ \mathrm{VP}\left[\textsc{subcat}\ \left\langle \mathrm{NP}_{\boxed{1}}\right\rangle,\ \textit{inf}\right]{:}\boxed{2}\right\rangle\end{bmatrix}\\ \textsc{cont} & \begin{bmatrix}\textit{try} & \\ \textsc{tryer} & \boxed{1}\\ \textsc{psoa-arg} & \boxed{2}\end{bmatrix}\end{bmatrix}\right]$$

b. $$\left[\textsc{ss}|\textsc{loc}\ \begin{bmatrix}\textsc{cat} & \begin{bmatrix}\textsc{head} & \textit{verb}\\ \textsc{subcat} & \left\langle \mathrm{NP}[\textit{nom}]_{\boxed{1}}\right\rangle\end{bmatrix}\\ \textsc{cont} & \begin{bmatrix}\textit{try} & \\ \textsc{tryer} & \boxed{1}\\ \textsc{psoa-arg} & \boxed{2}\end{bmatrix}\end{bmatrix}\right]$$

Here, in the resulting lexical entry, the subject is assigned a semantic role; its index appears as an argument of the *try*-relation in the TRYER slot. So the RP is not violated and (27a) will be licensed.

It is not hard to see that the ACP makes the same predictions. The input object can be a raising verb as much as an equi verb. In the case of a raising verb, not only the indices of the input's subject and that of the verbal complement would be identical, but rather the *synsem* objects as a whole would be structure shared. But so, trivially, would then be the index. The index of the subject is specified as referential.[22] Thus it is non-expletive and, in the case of raising verbs, also a non-role-filler. So, by the ACP, the *synsem* object that bears this index (here, the subject) would have to be on the SUBCAT list of some further element on the SUBCAT list of the output. Since there is no other such element on the rest of list $\boxed{1}$ (as it is tagged in the lexical rule), the output does not satisfy the raising principle

[22]If this were not the case, the rule would license sentences like *There/it seem(s)* for expletive *there*/*it*: Since the rule removes the VP complement, there would be nothing left that the CVP could require to be token-identical or index-shared with the subject, and since the subject is expletive, the ACP would not have to say anything either. So the sentence would be licensed.

and hence is not licensed. As desired, the output of the rule is licensed for equi verbs like *try*. In their CONTENT values, the subject's index fills a role, so the VP complement may be missing.

4.3 Conclusions

I have shown the informal meta-level RAISING PRINCIPLE of Pollard and Sag (1994) to be replaceable by a fully formalised description-level alternative while preserving the positively intended effects of the original RP. Having a precise account of the enforcement of the raising configuration and likewise the possibility to employ it when reasoning about the output values of description-level lexical rules might make more detailed discussions, analyses of generalisations about raising phenomena in the framework of HPSG possible.

References

De Kuthy, Kordula. 2000. *Discontinuous NPs in German – A Case Study of the Interaction of Syntax, Semantics and Pragmatics*. Ph. D.thesis, Universität des Saarlandes, Saarbrücken.

Meurers, Walt Detmar. 2001. On Expressing Lexical Generalizations in HPSG. *Nordic Journal of Linguistics* 24(2).

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

Przepiórkoswki, Adam and Rosen, Alexandr. 2004. Czech and Polish Raising/Control with or without Structure Sharing. In *Research in Language 3*, pages 33–66.

Richter, Frank. 2004. *A Mathematical Formalism for Linguistic Theories with an Application in Head-Driven Phrase Structure Grammar*. Phil. dissertation (2000), Eberhard-Karls-Universität Tübingen.

Tseng, Jesse L. 2007. English prepositional passive constructions. In Stefan Müller (ed.), *The Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar*, pages 271–286, Stanford: CSLI Publications.

On Predication

Stefan Müller

Freie Universität Berlin

Proceedings of the HPSG09 Conference

University of Göttingen

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

This paper discusses copula constructions in English, German, and Danish and argues that a uniform analysis of all copula constructions is inappropriate. I provide evidence from German that there should be a raising variant of the copula in addition to an identificational copula. A unary schema is provided that maps referential NPs that can be used as arguments onto predicational NPs. Data from Danish shows that predicational NPs can be subjects in specificational structures. An account for such specificational structures is provided and the different behaviour of predicational and specificational structures with regard to question tags is explained. A similar contrast can be found in German left dislocation structures, which follows from the assumptions made in this paper.

A modified treatment of complex predicate formation allows for a reduction of selectional features (that is abolishing of XCOMP or VCOMP) and for a uniform treatment of predicational phrases in copula constructions and resultative secondary predicates. This yields an account for constituent order variants that remained unexplained by earlier analyses.

1 The Phenomena

Research on copula structures has a long tradition (see Mikkelsen, To appear for an overview). One important question is the question of how many copulas are needed for the observable syntactic patterns and the respective meanings that can be expressed. I follow recent research in assuming that there are basically three types of copula constructions, two of which are order variants of each other (Section 1.1). Section 1.2 discusses V2 languages like Danish and German and compares English and Danish to German, which has rather free constituent order in general. Section 1.3 shows that one of the copula constructions is a raising construction and Section 1.4 discusses the formation of predicate complexes.

1.1 Equational, Predicational, and Specificational Constructions

Recent research on predication distinguishes three types of copula structures: equational, predicational, and specificational structures (Mikkelsen, To appear). In equational structures two expressions of the same type are equated. Examples of this type are given in (1):

(1) a. Cicero is Tully.
b. That must be her.

[†]I want to thank the audience of the HPSG conference for discussions. Special thanks go to Frank van Eynde, Doug Arnold, and Berthold Crysmann for discussion and pointing out interesting data. Thanks to Jakob Maché for comments on an earlier version of this paper.

I thank Philippa Cook for proof reading.

The work reported in this paper was supported by grants by the Deutsche Forschungsgemeinschaft (MU 2822/2-1 and MU 2822/1-1).

In (1a) two proper nouns are equated: that is, it is expressed that the referents of the two referential NPs are identical. Similarly, two pronouns are equated in (1b).

Mikkelsen gives the following examples for predicational constructions:

(2) a. Harvey/my brother/the guest of honor/she/everyone/noone was [happy].
  b. Sylvia is [from Seattle].
  c. Sylvia is [an architect].
  d. Sylvia is [the architect on that project].
  e. Sylvia is [my friend].
  f. Sylvia is [mayor of Seattle].

As the examples show, the predicate complement can be an AP, PP, NP or a noun with a complement. Mikkelsen claims that (2f) is an instance of an $\overline{\mathrm{N}}$ predicate (NP in her terminology), but the class of such predicates is smaller: It is basically nouns with their complements, but without modifiers:

(3) * He is new mayor of Seattle.

In English there seems to be a uniqueness restriction on determinerless predication. Sentences like those in (4) are ungrammatical:

(4) * He is sanator/teacher.

In comparison, the equivalents of (4) are possible in German:

(5) Er ist Lehrer.
  he is teacher
  'He is a teacher.'

As Mikkelsen (2005, p. 70–72) points out, question tags agree with the subject in predicational constructions in gender as they do in non-predicational structures:

(6) a. The guest of honor was happy, wasn't she/he/*it?
  b. The guest of honor spoke after dinner, didn't she/he/*it?

Apart from equational and predicative constructions a third type is identified in the literature. Mikkelsen gives the following example for what she calls a specificational construction:

(7) a. The director of *Anatomy of a Murder* is Otto Preminger, isn't it?
  b. The director of *Anatomy of a Murder*, that's Otto Preminger.

Here the post-copular NP is a proper name, that is, clearly referential. The pre-copular constituent contributes the predication. Interestingly, the pronoun *it* is used in question tags and the pronoun *that* in left dislocation structures. This test shows that the subject in (7) is not referential, but rather predicational. Specificational structures can be regarded as a variant of predicational structures with the predicational NP realized in pre-copula position.

While predicational structures are possible with verbs like *consider*, specificational and equational structures require the copula to be present (Rothstein, 1995, p. 32):

(8) a. I consider [Sylvia my best friend]. (predicational)
b. I consider [my best friend *(to be) Sylvia]. (specificational)
c. I believe [that/her *(to be) Sylvia]. (equational)

1.2 German, English, Danish: Specificational Constructions, Question Tags, and Left Dislocation

Evidence from question tags was used to argue for a special type of copula construction in English: Specificational constructions. The situation is more complicated in a language like Danish: Danish is a V2 language, so the orders with a predicative element in pre-copula position could be derived by fronting the predicate rather than the subject of a canonical predicational construction. However, there is a test that helps to identify which element is the subject: The negation attaches to the VP. For subordinate and main clauses we get the following structures:

(9) a. subject negation verb complements (subordinate)
b. verb subject negation complements (main clause, V1)

A V2 clause is derived from (9b) by fronting one constituent. Given this background we can show that Danish also has specificational structures in which the subject of the clause is the predicate. Since the post-negation position in (10b) is filled by *Max*, *Vinderen* has to be extracted from the pre-negation position and hence, it has to be the subject of the clause.

(10) a. Max$_i$ er $\_{i}$ ikke vinderen, er han vel. (Max= Subj, vinderen = Comp)
Max is not winner.DEF is he not
'Max is not the winner.'
b. Vinderen$_i$ er $\_{i}$ ikke Max, er det vel. (Max= Comp, vinderen = Subj)
winner.DEF is not Max is it not
c. Vinderen$_i$ er Max ikke $\_{i}$, er han vel. (Max= Subj, vinderen = Comp)
winner.DEF is Max not is he not

Interestingly, this corresponds to the question tags used in the sentences.

German differs from both English and Danish in another dimension: It is a language with rather free constituent order, so a test like the position of negation cannot be used for German. However, predicative elements can still be distinguished from referential ones: In left dislocation structures *das* is used for predicational elements and the genus agreeing *der*/*die*/*das* for referential elements.

(11) a. Klug / ein Mörder, das / *der ist Peter. (predicational)
smart a murderer that that is Peter
'Peter is smart / a murderer.'

b. Ja, aber Peter, der ist ein Mörder / nicht Klaus.
Yes, but Peter that is a murderer not Klaus
'Yes, but Peter is a murderer / not Klaus.' (predicational/equational)

So, there is evidence for a predication/equation difference in German, but not for a predication/specification distinction.

1.3 Raising

The predicative copula is usually analyzed as a raising predicate that does not contribute semantically, except for tense information in the case of finite forms of the copula (Paul, 1919, p. 41). One property of raising verbs is that they are not sensitive to the type of their arguments, for instance they allow for expletive subjects, which is – of course – compatible with the fact that they do not assign semantic roles to their arguments. An example for an adjective that allows for an expletive subject is *laut* ('loud'):

(12) In der Mensa ist es laut.
in the commons is it.EXPL loud
'It is loud in the commons.'

The adjective *laut* also has a non-expletive version, and (12) is actually ambiguous between the expletive and the non-expletive reading. With the expletive predicate, (12) means that the people, machines, or whatever, in the commons are loud, whereas in the non-expletive reading the *es* ('it') could refer to a child.

German is a language that has subjectless verbs and adjectives. Müller (2002, p. 72–73) discusses the following examples:[1]

(13) a. weil schulfrei ist
because school.free is
'because there is no school.'

b. weil ihm schlecht ist
weil him.DAT bad is
'because he is sick'

c. Für dich ist immer offen.
for you is always open
'It is always open for you.'

Again such data is consistent with a raising analysis that raises the subject of an embedded predicate if there is one but does not rule out embedded predicates that do not have a subject at all.

1.4 Predicate Complex Formation

Certain verbs form a predicate complex in languages like German, Dutch, Persian, and Hindi. The arguments of the verbs that are involved in complex formation can

[1] (13c) is quoted from Haider, 1986, p. 18.

be scrambled according to the general rules of the respective language. In addition parts of the predicate complex can be fronted while arguments of the fronted heads may be left behind. Adjuncts in pre-complex position can scope over different elements of the predicate complex. An industrial-strength overview of the phenomenon in German can be found in Bech, 1955. Bech coined the term coherent construction for verbal complexes. Analyses of the data in the framework of HPSG can be found in Hinrichs and Nakazawa, 1994; Kiss, 1995; Bouma and van Noord, 1998; Meurers, 2000; Kathol, 2000; Müller, 2002. Müller (2002) extended the verb complex analysis to verb adjective combinations. Since the focus of this paper is predicational constructions, I exclusively discuss copula constructions and other predicational structures here.

As within coherent combinations of verbs, different scopings can also be observed in copula constructions:

(14) weil ihr der Mann immer treu sein wollte.
because her.DAT the man.NOM always faithful be wanted.to
'because the man always wanted to be faithful to her.'
'because the man wanted to be always faithful to her.'

The sentence in (14) has the two readings that are indicated in the translation, but here the situation is less clear since the two readings may be due to the ambiguity between the modification of the copula and the modal. However, there are sentences like (15) where the adjective is fronted together with the adverbial.

(15) Immer treu wollte er ihr sein.
always faithful wanted.to he.NOM her.DAT be
'He wanted to be faithful to her forever.'

Due to the existence of such sentences, the possibility of adverbs modifying adjectives directly cannot be ruled out in general. Note furthermore, that the sentence in (15) is not ambiguous.

What is clear, however, is that the phrase *ihr immer treu* in (14) and (16) cannot be a closed AP in the wide scope reading since then the scoping of the adverb over a predicate outside the domain of the AP could not be explained.

(16) weil der Mann ihr immer treu sein wollte.
because the man.NOM her.DAT always faithful be wanted.to
'because the man always wanted to be faithful to her.'
'because the man wanted to be faithful to her forever.'

The example in (14) also shows that the subject of the adjective, which is also the subject of the modal, can appear between the adjective and its complement. The alternative order in (16) is also possible. See also den Besten, 1985, p. 60 on this point.

The examples discussed so far show that copula constructions with adjectives fulfill the criteria for so-called coherent constructions: Adjuncts can scope over predicates in the predicate complex, predicates can be fronted without their arguments, arguments of several heads can be scrambled with respect to each other.

However, there are also examples that are reminiscent of incoherent constructions: In (17) the adjectives are not adjacent to the copula but intraposed in the Mittelfeld:

(17) a. Sie wuchsen in einem gesellschaftlichen Klima auf, das *freier*
they grew in a social climate PART(up) that freer
in Deutschland nie *war*.[2]
in Germany never was
'They grew up in a social climate that was freer than ever in Germany.'

b. daß *ausschlaggebend für die Interpretation abgeleiteter Verben* bestimmte
that decisive for the interpretation derived verbs certain
semantische Interpretationsmuster *sind*, die sich [...][3]
semantic interpretation.models are which self
'that certain semantic interpretation models that are [...] are decisive for the interpretation of derived verbs.'

Due to space limitations the discussion of the data remains sketchy here, but a thorough discussion of the data can be found in Müller, 2002, Chapter 2.1.9.

In 2002, I focussed on adjectival predication, but of course the copula can be combined with predicative NPs and PPs as well. In contrast to adjectival predication, predicative NPs and PPs do not enter the predicate complex in the sense that the noun or preposition forms a complex with the copula. Instead nouns and prepositions that are used predicatively have to form full phrases and hence can be intraposed (that is, scrambled) (Müller, 1999, p. 173).

Resultative constructions with adjectival predicates behave similarly to copula constructions. Partial fronting and scrambling of arguments is allowed. However, PPs can be predicates in resultative constructions as well. Resultative constructions with PPs resemble incoherent constructions, while resultative constructions with adjectives allow for coherent constructions.

This section showed that predicative constructions can take part in cluster formation (primary and resultative predication with adjectives) but that there are also cases in which no complex formation takes place (primary predication with NPs and PPs, and resultative predication with PPs). An analysis should provide a unified account of these phenomena.

2 Previous Accounts

This section discusses previous proposals in the literature. I start with a lexical rule-based proposal to predication, continue with van Eynde's non-raising approach, and finish the section with a discussion of my earlier treatment of primary and secondary adjectival predication.

[2] taz, 01.07.1995, p. 10.

[3] In the main text of Kaufmann, 1995, *Konzeptuelle Grundlagen semantischer Dekompositionsstrukturen*, p. 162.

2.1 Pollard and Sag 1994 and Sag and Ginzburg 2000

Pollard and Sag (1994, p. 360) sketch the lexical rule in (18) that takes nouns as used in normal referential NPs like *a teacher* in (19a) and maps them onto another lexical item that can be used predicatively like in (19b).

(18) $\mathrm{N}[-\textsc{prd}, \textsc{subj}\ \langle\rangle]{:}[\textsc{restriction}\ \{\boxed{2}\}]_{\boxed{1}} \mapsto \mathrm{N}[+\textsc{prd}, \textsc{subj}\ \left\langle \mathrm{XP}_{\boxed{1}} \right\rangle]{:}\boxed{2}$

(19) a. A teacher laughs.
b. John is a teacher.

Ginzburg and Sag (2000, p. 409) give the following variant of the rule in (18):

(20) Singular Predicative Noun Lexical Rule:

$$\begin{bmatrix} \textsc{ss}|\textsc{loc}|\textsc{cat}|\textsc{head}\ n \\ \textsc{arg-st}\ \langle \boxed{1} \rangle \oplus \boxed{A} \\ lx \end{bmatrix} \Longrightarrow_{LR} \begin{bmatrix} \textsc{ss}|\textsc{loc}|\textsc{cat} & \begin{bmatrix} \textsc{head} & \begin{bmatrix} \textsc{agr}|\textsc{num}\ sg \\ \textsc{pred}\ + \end{bmatrix} \\ \textsc{spr} & \langle \boxed{1} \rangle \\ \textsc{subj} & \langle \boxed{2} \rangle \end{bmatrix} \\ \textsc{arg-st}\ \langle \boxed{2}, \boxed{1} \rangle \oplus \boxed{A} & \\ word & \end{bmatrix}$$

The lexical rule in (18) adds a subject to the valence features of the noun and by doing so makes it parallel to predicative adjectives. The copula and verbs like *seem* and *consider* are treated as raising verbs that raise the element in SUBJ and make it their own subject or – in the case of *consider* – object. Such a raising analysis of the copula and verbs like *consider* is also assumed by other researchers working on different languages (see for instance Müller, 2002, Chapter 2.2.7–8).

Pollard and Sag suggest that the element in the set of restrictions of the noun in the input of the rule is represented as the main semantic contribution of the resulting noun. So the contribution of *teacher* in (19b) is *teacher*'($\boxed{1}$), while it is $\boxed{1}|\{$*teacher*'($\boxed{1}$)$\}$ for (19a).[4] As Pollard and Sag point out, this analysis does not extend to proper nouns like those in (1a) for semantic reasons. Like most researchers Pollard and Sag (1987, p. 66) distinguish between the *be* of predication and the *be* of identity, and hence the lexical rule does not have to account for cases with two proper names or two pronouns.

As Kasper (1995) pointed out in unpublished work[5], the lexical rule-based analysis fails for examples that contain modifiers in the predicative phrase:

(21) He is a good candidate.

The classical analysis of adjuncts assumes that nominal modifiers attach to an $\overline{\mathrm{N}}$ and identify their referential index with the referential index of the noun. But if the

[4] The curly brackets around $\boxed{2}$ in the input are missing in Pollard and Sag's version of the lexical rule.

[5] See also Gerbl, 2007, p. 241.

semantic contribution of *candidate* is a predicate rather than an index, modification cannot apply as usual.[6]

2.2 Van Eynde 2008

Van Eynde suggests the following alternative to the raising analysis: Lexical items for *seems* as in (22a) are constrained by (23) and items like *consider* in (22b) are constrained by (24).

(22) a. John seems a nice guy.

b. Bob considers his brother a genius.

(23) a1-pred-lex $\Rightarrow$

$$\left[\begin{array}{ll} \textsc{arg-st} & \left\langle \text{NP}_{\boxed{1}} \text{ (, PP}_{\boxed{2}}\text{), Z}_{\boxed{3}} \right\rangle \\ \textsc{ss}|\textsc{loc}|\textsc{cont}|\textsc{nucl} & \left[\begin{array}{ll} \textsc{experiencer} & \boxed{2} \\ \textsc{soa-arg}|\textsc{nucl} & \left[\begin{array}{ll} \textsc{inst} & \boxed{3}\ \textit{index} \\ \textsc{theme} & \boxed{1}\ \textit{index} \\ \textit{coref-rel} & \end{array}\right] \\ \textit{exp-soa-rel} & \end{array}\right] \end{array}\right]$$

(24) a2-pred-lex $\Rightarrow$

$$\left[\begin{array}{ll} \textsc{arg-st} & \left\langle \text{NP, NP}_{\boxed{2}}\text{, Z}_{\boxed{3}} \right\rangle \\ \textsc{ss}|\textsc{loc}|\textsc{cont}|\textsc{nucl} & \left[\begin{array}{ll} \textsc{soa-arg}|\textsc{nucl} & \left[\begin{array}{ll} \textsc{inst} & \boxed{3}\ \textit{index} \\ \textsc{theme} & \boxed{2}\ \textit{index} \\ \textit{coref-rel} & \end{array}\right] \\ \textit{soa-rel} & \end{array}\right] \end{array}\right]$$

By assuming these lexical entries van Eynde can analyze the sentences in (22) with normal nouns without having to assume a separate predicative lexical item for the predicative usage of the noun.

Van Eynde assumes that all predicate selectors contribute such semantic information and explicitly includes the copula *be* here. He argues that the dative of judgment depends on the copula, which he takes as evidence for its relational status:

(25) Es ist mir zu kalt.
it is me.DAT too cold
'It is too cold for me.'

Traditionally it is said that this dative depends on the *zu* (How this is captured in HPSG is a different question. The analysis is not trivial since dative and *zu* can be discontinuous). Note, however, that van Eynde would be forced to assume empty

[6]This may not be an issue if an MRS semantics (Copestake et al., 2005) is assumed. However, one would have to be willing to claim that the type of the index of *candidate* is not changed by the predication lexical rule.

copulas in prenominal position if he were to apply his argument to the following data:

(26) a. bis auf das mir zu kalte Ziel Spitzbergen
until on the me.DAT too cold goal Spitsbergen
'except for the goal Spitsbergen, which is too cold for me'

b. die mir zu warme Book-Unterseite
the me.DAT too warm bottom.of.the.Book
'the bottom of the Book, which is too warm for me'

Here we have *mir zu warme* and *mir zu kalte*, with *zu* present but in a prenominal context in which copulas are never present.

There are examples of copula constructions with a dative without a degree word like *zu* ('to') or *genug* ('enough') being present:

(27) Du bist mir ja ein schöner Vorsitzender!
you.NOM are me.DAT PART a nice chair
'You are a nice chair to me.'

Such sentences are used to express that the speaker thinks that the addressee does not have all properties that are usually assigned to the predicative noun. Such datives should be handled as scopal modifiers that encapsulate the meaning of the predication similar to the way suggested by van Eynde in (23).

Another example of datives in copula constructions is shown in (28):

(28) Er war dem König ein treuer Diener.
he.NOM was the king.DAT a faithful servant
'He was a loyal servant of the king.'

I would argue that such datives are adjuncts as well. They are of the type we see in (29):

(29) Er bemalt dem König den Tisch.
he.NOM paints the king.DAT the table.ACC
'He paints the table for the king.'

The verb *bemalen* ('paint') is a transitive verb and the dative is a modifier that can be used to express the benefactive/malefactive of the event (Wegener, 1985).[7]

Van Eynde's analysis works for the given examples, but the argumentation against the raising analysis is not convincing. In addition, the identity analysis faces several problems.

The first problem is that pronouns and proper names cannot be used as predicates in such constructions:

(30) a. * He seems him.

b. * He seems John Malkovich.

[7]Since such datives interact with the dative passive, they are probably licensed by a lexical rule that adds the dative to the argument list of a verb.

Here the copula has to be used:

(31) a. He seems to be him.
     b. He seems to be John Malkovich.

The same is true for gerunds and infinitives:

(32) a. * The greatest pleasure on earth seems eating oysters . . . .
     b. * His main worry now seems to get rid of his detractors.
     c. The greatest pleasure on earth seems to be eating oysters . . . .
     d. His main worry now seems to be to get rid of his detractors.

This difference is captured by an analysis that treats *seem* as a raising verb and assumes that there is an equational copula *be*. Since *seem* does require a predicative phrase as complement, gerunds and infinitives are excluded and since the identity copula can be combined with gerunds and infinitives, examples like (32c,d) are well-formed.

Secondly, there seems to be no way to account for the differences in question tags and pronouns in left dislocation structures that were discussed in Section 1.1.

In addition there is a very general problem of the analysis: It does not extend to predicates with an expletive subject (12) or predicates that do not have a subject at all (13). In both cases there is nothing present that could be "coreferential" with the adjectival predicate.

Van Eynde (presentation at HPSG 2009) suggests that the THEME role of the *coref-rel*' is optionally filled: that is, in the case of expletives there is no index linked to THEME. He argues that this is parallel to cases like (33):

(33) a. He eats pizza.
     b. He eats.

In (33b) the object of *eats* remains implicit. Note that this analysis introduces a disjunction in the lexical item for the copula, namely a disjunction between referential and expletive indices of the subject NP. In addition one would need another disjunction that accounts for the fact that the subject can be missing altogether. Therefore one would have to have three versions of the copula: one for clauses with referential subjects, one for clauses with expletive subjects, and one for clauses without subject. The big problem for such a proposal is that it has to be ensured that the right copula is used with the right embedded predicate. For instance it is impossible to use (13b) with a subject:

(34) * weil der Mann ihm schlecht ist
       because the man.NOM him.DAT sick is

Similarly, expletives are impossible in normal prediative constructions:

(35) Es ist klug.
     it is smart
     'He/she is smart.'

(35) does not have a reading in which nobody is smart or there is generic smartness. The *es* has to be referential and it has to refer to something that has neuter gender as for instance *Mädchen* ('girl') or *Bürschlein* ('boy'). This means that the subject of the copula has to be expletive if and only if the embedded predicate allows for an expletive. It can be missing if and only if the embedded predicate does not require a subject. This is best captured by a raising analysis.

2.3 Müller 2002

Some authors have suggested using a special valence feature called XCOMP or VCOMP for the selection of an argument that enters predicate complex formation (Chung, 1993; Rentier, 1994; Müller, 1997; Kathol, 1998). Müller (2002, p. 103) extended the verb complex analysis of other authors to copula constructions and resultative secondary predicates. He gave the following lexical item for the copula:

(36) *sein* (predicative copula, according to Müller (2002, p. 103)):

$$\begin{bmatrix} \text{SUBCAT} & \boxed{1} \oplus \boxed{2} \\ \text{XCOMP} & \left\langle \begin{array}{l} \text{ADJ}[\text{MOD } \textit{none}, \text{ PRD } +, \text{ SUBJ } \boxed{1}, \text{ SUBCAT } \boxed{2}, \\ \qquad \text{XCOMP } \langle\rangle, \text{ LEX } +] \end{array} \right\rangle \end{bmatrix}$$

The copula raises both the subject, if there is one ($\boxed{1}$), and other arguments of the embedded adjective ($\boxed{2}$). The predicative adjective is required to be LEX+. Therefore it forms a complex with the copula directly and all its arguments are raised.

The problem with this lexical item is that it specifically selects a predicative adjective. Müller selected all verbs that take part in complex formation via XCOMP, but those that were realized as full phrases – that is in so-called incoherent constructions – were selected via SUBCAT. The problem that results from this treatment is that two lexical items for the predicative copula are needed, one that selects NP and PP predicates and one for adjectival predicates. Similarly the lexical rule for resultative predication selects the result predicate via XCOMP. Since both PPs and adjectives can function as the result predicate in German but only structures with adjectives fulfill the criteria for coherent constructions, a more general treatment of the facts is desirable.

3 The Analysis

As was discussed in Section 2.1, lexical rule-based approaches to predicative NPs have a problem with the semantic type of predicative nouns. NP internally the nouns behave like normal nouns, only the complete NP has a predicative function. The problem can be solved by assuming Schema 1 instead of the lexical rule in (20).[8] This unary projection applies to a full NP and licenses the predicative NP

[8]Gerbl (2007, p. 241–242) independently suggested a similar solution. See also Partee, 1987.

Schema 1 (Predicative NP Projection Schema)
np-pred-phrase ⇒

$$
\begin{bmatrix}
\text{SYNSEM}|\text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{PRD} & + \\ \text{MOD} & \mathit{none} \\ \text{SUBJ} & \left\langle \text{NP}_{\boxed{1}} \right\rangle \\ \mathit{noun} & \end{bmatrix} \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \text{IND} & \boxed{0} \end{bmatrix} \end{bmatrix} \\
\text{C-CONT} & \begin{bmatrix} \text{RELS} & \left\langle \begin{bmatrix} \text{ARG0} & \boxed{0}\ \mathit{event} \\ \text{ARG1} & \boxed{1} \\ \text{ARG2} & \boxed{2} \\ \mathit{equal\text{-}rel} & \end{bmatrix} \right\rangle \\ \text{H-CONS} & \langle\rangle \end{bmatrix} \\
\text{NON-HEAD-DTRS} & \left\langle \begin{bmatrix} \text{SYNSEM}|\text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{MOD} & \mathit{none} \\ \mathit{noun} & \end{bmatrix} \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \text{IND} & \boxed{2}\ \mathit{index} \\ \mathit{npro} & \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle
\end{bmatrix}
$$

(PRD+) with an appropriate SUBJ value. The referential index of the subject NP ($\boxed{1}$) is related to the referential index of the daughter NP ($\boxed{2}$). The relation is introduced constructionally via C-CONT (see Copestake, Flickinger, Pollard and Sag, 2005 on semantic composition and C-CONT). The unary branching rule cannot apply to its output since the daughter NP has to have an IND value of type *index* and the resulting sign has an IND value of type *event*.

I assume that SUBJ is not a valence feature (Pollard, 1996; Kiss, 1992). In configurational languages like Danish and English the subject is mapped to SPR for those heads which allow direct combination with their subject. For non-configurational languages the subject of finite verbs is mapped to the COMPS list and the one of non-finite verbs is mapped to SUBJ, since it is never combined with the verb directly.

Note that in this analysis there is still ambiguity between NPs that can function as complements and NPs that can function as predicates – something that van Eynde criticized – but the ambiguity is reduced considerably since it is only present at the NP level and not for all nominal projections. So there is no predicative version of *good candidate*.

The analysis changes the semantic type of an NP and its syntactic properties. It is interesting to note that a similar analysis is necessary for temporal NPs: As

Flickinger (2008, p. 91–92) points out, it is not just simple NPs that can act as modifiers of verbs. The time nouns can be embedded inside of a more complex NP, as (37) shows.

(37) a. Kim disappears those days.
b. Kim disappears some of those days.

Therefore a treatment in which the time noun has a MOD value that allows it to modify a verb is not appropriate. Further evidence for an analysis as unary projection is provided by parallel German examples:

(38) a. Er arbeitete den größten Teil der Nacht.
he worked the.ACC largest part of.the.GEN night
'He worked almost all night.'

b. Er arbeitete die halbe Nacht.
he worked the.ACC half.ACC night
'He worked half of the night.'

In (38a) the time expression *der Nacht* is genitive but the whole NP is accusative. This accusative is called a semantic case. It is connected to the function of the NP and is not assigned by the verb. It is clear from data like (38a) that an analysis like the one suggested by Müller (2007, p. 226) that assigns both function (i.e. MOD value) and case lexically cannot explain the data in (38a). Hence we have evidence from another area of grammar that type shifting phrasal schemata are needed.

In addition to the unary branching schema one would keep the lexical rule for sentences with determinerless predication like (2f). The noun *mayor* is mapped to a predicative version. This predicative version can be combined with its arguments but since the index is of the wrong type it cannot be combined with adjuncts.

Turning to the lexical item for the copula, I suggest the following for German: This lexical entry is similar to the one suggested by Müller (2002, p. 103) in that

(39) Entry for the predicative copula for German:

$$\begin{bmatrix} \text{ARG-ST} & \boxed{1} \oplus \boxed{2} \oplus \left\langle \begin{bmatrix} \text{HEAD} & \begin{bmatrix}\text{PRD} & + \\ \text{SUBJ} & \boxed{1}\end{bmatrix} \\ \text{COMPS} & \boxed{2} \\ \text{CONT} & \begin{bmatrix}\text{IND} & \boxed{3}\\ \text{LTOP} & \boxed{4}\end{bmatrix}\end{bmatrix} \right\rangle \\ \text{CONT} & \begin{bmatrix}\text{IND} & \boxed{3}\\ \text{LTOP} & \boxed{4}\end{bmatrix} \\ \text{RELS} & \langle\rangle \end{bmatrix}$$

both the elements of SUBJ and of COMPS of the embedded predicate are raised to the ARG-ST list of the copula. The elements at the COMPS list of the embedded predicate are raised in addition to the elements in SUBJ since German forms a verbal complex and predicative constructions like copula constructions and resultative constructions take part in complex formation. The formation of verbal complexes is analyzed via argument attraction (Hinrichs and Nakazawa, 1994; Kiss, 1995).

Note that nothing is said about the actual members of the lists. It is therefore possible to handle the cases in (40) as well as the subjectless examples that were given in (13).

(40) a. weil er auf seinen Sohn stolz ist
because he.NOM on his son proud is
'because he is proud of his son'

b. weil er klug ist
because he.NOM smart is
'because he is smart'

In the analysis of (40a), $\boxed{1}$ contains the subject (*er*) and $\boxed{2}$ the PP (*auf seinen Sohn*). In the analysis of (40b), $\boxed{1}$ contains the subject (*er*) and $\boxed{2}$ is the empty list. In the analysis of (13b), $\boxed{1}$ is the empty list and $\boxed{2}$ contains the dative object *ihm* ('him'). In the analysis of (13a), both $\boxed{1}$ and $\boxed{2}$ are the empty list.

The same lexical item can be used for English if one assumes that head-complement phrases require their non-head daughter to be saturated. If this assumption is made, it follows that the COMPS list of the predicative argument ($\boxed{2}$) has to be the empty list if this argument is used in a head-complement phrase. Hence, nothing but the subject is raised from the predicative element. German and Dutch differ from English and Danish in allowing complex formation (see Section 3.1). When predicate complexes are formed, $\boxed{2}$ can be non-empty, since the predicate complex schema does not impose any restrictions on the length of the COMPS list of its non-head daughter.

The copula does not contribute semantically, hence the RELS list is empty. The INDEX value is shared with that of the embedded predicate. The copula enters inflectional lexical rules and these rules introduce relations that provide information about tense. The arguments of the respective relations are of type *event*.[9] Therefore, the INDEX value of the copula in (39) is *event* and hence the INDEX value of the embedded predicate has to be of type *event* as well. The requirement that the predicative element is of type *event* will play an important role in Section 3.4 on raising nouns in English.

3.1 Raising and Complex Formation

There is another important aspect regarding the lexical item in (39): The predicate is selected via COMPS rather than VCOMP or XCOMP (see Section 2.3). With a uniform selection of verbal complements via COMPS it is possible to treat optionally coherent verbs like *versuchen* with one lexical item (Kiss, 1995, p. 178). The control verb does not specify whether it forms a verbal complex with the embedded verb or not. It does not mention the LEX value of the embedded verbal element.

[9] *event* is to be understood as the most general type referring to situations. The only thing that is important here is that the type differs from the type used to refer to objects.

Because of this we can analyze examples with a predicate complex as in (41a) and examples like (41b) with so-called intraposition:

(41) a. Karl hat das Buch nicht [zu lesen versucht].
Karl has the book not to read tried
'Karl did not try to read the book.'

b. Karl hat [das Buch zu lesen] nicht versucht.
Karl has the book to read not tried
'Karl did not try to read the book.'

In comparison verbs like *scheinen* ('to seem') or modals, that obligatorily construct coherently, select a verbal complement that is LEX+. Consequently they do not allow for intraposition of a VP complement, but require complex formation.

Müller (2002, p. 112) criticized Kiss's analysis of optional coherence because it also licences unwanted structures like (42) and hence results in spurious ambiguities.

(42) weil Karl das Buch [[dem Mann zu geben] verspricht].
because Karl the book the man to give promises
'because Karl promises to give the book to the man.'

In (42) *versprechen* is combined with a partly saturated verbal projection *dem Mann zu geben* and the non-saturated argument *das Buch* is raised and combined with *dem Mann zu geben verspricht* in a later step. However, this structure is excluded if arguments are required to be saturated and elements of the predicate complex are required to be LEX +.[10] Hence, I assume the Schemata 2 and 3.

Schema 2 (Head-Complement-Schema)
head-complement-phrase ⇒

$$\begin{bmatrix} \text{SYNSEM|LOC|CAT|COMPS } \boxed{1} \oplus \boxed{3} \\ \text{HEAD-DTR|CAT|COMPS } \boxed{1} \oplus \langle \boxed{2} \rangle \oplus \boxed{3} \\ \text{NON-HEAD-DTRS } \left\langle \begin{bmatrix} \text{SYNSEM } \boxed{2} \begin{bmatrix} \text{LOC|CAT|COMPS } \langle\rangle \\ \text{LEX } - \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

Schema 2 shows the version of the schema for languages with free constituent order. In languages like English, that have a strict order, $\boxed{3}$ is the empty list (Müller, In Preparation). With the new treatment of predicate selection via COMPS, it is not required that predicative PPs or NPs are part of the predicate complex as was suggested by Müller (2002) for PPs in resultative constructions. Instead they can be analyzed as head-complement structures.

Returning to the copula, it allows the embedding of fully saturated phrases like predicative NPs and PPs but also allows for the formation of a predicate complex

[10]This is a simplification, since I assume that the so-called Third Construction is also an instance of predicate complex formation. Schema 3 has to be refined in order to allow non-lexical material in the complex if the conditions of the Third Construction are met. See Müller, 1999 for details.

Schema 3 (Predicate Complex Schema)
head-cluster-phrase ⇒

$$\begin{bmatrix} \text{SYNSEM} & \begin{bmatrix}\text{LOC}|\text{CAT}|\text{COMPS}\ \boxed{1}\end{bmatrix} \\ \text{HEAD-DTR} & \begin{bmatrix}\text{SYNSEM}|\text{LOC}|\text{CAT}|\text{COMPS}\ \boxed{1} \oplus \langle\ \boxed{2}\ \rangle\end{bmatrix} \\ \text{NONHEAD-DTRS} & \langle\ [\ \text{SYNSEM}\ \boxed{2}\ [\ \text{LEX}\ +\ ]\ ]\ \rangle \end{bmatrix}$$

consisting of adjective and copula. Since coherence is optional we can explain so-called focus movement of adjectives as in (17), something that was noted by Müller (2002, p. 69) but not treated in his analysis.

3.2 German, English, Danish: Specificational Constructions, Question Tags, and Left Dislocation

The difference between specificational and predicational structures is best captured by generalizing the German lexical item for the copula: Instead of using the append operator ($\oplus$) to concatenate two lists as in (39), the more general version of the copula uses the shuffle operator ($\bigcirc$):

(43) Entry for the Danish and English predicational and specificational copula:

$$\begin{bmatrix} \text{ARG-ST}\ \ (\boxed{1} \oplus \boxed{2}) \bigcirc \left\langle \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{PRD}\ + \\ \text{SUBJ}\ \boxed{1} \end{bmatrix} \\ \text{COMPS} & \boxed{2} \end{bmatrix} \right\rangle \end{bmatrix}$$

Since English and Danish do not form predicate complexes there is just the Head-Complement Schema, which requires complements to be fully saturated. Hence $\boxed{2}$ is the empty list. $\boxed{1}$ is a list containing exactly one element, since neither English nor Danish allows for subjectless constructions. Shuffle combines the elements of two lists in any order provided the order of the elements in the respective lists is preserved. In the example above we have a trivial case: Two lists with exactly one element are shuffled. The result is that the predicative argument is ordered first or last. The lexical item for the copula gets inflected and the first element of the ARG-ST list is mapped to SPR and the rest of the list to COMPS.

Gerbl (2007, p. 102, 190–191) pointed out that there are additional constraints regarding extraction of or extraction out of the post-copular phrase in specificational structures. These can be formalized by an additional implicational constraint with a complex antecedent, which is not given here due to space limitations.

3.3 Raising and Nonlocal Dependencies

The treatment of raising in (39) differs in an interesting way from the characterization of raising as it is given in Ginzburg and Sag (2000, p. 22). Ginzburg and Sag assume the following constraint:

(44) $[\text{ARG-ST}\ \langle\ [\ \text{LOC}\ \boxed{1}\ ],\ [\ \text{SUBJ}\ \langle\ [\text{LOC}\ \boxed{1}\ ]\rangle\ ]\ \rangle\ ]$

This version of raising differs from earlier proposals in that only LOCAL values are shared instead of whole *synsem* objects. The reason for this treatment is that one would get problems with the lexical SLASH amalgamation that was suggested by Bouma et al. (2001): if the whole *synsem* object was shared there would be SLASH amalgamation in the subject and in the phrase from which the subject is raised, an unwelcome result (Ginzburg and Sag, 2000, p. 21, fn. 8). So if one were to assume an amalgamation account of nonlocal dependencies for German, one would be forced to use a relational constraint that walks through lists and produces a copy of the list that contains elements that share the LOCAL values with the elements of the list from which they are raised. Note that assuming a disjunction that refers to the arity of the SUBJ list is not sufficient for German since complements are raised as well and the number of elements on the COMPS list is restricted by performance factors only (Müller, 2004, p. 220).

Rather than complicating the analysis of raising, I will drop the amalgamation analysis and return to an analysis that introduces nonlocal dependencies in syntax (through a trace or a unary branching projection).[11] As Bouma, Malouf and Sag (2001, p. 29) point out, the amalgamation analysis is not necessary to account for extraction path marking phenomena. If adjuncts are registered at a head (either in an adjunct as dependents analysis or via a mechanism of the kind suggested by Levine and Hukari (2006, Chapter 3.7.2)), a pathway marking element can attach to the head and check its INHER|SLASH value and the SLASH values that are contributed by the elements in the COMPS list and the SLASH values of the registered adjuncts.

3.4 Predicative Raising-Nouns

Doug Arnold brought the following kind of predicative noun phrases to my attention:

(45) a. He is a dead cert/a certainty to win.
b. This is a cinch to prise off.

These nouns are raising nouns and can only be used predicatively:

(46) a. * A dead cert/a certainty to win came into the room.
b. * A cinch to prise off came into the room.

I assume the lexical entry in (47) for a noun like *cert*. This noun is similar to normal nouns in that its semantic contribution is a referential index with person and number features and in that it takes a determiner as specifier that has to agree with the noun in number. The noun takes as its complement a VP and raises the missing specifier of this VP (the subject) to its own SUBJ list. The referential index of the noun is linked to the first argument of the relation that is contributed by the noun and the semantic contribution of the VP is linked to the second argument.

[11] See Bender, 2002, Müller, To appear, and Sag, Wasow and Bender, 2003, p. 463–464 for arguments that empty elements actually simplify grammatical descriptions.

(47) *cert*:

$$\begin{bmatrix} \textsc{cat} & \begin{bmatrix} \textsc{head} & \begin{bmatrix} \textsc{prd} & + \\ \textsc{subj} & \langle \boxed{1} \rangle \\ \multicolumn{2}{l}{\textit{noun}} \end{bmatrix} \\ \textsc{spr} & \left\langle \text{Det}[\textsc{num}\ \boxed{2}] \right\rangle \\ \textsc{comps} & \left\langle \text{VP}[\textsc{spr}\ \langle \boxed{1} \rangle]{:}\boxed{3} \right\rangle \end{bmatrix} \\ \textsc{cont} & \begin{bmatrix} \textsc{ind}\ \boxed{4} & \begin{bmatrix} \textsc{per} & \textit{3} \\ \textsc{num} & \boxed{2}\ \textit{sg} \\ \multicolumn{2}{l}{\textit{index}} \end{bmatrix} \end{bmatrix} \\ \textsc{rels} & \left\langle \begin{bmatrix} \textsc{arg0} & \boxed{4} \\ \textsc{arg1} & \boxed{3} \\ \multicolumn{2}{l}{\textit{cert}} \end{bmatrix} \right\rangle \end{bmatrix}$$

Since the noun is specified to be PRD+, all projections of this noun are excluded in positions in which non-predicative NPs are required and hence sentences like (46) are ruled out.

After combination of this lexical item with the VP complement, the determiner, and possibly some adjuncts, the resulting phrase can function as the daughter in the Predicative NP Projection Schema. It is then projected to an NP that has an index of type *event*. The resulting NP is compatible with the requirement of the copula that the predicative argument has to have an index of type *event*.

One thing is missing to make the analysis of sentence like (45) complete: The Predication Schema does not identify the HEAD value of the non-head daughter with the HEAD value of the mother. After all it usually applies to non-predicative NPs and hence, sharing of the HEAD values would cause conflicts in these cases. Therefore the SUBJ value of the raising noun NP is not identified with the SUBJ value in the mother node. This has to be stated explicitly for the cases under discussion:

(48)

$$\begin{bmatrix} \textsc{non-head-dtrs}\ \langle\ [\ \textsc{synsem|loc|cat|head|prd}\ +\ ]\ \rangle \\ \textit{np-pred-phrase} \end{bmatrix} \Rightarrow \begin{bmatrix} \textsc{synsem|loc|cat|head|subj}\ \boxed{1} \\ \textsc{non-head-dtrs}\ \left\langle [\ \textsc{synsem|loc|cat|head|subj}\ \boxed{1}\ ] \right\rangle \end{bmatrix}$$

The constraint in (48) is the only stipulative part of the analysis, but I see no other way to acount for this data if one does not want to employ several semantic features for external and internal content of phrases as was done by Kasper (1995).

4 Conclusion

This paper provided the basic building blocks for predicational and specificational constructions. An entry for the equational copula was not given, but I consider this trivial.

I have shown that the arguments provided by van Eynde for an identity analysis without raising are not convincing. In addition, in his analysis there are problems with pronouns in predication structures, the analysis cannot account for question tags and pronouns in left dislocation structures, and the analysis does not extend to subjectless constructions.

I suggest returning to a raising analysis of predication that raises the complete value of SUBJ of the embedded predicate rather than identifying LOCAL values of raised subjects. The predication lexical rule was recoded as a unary branching immediate dominance schema, which allows the inclusion of modifiers in the NP. In addition it was suggested to dispense with the XCOMP feature and to return to a COMPS-based analysis in which predicative and non-predicative arguments are selected uniformly via COMPS. This makes it possible to treat the various predication structures as optionally coherent constructions.

The analysis has been implemented in the TRALE system as part of grammar fragments of German and Danish. These grammars share a core grammar with grammars for Persian, Mandarin Chinese, and Maltese. The respective grammars can be downloaded at http://hpsg.fu-berlin.de/Software/.

References

Bech, G. 1955. *Studien über das deutsche Verbum infinitum*. Tübingen: Niemeyer.

Bender, E. M. 2002. *Syntactic Variation and Linguistic Competence: The Case of AAVE Copula Absence*. Ph. D.thesis, Stanford University.

Bouma, G., Malouf, R. and Sag, I. A. 2001. Satisfying Constraints on Extraction and Adjunction. *Natural Language and Linguistic Theory* 19(1), 1–65.

Bouma, G. and van Noord, G. 1998. Word Order Constraints on Verb Clusters German and Dutch. In Hinrichs et al. (1998), pp 43–72.

Chung, C. 1993. Korean Auxiliary Verb Constructions Without VP Nodes. In *Proceedings of the 1993 Workshop on Korean Linguistics*, pp 274–286, Cambridge, Massachusetts: Harvard University, Department of Linguistics.

Copestake, A., Flickinger, D. P., Pollard, C. J. and Sag, I. A. 2005. Minimal Recursion Semantics: an Introduction. *Research on Language and Computation* 4(3), 281–332.

den Besten, H. 1985. Some Remarks on the Ergative Hypothesis. In W. Abraham (ed.), *Erklärende Syntax des Deutschen*, pp 53–74, Tübingen: Gunter Narr Verlag.

Flickinger, D. P. 2008. Transparent Heads. In Müller (2008).

Gerbl, N. 2007. *An Analysis of Pseudoclefts and Specificational Clauses in Head-driven Phrase Structure Grammar*. Ph. D.thesis, Georg-August-Universität Göttingen.

Ginzburg, J. and Sag, I. A. 2000. *Interrogative Investigations: the Form, Meaning, and Use of English Interrogatives*. Stanford, CA: CSLI Publications.

Haider, H. 1986. Fehlende Argumente: vom Passiv zu kohärenten Infinitiven. *Linguistische Berichte* 101, 3–33.

Hinrichs, E. W., Kathol, A. and Nakazawa, T. (eds.). 1998. *Complex Predicates in Nonderivational Syntax*. San Diego: Academic Press.

Hinrichs, E. W. and Nakazawa, T. 1994. Linearizing AUXs in German Verbal Complexes. In J. Nerbonne, K. Netter and C. J. Pollard (eds.), *German in Head-Driven Phrase Structure Grammar*, pp 11–38, Stanford, CA: CSLI Publications.

Kasper, R. T. 1995. Semantics of Recursive Modification. Talk at HPSG-Workshop in Tübingen 1995.

Kathol, A. 1998. Constituency and Linearization of Verbal Complexes. In Hinrichs et al. (1998), pp 221–270.

Kathol, A. 2000. *Linear Syntax*. New York, Oxford: Oxford University Press.
Kiss, T. 1992. Variable Subkategorisierung. *Linguistische Berichte* 140, 256–293.
Kiss, T. 1995. *Infinite Komplementation*. Tübingen: Niemeyer.
Levine, R. D. and Hukari, T. E. 2006. *The Unity of Unbounded Dependency Constructions*. Stanford University: CSLI Publications.
Meurers, W. D. 2000. Lexical Generalizations in the Syntax of German Non-Finite Constructions. Arbeitspapiere des SFB 340 No. 145, Eberhard-Karls-Universität, Tübingen.
Mikkelsen, L. 2005. *Copular Clauses: Specification, Predication and Equation*. Amsterdam: John Benjamins Publishing Co.
Mikkelsen, L. To appear. Copular Clauses. In K. von Heusinger, C. Maienborn and P. Portner (eds.), *Handbook of Natural Language Meaning*, Berlin: Mouton de Gruyter.
Müller, S. 1997. Yet another Paper about Partial Verb Phrase Fronting in German. Research Report RR-97-07, DFKI, Saarbrücken, also in *Proceedings of COLING 96*, pp 800–805.
Müller, S. 1999. *Deutsche Syntax deklarativ. Head-Driven Phrase Structure Grammar für das Deutsche*. Tübingen: Niemeyer.
Müller, S. 2002. *Complex Predicates: Verbal Complexes, Resultative Constructions, and Particle Verbs in German*. Stanford, CA: CSLI Publications.
Müller, S. 2004. Continuous or Discontinuous Constituents? *Research on Language and Computation* 2(2), 209–257.
Müller, S. 2007. *Head-Driven Phrase Structure Grammar: Eine Einführung*. Tübingen: Stauffenburg Verlag, first edition.
Müller, S. (ed.). 2008. *Proceedings of the 15th International Conference on Head-Driven Phrase Structure Grammar, NICT, Keihanna, Japan*, Stanford, CA, CSLI Publications.
Müller, S. In Preparation. Head-Driven Phrase Structure Grammar. In A. Alexiadou and T. Kiss (eds.), *Syntax – Ein internationales Handbuch zeitgenössischer Forschung*, Berlin: Walter de Gruyter Verlag, second edition.
Müller, S. To appear. Elliptical Constructions, Multiple Frontings, and Surface-Based Syntax. In G. Jäger, P. Monachesi, G. Penn and S. Wintner (eds.), *Proceedings of Formal Grammar 2004, Nancy*, Stanford, CA: CSLI Publications.
Partee, B. H. 1987. Noun Phrase Interpretation and Type-Shifting Principles. In J. A. G. Groenendijk, D. de Jongh and M. J. B. Stokhof (eds.), *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*, pp 115–143, Dordrecht: Foris.
Paul, H. 1919. *Deutsche Grammatik. Teil IV: Syntax*. Halle an der Saale: Niemeyer.
Pollard, C. J. 1996. On Head Non-Movement. In H. Bunt and A. van Horck (eds.), *Discontinuous Constituency*, pp 279–305, Berlin, New York: Mouton, published version of a Ms. dated January 1990.
Pollard, C. J. and Sag, I. A. 1987. *Information-Based Syntax and Semantics*. Stanford, CA: CSLI Publications.
Pollard, C. J. and Sag, I. A. 1994. *Head-Driven Phrase Structure Grammar*. Chicago, London: University of Chicago Press.
Rentier, G. 1994. Dutch Cross Serial Dependencies in HPSG. In *Proceedings of COLING 94*, pp 818–822, Kyoto, Japan.
Rothstein, S. D. 1995. Small Clauses and Copula Constructions. In A. Cardinaletti and M. T. Guasti (eds.), *Small Clauses*, pp 27–48, San Diego, CA: Academic Press.
Sag, I. A., Wasow, T. and Bender, E. M. 2003. *Syntactic Theory: A Formal Introduction*. Stanford, CA: CSLI Publications, second edition.
van Eynde, F. 2008. Predicate Complements. In Müller (2008), pp 253–273.
Wegener, H. 1985. *Der Dativ im heutigen Deutsch*. Tübingen: Gunter Narr Verlag.

Serial Verb Constructions in Chinese: An HPSG Account

Stefan Müller and Janna Lipenkova

Freie Universität Berlin

Proceedings of the HPSG09 Conference

Georg-August-Universität Göttingen, Germany

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

This paper gives an account of Serial Verb Constructions (SVCs) in Mandarin Chinese. After a typological presentation of the phenomenon, we give an overview of the Chinese data. The inventory of SVC types is classified according to causal and temporal relations between the components. We discuss pragmatic conditions on the use of SVCs and alternative, semantically equivalent constructions. An HPSG analysis is proposed for marked SVCs which uses the interaction between aspect marking and the set of possible subordinative relations to deduce the extra-lexical meaning of the construction. Particular attention is paid to the syntactically peculiar SVC with shared internal arguments, which is accounted for by a non-cancellation approach to valence requirements.

1 Introduction

This paper proposes an account of Serial Verb Constructions with special focus on Chinese. The Serial Verb Construction is a complex predicate structure formed by two or more verbal phrases which select for the same subject. There is no syntactic marking available for the specification of the relation between the verbs. Semantically, a specific relation holds between the described events:

(1) a. Sranan: mi teki a nefi koti a brede
I take the knife cut the bread
'I cut the bread with a knife.'

b. Saramaccan: Kofi bay soni da di mujee
Kofi buy something give the woman
'Kofi bought something for the woman.'

The SVC has a complex event meaning, which is composed of the meanings of the single VP components and the extra-lexical causal relation between the sub-events.

SVCs are a typical example for syntactic underspecification in Chinese which results from the surface indeterminacy of the language. Thus, Chinese shows a high degree of context-sensitivity, which necessitates the systematic involvement of world and context knowledge for interpretation.

We present the Chinese data after a cross-linguistic consideration of general characteristics and types of SVCs in Section 4; we will see that, compared to other languages with strongly lexicalized and less productive SVCs, Chinese imposes weaker restrictions on the semantic properties of SVCs which are discussed in Section 3.2. The meaning of SVCs in Chinese is determined by semantic compositionality on the one hand and extra-lexical meaning components on the other

[†]We want to thank the group for German grammar at the FU Berlin, Andreas Guder, and Wang Lulu for comments and discussions. We thank Philippa Cook for proof reading.

hand. Together with the syntactic underspecification of the relation between the VP constituents, this represents the basic problem for their interpretation: Chinese SVCs are ambiguous with respect to the causal semantic relation between their VP components. This relation can be deduced on the basis of four interacting devices: on the level of surface structure, aspect markers can be used to mark a temporal relation between the events, which allows for the deduction of a subordinative relation manifesting the relevant temporal structure. On the other hand, the ordering of the VPs also indicates the relationship between the subevents. Semantically, combinations of specific, SVC-typical verbs may impose a fixed interpretation of the construction. Finally, context and world knowledge are often necessary for a correct understanding of the SVC; thus, SVCs for which an interpretation cannot be derived on the basis of syntactic and semantic constraints are apparently only used in situations in which the speaker assumes the receiver to be able to interpret the SVC correctly based on world and context knowledge.

The HPSG analysis proposed in Section 5 treats the SVC as syntactic coordination. The additional causal relation between the constituents is added on the mother node with the C-CONT (constructional content) feature. It is deduced based on semantic constraints on the aspect marking constellations for possible SVC types. A separate constraint is posited for the SVC with shared internal arguments. As it is assumed that a semantic role cannot be assigned twice to different arguments, we propose the projection of already satisfied selectional requirements up to the mother node. Thus, verbs with syntactically unrealized arguments can access already satisfied complements at phrase level.

2 Typological situation and cross-linguistic studies of SVCs

2.1 Typological situation

SVCs are found in four groups of languages distributed in geographically delimitable areas: West Africa, Central America, South-East Asia, and Oceania. These languages manifest structural similarities: SVCs are mostly used in SVO languages, although a few VSO and SOV languages (Ijo, Kwa, Ravüa) also allow for serialization ((Kroeger, 2004, p. 237), Seuren (1990)). On the other hand, serializing languages show deficient systems for the expression of semantic relations. They often manifest poor inflectional and prepositional components, which might represent an argument for the motivation of SVCs. An explanation for this correlation is proposed by Schiller (1990), who states the Semantic Case Instantiation Principle claiming that a language uses the most concrete mechanism available to express semantic relations. He posits the following preference hierarchy:

(2) Inflectional marking → Prepositional phrases → Serial verb constructions

Following this line, the existence of SVCs is explained by the incomplete systems of semantic specification in certain types of languages. These restrictions in semantic expressiveness are typical for creole and pidgin languages; besides, they also appear in isolating languages like Chinese, which, according to Tai (1989), exhibits a number of grammatical properties of child language, but also of creoles and pidgins. Thus, SVCs are semantically underspecified and context-dependent constructions which seem to occur as provisional grammatical structures in languages evolving towards more elaborated states. They are often subject to grammaticalization and lexicalization processes and develop into prepositional or coverbial expressions and lexical compounds.

Cross-linguistically, SVCs can have different formal and functional instantiations. Syntactically, we distinguish between two basic forms of SVCs: on the one hand, the SVC can be constructed out of two canonical verbal phrases directly adjoined to each other, as is the case in the examples in (1). On the other hand, in some languages, the different VPs are reordered: the SVC consists of two clusters, one containing the verbs and the other containing the objects of these verbs (Kroeger, 2004, p. 239-240). This is illustrated in the following examples:

(3) a. Jeh: Mi ruat doh au phei.
you buy give me rice
'You buy rice for me.'

b. Barai: Fu burede ije sime abe ufu.
he bread the knife take cut
'He cut the bread with the knife.'

Semantically, SVCs manifest different degrees of productivity, which is mainly due to restrictions on verbal combinations which can be conceptualized as single events. A number of prototypical functions can be discerned. According to Seuren (1990), the following meanings are often instantiated by verbal constituents of SVCs:

- Instrumental ('take')
- Dative or benefactive ('give')
- Comparative ('surpass')
- Reported speech ('say')
- Aktionsart: termination of an event ('finish')
- Directional adjunct ('go'/'come')

2.2 Survey of the literature on SVCs

The SVC has been extensively discussed in the literature on African and Chinese linguistics. For African languages, early accounts have been proposed by Stahlke (1970), Schachter (1974), Sebba (1987), and Baker (1989). Their analyses and definitions were subsequently used as a basis for analyses of Chinese SVCs. However, analyses of African SVCs can only in part be projected onto Chinese data, as Chinese SVCs are differently motivated and also manifest a number of peculiar characteristics not found in African languages. In Chinese linguistics, the serial verb construction was first discussed in Li and Thompson, 1981. It should be noted that earlier grammars also include examples of SVCs which are, indeed, subsumed under other more canonical grammatical structures such as coordination or complementation. Initially, some difficulties arose with respect to the delimitation of the relevant constructions: in their account of SVCs, Li and Thompson (1981) consider all predicates containing more than one verb. Thus, focussing on the surface form of the constructions, they also include control verb structures, clausal subjects and objects as well as descriptive clauses. These problems left aside, most subsequent analyses (Dai, 1990; Chang, 1990; LIU, 1997) concentrated on the syntactic properties of SVCs. This again led to incomplete descriptions: the semantic composition and, particularly, the ambiguity of SVCs, which we take as basic characteristics distinguishing canonical SVCs from verbal coordination, were often disregarded. Thus, the status of the SVC as an autonomous construction was challenged by authors who attempted to subsume it under other syntactically similar structures (coordination in Wippermann, 1993, complementation in Paul, 2005; Seuren, 1990). This tendency is also manifested in African linguistics: Bodomo (1993) states that SVCs are usually categorized either as coordinative structures with suppressed conjunctions, or as subordinative constructions containing embedded clause complements with suppressed complementizers.

In the following, we will attempt to make a short synthesis of the SVC definitions proposed for Chinese. We will also refer to the extensive literature on African SVCs, hoping to provide a set of characteristics that delimits accurately a type of construction that can be well-handled in a constraint-based analysis. However, we will also see that SVCs are related to pragmatic, cultural and conceptual restrictions that cannot be completely captured in a formal account.

3 Overview of the Chinese data

3.1 Syntax

The Chinese SVC is composed of two verbal phrases. They follow each other without an overt syntactic marking of the semantic relation between the described events:

(4) Ta1 qi3 chuang2 chuan1 yi1fu4.
he get.up bed dress clothes
'He gets up and puts on his clothes.'

Whereas the conjunction *and* is used in English to mark a simple coordination or temporal succession between the VPs, Chinese simply adjoins the two VP constituents. The relation has than to be inferred from speech context, conceptual knowledge, and constructional meaning.

The VPs in an SVC share their subject. It is realized only once in sentence-initial position and understood to be the subject of the second VP.

Additionally, the verbs may also share their direct object:

(5) Ta1 zhong3 cai4 mai4.
he plant vegetables sell
'He plants vegetables to sell them.'

In this example, *cai4* is the object both of *zhong3* and of *mai4*. It is only realized in the first VP. In this type of SVC, a relation of purpose holds between the two events. LIU (1997) proposes an explanation for this structure in terms of Ross' directionality constraint (1967): deletion is directed forward if the identical elements are left-branching, but backward if they are right-branching.

3.2 Semantics

The SVC is used to describe a single overall event, which is composed of two subevents. This general description of the semantic composition of SVCs bears some degree of arbitrarity, as the possible conceptual combinations of events are often conditioned by cultural as well as individual perceptions of the world:

> [...] in order for SVCs to be grammatical, it must be possible for speakers of the language to interpret the various actions as comprising a single coherent event. It appears that different languages impose different restrictions as to which specific combinations of verbs are permissible, and that these restrictions are sometimes due to cultural factors. (Kroeger, 2004, p. 234)

SVCs are often translated by single mono-verbal clauses in non-serializing languages. As is pointed out in Durie, 1997, p. 321, the codification of a situation by a separate verb indicates that this situation is perceived as a salient event type: "the verbal system of a language evolves as a categorization of the event-types that are [...] communicatively in demand for the speech community." In serializing languages with poor verbal systems, SVCs are used as a means to enrich the inventory of possible event-types by verbal series with recurring components. The SVCs in these languages show a strong tendency towards lexicalization: on the one hand, single verbs often develop distinct meanings when they are used in SVCs. On the other hand, verbal combinations often take semantically unanalyzable meanings.

In light of this close relation between SVC verbs in other serializing languages, the constituents of Chinese SVCs manifest a certain autonomy in that each of the VPs can occur on its own as an independent predicate (with limitations for the shared-object SVC, in which the object has to be overtly realized if the second VP is used independently). In this case, the isolated "subevent" can be perceived as a conceptual whole. However, the meaning of the SVC is not merely a combination of the two VP meanings. As a specific, but underspecified semantic relation holds between the two subevents, additional content is created at the level of the mother node. Therefore, a switch of the VP positions changes the meaning of the construction. This contrasts with instances of VP coordination, where an unspecified temporal relation holds between the events, allowing for the inversion of the constituents without significant change of the meaning:

(6) a. Ta1 xie3 xin4 hui4 ke4.
he write letter receive guest
'He writes letters and receives guests.'

b. Ta1 hui4 ke4 xie3 xin4.
he receive guest write letter
'He receives guests and writes letters.'

An unmarked SVC does not specify the relation between the two events. Thus, multiple interpretations are possible. The correct reading is to be inferred under consideration of world and context knowledge and the lexical semantics of the verbs. Figure 1 shows the possible relations between the subevents of an SVC.

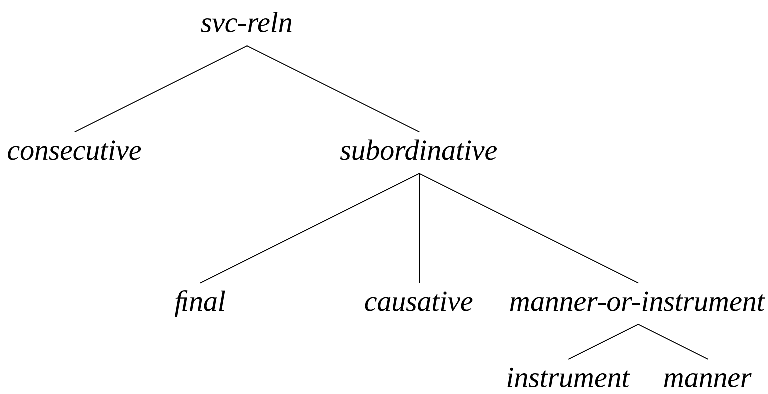


Figure 1: Possible relations between events expressed by SVCs

3.3 Shared object SVCs

In this section, we describe in more detail the specific syntactic and semantic properties of the SVC with shared direct object. SVCs of this type are formed out of two transitive verbs. However, only the first verb takes an overtly realized direct object. The unrealized object of the second verb is understood to be coreferenced with the object of the first verb. The shared-object SVC involves no semantic ambiguity: it only allows for a final reading and thus also has the semantic constraints imposed on canonical final SVCs. However, shared-object SVCs are limited in productivity, as they impose further lexical constraints on the possible combinations of verbs. These restrictions are discussed in Section 4.2.

Liu (2009) argues that the described constellations with shared objects are not instances of SVCs. He motivates this by the different properties of the constructions with respect to perfective aspect marking: in an SVC with two complete VPs, both verbs can be marked by the perfective aspect marker *le*, whereas only the first VP can be marked in the shared-object SVC. This argument results from a different understanding of SVCs; in fact, both VPs in canonical SVCs can take *le* without challenging the syntactic acceptability of the construction. However, the notion of SVC adopted in our paper relies on the semantic relations between subevents. This relation in turn interacts with aspectual properties: the distribution of aspect markers is restricted for subtypes on semantic grounds. For the final SVC – whether canonical or shared-object – we assume that the second VP cannot be marked by *le*, as it is an irrealis clause.

4 Extra-lexical meaning components in SVCs

The challenges posited by SVCs are to a great part semantic in nature. On the one hand, we have to deal with the non-compositionality and underspecification of meaning and the resulting ambiguities. On the other hand, we will see that SVCs show different degrees of specificity of meaning and, therefore, of productivity: possible SVCs go from fully productive structures with free lexical instantiations to collocational expressions reflecting grammaticalization and lexicalization tendencies. In a typological perspective, SVCs show systematic restrictions on possible meaning combinations, which have to be integrated into the analysis in addition to syntactic constraints on the form and argument structure of the VPs. Finally, SVCs show interesting effects of interaction between the argument structures of the constituent verbs, which also contribute a part of their non-compositional meaning.

4.1 Surface ambiguity and disambiguation of the SVC

We have seen that SVCs come with a set of possible semantic relations between the subevents. They are not marked on the surface and thus are determined at phrase level. The semantic ambiguity of an unmarked SVC results from an underspecification, as the correct relation between the parts of the SVC is to be deduced from world and context knowledge and from lexical and iconic properties of the verbal combinations.

We hypothesize that three types of knowledge – with different degrees of specificity with respect to the speech situation – are involved in the interpretation of an utterance: 1) Linguistic knowledge (default: semantic compositionality), 2) World knowledge (presupposes concrete receiver), and 3) Context (presupposes concrete speech situation). The presumed availability of these knowledge components impacts on the choice of a construction with which the speaker intends to express a semantic relation. In line with the argument of Goldberg (1995, p. 68), who claims that two constructions cannot be both semantically and pragmatically equivalent, the following constructions are available to express the set of relations postulated for SVCs in different pragmatic settings:

- Lexical / syntactic meaning → complex clause with subordinate conjunction
- World knowledge → SVC with aspectual marking
- Context → unmarked SVC

We see a decrease in "heaviness" of the constructions: the more information available, the less complex and elaborate the syntactic structure. It is assumed that the speaker chooses the most economic form of expression allowing for a correct interpretation.

In the case of the complex clause, the meaning can be deduced compositionally: it is contributed by the meanings of the lexical items and their syntactic combination. The subordinate relation is unambiguously specified by an overt conjunction. For the use of SVCs, we assume that speakers of the language have knowledge about the set of possible causal SVC-relations as part of their language capacity. If world knowledge is assumed on the side of the hearer which allows the perception of the described events as a conceptual whole, the SVC with aspect marking is used: as we will see in the next section, causal relations that hold in SVCs also contain a temporal component, which can be specified by aspect markers. The mapping of the aspect values onto the set of possible relations allows the deduction of the correct causal relation. Finally, if an SVC-relation is to be expressed which fits in a specific context known to the hearer, an unmarked, completely underspecified SVC will be used.

In the following paragraphs, we illustrate the semantic correspondences between unmarked SVCs, marked SVCs and complex clauses. It will be shown that these constructions differ on the level of pragmatics: their use is conditioned by presuppositions of the speaker about the presence or absence of world and context knowledge on the side of the receiver.

4.1.1 Aspect marking in SVCs

The relation between the two events can be disambiguated by use of the particles *le* (perfective) and *zhe* (durative). These particles are commonly claimed to be aspect markers. However, they can also act as markers of temporal reference: although Chinese does not have a grammaticalized tense component, aspect markers in complex clauses are interpreted as markers of temporal relations between the events.

In SVCs, aspect markers perform a pragmatic function similar to subordinative conjunctions. Their temporal reference function can be related to the semantic relations in SVCs in the following way: subordinative relations are complex relations in the sense that they also contain a temporal component. They expose the following correspondences:

- Final → succession
- Causative → underspecified relation (succession or simultaneity)
- Manner, instrument → simultaneity

Thus, by mapping the temporal function of aspect markers onto the set of possible subordinative relations, we get the following interpretations for SVCs:

- VP1[perf] VP2 → VP1 *in order to* VP2

(7) Ta1 qu3 le qian2 qu4 guang1jie1.
he withdraw PERF.ASP money go shopping
'He withdrew money to go shopping.'

- VP1 VP2[perf] → VP2 *because of* VP1

(8) Ta1 zhu4 Zhong1guo2 xue2 le Han4yu3.
he live China learn PERF.ASP Chinese
'He acquired Chinese because he lived in China. '

- VP1[dur] VP2 → VP2 *by means of* VP1

(9) Ta1 na2 zhe kuai4zi chi1 fan4.
he take DUR.ASP chopsticks eat meal
'He eats with chopsticks.'

4.1.2 Interrelations of SVCs with complex clauses

The causal relations in SVCs can also be expressed by complex clauses with subordinate conjunctions (e. g. *yin1wei4* ('because of'), *wei4le* ('in order to'), *yi3hou4* ('after')). The following examples demonstrate such semantic equivalences:

(10) a. Ta1 xie3 zi4 ai4 ma3.
he write characters suffer critics
'He wrote characters and suffered critics.' or
'He suffered critics for writing characters.'

b. Ta1 yin1wei4 xie3 zi4 ai4 ma3.
he because write characters suffer critics
'He suffered critics for writing characters.'

(11) a. Ta1 qu3 qian2 qu4 guang1 jie1.
he withdraw money go shopping
'He withdraws money to go shopping.'

b. Ta1 wei4le qu4 guang1 jie1 qu3 qian2.
he in order to go shopping withdraw money
'He withdraws money to go shopping.'

4.1.3 Ordering of the VPs

The ordering of the VPs in an SVC also makes a contribution to its extra-lexical meaning: the subevents are sequenced according to the order of occurrence in the real world (Temporal Sequence Principle, Tai, 1988) as well as to their direction of causation (Durie, 1997, p. 330). Both criteria apply for SVCs with a consecutive ordering of the events: in final SVCs, the purpose VP follows the action VP. In causative SVCs, the cause VP precedes the effect VP. Instrument SVCs, which bear a temporal relation of simultaneity, are interpreted according to causal priority between the events: the use of an instrument is prior to the effect which is achieved with it; thus, the instrument VP precedes the main event VP.

4.2 Specificity of meaning and productivity in SVCs

In this section, we will show that SVCs show different degrees of specificity of meaning, which are interrelated with restrictions in productivity of the possible lexical constellations: a range of SVCs can only be formed with verbs from restricted classes. These restrictions, in turn, interact with the choice of a "preferred" construction by the speaker described in the previous section: the hierarchy of constructions applies fully only in the case of freely productive SVCs (causative / final SVCs with unshared objects). We find two basic kinds of SVC productivity in Chinese: first, SVCs can manifest combinations of verbs of semantic classes which seem to be representative for the causal relations included in the event structure of SVCs. Such combinations are found in final SVCs with shared objects as well as in causative SVCs. On the other hand, SVCs may include one verb that is frequently used in series. This kind of serialisation is also found in a number of other serialising languages (e. g. Sranan, Sebba, 1987). It is used to describe event-types with “identifiable recurrent subcomponents” (Durie, 1997). We find this type of serialisation in Chinese manner, instrument and deictic-final SVCs.

In shared-object SVCs, both verb positions are restricted: the V1 is obligatorily volitional and denotes the creation or acquisition of its object; thus, two semantic classes are available for V1: Verbs of creation (ex. *chuang4zuo4* (‘create’), *chao3* (‘cook’), *zhong3* (‘plant’)) and Verbs of acquisition (ex. *mai3* (‘buy’), *zhao3* (‘find’)). These verbs can also occur in the ditransitive *gei3*-construction with a benefactory argument. Assuming that a benefactory role is inherently contained in their lexical semantics, the agent of the shared-object SVC can be understood as an implicit beneficient.

The V2 expresses how the object is to be disposed of after the action of V1. The disposal meaning is also relevant for other syntactic constructions in Chinese; thus, the *ba*-construction, which licenses preposed objects, is only grammatical with verbs containing a disposal component.

The overall meaning of the shared-object SVC can be illustrated as follows:

(12) SUBJ V1 OBJ V2
agent creates/gains possession over theme/patient$_i$ in order to dispose of *i*

The following set of examples shows possible instantiations of this semantic constraint:

(13) a. Ta1 chao3 yi4 pan1 cai4 chi1.
he cook one CL dish eat
‘He cooked a dish to eat it.’

b. Ta1 zhong3 cai4 mai4.
he plant vegetable sell
‘He plants vegetables to sell them.’

c. Ta1 chuang4zao4 yue4qu3 yan2chu1.
he create music.work perform
‘He writes musical works to perform them.’

In causative SVCs, the first verb is obligatorily volitional, whereas the second verb is mostly unaccusative; the second VP can also take a passive form with the particle *bei4* (14b):

(14) a. Ta1 zuo4 zai4 di4shang4 gan3mao4 le.
he sit on floor get.cold PERF.ASP
‘He caught a cold because he was sitting on the floor.’

b. Ta1 tou1 che1 bei4 jing3cha2 zhua1 le.
he steal car BEI police arrest PERF.ASP
‘He was arrested by the police for stealing a car.’

In manner SVCs, the verb in the first VP is restricted to verbs which are canonically used to express means or manner; these are: Verbs of position as in (15), verbs of motion as in (16), and *zuo* (‘sit’), which takes as object a transport medium and expresses the means by which one gets to a location (17). In the latter case, V2 is also restricted to the two verbs *qu* (‘go’) and *lai* (‘come’), which attributes a collocational character to the SVC.

(15) Ta1men zhan4 zai4 men2kou3 liao2tian1.
they stand at door chat
‘They chat standing by the door.’

(16) Ta1 qi2 zhe zi4xing2che1 da3 dian4hua4.
he ride PERF.ASP bike call phone
‘He phones cycling on his bike.’

(17) Ta1 zuo4 huo3che1 qu4 Bei3jing1.
he sit train go Pekin
‘He goes to Beijing by train.’

Another kind of SVC with collocational meaning is the final SVC in which the first VP describes the movement towards a location at which the action of the second VP is to be performed. The position of the first verb is restricted to a small class of verbs which can also act as directional complements:

(18) a. Ta1 lai2 Mo2si1ke1 xue2 E2yu3.
he come Moscow learn Russian
‘He comes to Moscow in order to learn Russian.’

b. Ta1 shang4 lou2 shui4jiao4.
he go.up house sleep
‘He goes upstairs to sleep.’

In this case, the meaning of the construction is:

(19) SUBJ V1 OBJ VP2
agent goes to/comes to goal$_i$ in order to perform some action at *i*

The object of V1 is assigned two thematic roles: it is the goal of V1 and the location of the event described by the second VP.

Finally, the instrument SVC can be formed only with the two verbs *na* (‘take’) and *yong* (‘use’). In these cases, the object of the first verb is understood to be the instrument argument of the second verb.

In this section, we have seen various ‘prototypical’ constellations of SVCs which impact on the constructional meaning and show that the meaning of SVCs in Chinese cannot be deduced lexically. Further evidence for the SVC as an autonomous construction is provided by languages in which SVCs bear semantically unanalyzable, strongly lexicalized meanings. We have also shown that the additional content of SVCs is often conditioned by overlapping argument structures, in that a sole argument gets assigned semantic roles from different verbs. The argument structure properties of SVCs are discussed in the following section.

4.3 Issues of argument structure in SVCs

The SVC shows two distinctive argument structure properties: on the one hand, it disallows the attribution of the same semantic role to different arguments. On the other hand, the same argument can receive multiple semantic roles from different verbs.

Durie (1997) points out that SVCs cannot contain duplicate semantic roles: a role cannot be attributed to two different arguments. He illustrates this with examples from White Hmong and Kalam, where two transitive verbs can only take distinct objects if one of these objects is an oblique argument. This property also applies for other verbal constructions, starting with simple clauses with single verbs. It justifies the overall event reading of the SVC as we assume that the same event does not allow for two distinct participants to be attributed the same semantic role. Thus, coinciding semantic role assignments of verbs must be realized on the same argument. In the following pair of examples, (20a) is an instance of coordination where the two verbs each have an independent theme argument; (20b) is an SVC, as both verbs attribute their theme role to the argument *cai*:

(20) a. Ta1 zhong3 cai4 mai4 shui3guo3.
he plant vegetable sell fruit
'He plants vegetables and sells fruits.'

b. Ta1 zhong3 cai4 mai4.
he plant vegetable sell
'He plants vegetables to sell them.'

To account for the assignment of multiple semantic roles to the same arguments, Durie (1997) proposes an approach with two levels of argument structure: alongside the independent argument structures of the single verbs, a "fused" argument structure is imposed for the whole construction. Durie points out that this additional level is necessary for the realization of the prohibition against the duplication of semantic roles, as it is illustrated by the following example:

(21) Ta1 na2 bi3 xie3 zi4.
he take pen write character
'He writes characters with a pen.'

On the level of lexical semantics, the verbs *na2* and *xie3* both assign a theme role to their direct object. However, the "fused" argument structure can be represented as [Agent, Instrument, Theme], whereby the noun *bi3* is assigned the instrument role instead of the theme role. Thus, the constraint against duplicate role assignment is satisfied at the level of the constructional argument structure. This level is also involved in the correct interpretation of argument roles, which can often only be deduced in the context of the whole SVC: we have seen that *na2* in the above example does not take an instrument argument when used independently. However, in the SVC context, it is used to mark an instrument.

5 HPSG analysis of Chinese SVCs

In this section, we describe an HPSG-analysis of Chinese SVCs. We first posit a general syntactic constraint that holds for all SVCs. In a second step, we deal with constraints on binary SVCs (unshared-object SVCs and shared-object SVCs) in more detail. The consecutive SVCs will not be dealt with in this paper. We propose complex implicational constraints relating the aspect marking constellations of SVCs to the semantic relations that were introduced in Section 3.2. Finally, we show how valence requirements in shared-object SVCs can be satisfied non-locally by projection to the constructional level.

5.1 General constraint for SVCs

We assume that all SVCs are instances of one of three types: *consecutive-svc*, *unshared-obj-svc*, and *shared-obj-svc*. These types are subtypes of the type *svc*. Structures of type *svc* have to obey the following constraint:

(22) $$\textit{svc} \rightarrow \begin{bmatrix} \text{SYNSEM}\,|\,\text{LOC}\,|\,\text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{SPR} & \langle \boxed{1}\ \text{NP} \rangle \end{bmatrix} \\ \text{C-CONT} & \begin{bmatrix} \text{IND} & \boxed{2} \\ \text{RELS} & \left\langle \begin{bmatrix} \textit{svc-reln} & \\ \text{ARG0} & \boxed{2} \\ \text{ARG1} & \boxed{3} \\ \text{ARG2} & \boxed{4} \end{bmatrix} \right\rangle \end{bmatrix} \\ \text{NH-DTRS} & \left\langle \begin{bmatrix} \text{SS}|\text{LOC}|\text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{SPR} & \langle \boxed{1} \rangle \\ \text{SUBCAT} & \textit{los} \end{bmatrix} \\ \text{CONT}\,|\,\text{IND} & \boxed{3} \end{bmatrix}, \begin{bmatrix} \text{SS}|\text{LOC}|\text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{SPR} & \langle \boxed{1} \rangle \end{bmatrix} \\ \text{CONT}\,|\,\text{IND} & \boxed{4} \end{bmatrix} \right\rangle \end{bmatrix}$$

We represent the SVC as a non-headed structure with two verbal daughters, whereby the first verbal daughter is always a complete VP. We assume a non-cancellation approach to valence. This approach was introduced by Meurers (1999) and Przepiórkowski (1999) for the analysis of case and fronting in German. It has subsequently been used by Müller (2008) for depictives in German and English, as well as by Bender (2008) for the explanation of constituent order in Wambaya. The gist of this proposal is that valents are still members of the SUBCAT list even if the respective argument has been combined with the head already. Whether this combination has taken place or not is registered by a binary feature REALIZED whose value is '+' if an argument is combined with its head and '–' if no such combination has taken place. A fully saturated head has a SUBCAT list that has only elements with the REALIZED value '+'. Meurers called such elements *spirits*. So, the value of the SUBCAT list in the first non-head daughter in (22) is *list-of-spirits* (*los*). This list contains the values of the arguments already realized in the VP. The two verbal daughters subcategorize for the same subject. Therefore, their SPR values are identified and projected to the mother node. The semantic relation between the VPs is contributed at the level of the mother node: we use the feature C-CONT (constructional content) proposed in Copestake, Flickinger, Pollard and Sag, 2005 to accommodate semantic relations contributed at construction level. The constraint above only says that there will be a relation between the two events expressed by the VPs. The relation is a subtype of *svc-reln* (see Section 3.2).

5.2 Analysis of SVCs with unshared objects

SVCs with unshared objects require that the arguments of the verb in the second VP are all realized, that is: the elements in the SUBCAT list of the second VP have to be spirits. This is what is formalized as the following constraint:

(23) $\textit{unshared-object-svc} \rightarrow \left[\text{NH-DTRS}\ \left\langle [\ ], \left[\text{SS}|\text{LOC}|\text{CAT}|\text{SUBCAT}\ \textit{los}\right]\right\rangle\right]$

The semantic interpretation of the construction depends on the aspect marking of the VPs. If the second VP is perfective, the relation between the two events is causative. We assume that the perfective aspect is analyzed as a lexical rule that combines a verb with the aspect marker *le* and contributes a *perfective*' relation to the beginning of the RELS list. Hence, the unshared object SVC can refer to this relation: if it is present in the RELS list of the second VP the relation that is contributed by the construction has to be *causative*':

(24) $\begin{bmatrix}\textit{unshared-object-svc}\\ \text{NH-DTRS}\ \left\langle [\ldots], \left[\text{RELS}\ \langle \textit{perfective}\rangle \oplus \textit{list}\right]\right\rangle\end{bmatrix} \rightarrow \left[\text{C-CONT}|\text{RELS}\ \langle \textit{causative}\rangle\right]$

If the first VP is perfective, the relation between the two events is final:

(25) $\begin{bmatrix}\textit{unshared-object-svc}\\ \text{NH-DTRS}\left\langle \left[\text{RELS}\ \langle \textit{perfective}\rangle \oplus \textit{list}\right]\right\rangle \oplus \textit{list}\end{bmatrix} \rightarrow \left[\text{C-CONT}\ |\text{RELS}\ \langle \textit{final}\rangle\right]$

Note that this analysis predicts that not both VPs can be marked for (perfective) aspect simultaneously, since if they were, conflicting constraints would be imposed on the constructional contribution of the SVC (*final* and *causative* are incompatible with each other, see Figure 1).

We assume that the relations that are contributed by linguistic objects are not represented inside of CONT, but at the outermost level of the sign. Since heads select only *synsem* objects and not complete signs, this makes it impossible for a head to select the semantic relations contributed by its dependents and hence results in a more local theory of selection. See also Sailer, 2004 on the locality of selection with regard to semantic information. However, the semantic contribution of daughters can be accessed on the constructional level as is demonstrated in the constraint in (25).

The durative marker *zhe* can only be used in the first VP.[1] It marks either a manner or an instrument relation between the two events:

(26) Ta1 chang4 zhe ge1 qu4 xue2xiao4.
he sing DUR.ASP song go school
'He goes to school singing a song.'

(27) $\begin{bmatrix} \textit{unshared-object-svc} \\ \text{NH-DTRS} \ \left\langle \left[\text{RELS} \left\langle \textit{durative} \right\rangle \oplus \textit{list}\right] \right\rangle \oplus \textit{list} \end{bmatrix} \rightarrow \left[\text{C-CONT|RELS} \ \left\langle \textit{manner-or-instrument} \right\rangle\right]$

We have described SVCs with the two verbs *na2* ('hold') and *yong4* ('use') as structures with a collocational character: the object of the first VP is understood to be the instrument for the action described by the second VP. The *instrument* relation is a subtype of *manner-or-instrument* relation. Thus, an SVC whose VP1 contains the durative marker in combination with a verb that contributes either a *hold*' or *use*' relation is interpreted as an instrumental SVC:

(28) Ta1 na2 zhe bi3 xie3 zi4.
he hold DUR.ASP pen write characters
'He writes characters with a pen.'

(29) $\begin{bmatrix} \textit{unshared-object-svc} \\ \text{NH-DTRS} \ \left\langle \left[\text{RELS} \left\langle \textit{durative, hold-use-rel} \right\rangle \oplus \textit{list}\right] \right\rangle \oplus \textit{list} \end{bmatrix} \rightarrow \left[\text{C-CONT|RELS} \ \left\langle \textit{instrumental} \right\rangle\right]$

Having explained SVCs with unshared objects, we now turn to SVCs with shared objects.

5.3 Analysis of SVCs with shared objects

In the basic SVC case, each of the two verbs takes its own object. We therefore posited a straightforward subtype *unshared-object-sc* with two VP daughters whose valence requirements are realized locally. For the *shared-object-svc*, we assume a subtype with a complete VP as first daughter and a single verb as second daughter. In this case, the object of the second verb is identical to the object inside the preceding VP.

In order to explain the details of the analysis, we have to elaborate the sketch of the raising spirits analysis that was provided in the previous section: As was mentioned above, we adopt a complex structure for the elements on the SUBCAT-list. The *synsem* objects are represented as the values of the feature ARGUMENT and the status of the argument is represented via the boolean feature REALIZED. The value of REALIZED is '+' for arguments that are realized in a head argument structure and '–' for unrealized arguments.

[1] *Zhe* can mark two adjoined VPs. However, the resultant structure is VP coordination as no specific relation holds between the two events.

(30) $\begin{bmatrix}\textsc{argument} & \textit{synsem}\\ \textsc{realized} & \textit{boolean}\end{bmatrix}$

This treatment of valence ensures that the elements on the SUBCAT-list are not deleted after their realization. Instead, they are simply marked as realized and projected to the mother node. With this machinery in place, we posit the following constraint for the *shared-object-svc*:

(31) *shared-object-svc* →

$$\begin{bmatrix}\textsc{nh-dtrs} & \left\langle \begin{bmatrix}\textsc{subcat} & \left\langle \begin{bmatrix}\textsc{argument} & \boxed{1}\\ \textsc{realized} & +\end{bmatrix}\right\rangle \oplus \textit{list}\end{bmatrix}, \begin{bmatrix}\textsc{subcat} & \left\langle \begin{bmatrix}\textsc{argument} & \boxed{1}\\ \textsc{realized} & -\end{bmatrix}\right\rangle \oplus \textit{list}\end{bmatrix}\right\rangle\end{bmatrix}$$

The object of the first verb is overtly realized, whereas the object of the second verb is not. Its ARGUMENT value is identified with that of the object of the first daughter.

The constraint in (31) refers to the first elements in the respective SUBCAT lists, but nothing is said about the length of this list. This allows for instance ditransitive verbs as the second part of an SVC. (32) shows an example:

(32)
Ta1	mai3	yi1	ben3	shu1	song4	gei3	wo3.
he	buy	one	CL	book	offer	for/to	me

'He buys a book to offer it to me.'

In contrast to unshared object SVCs the semantic contribution of SVCs with a shared object is fixed. It is always the *final* relation. This is captured by the following constraint on *shared-object-svc*:

(33) *shared-object-svc* → $\begin{bmatrix}\textsc{c-cont}|\textsc{rels} \ \langle \textit{final} \rangle\end{bmatrix}$

We have pointed out in Section 4.2 that the semantics of SVCs is not only constrained with respect to possible relations between the described events; rather, the set of possible meanings for the subevents is also limited. We thus posit a hierarchy of relevant semantic verb classes (*creation-or-acquisition*, *disposal*, *volitional*, *go-or-come*, *hold-or-use* etc.) and constrain the KEY values of the verbs to subtypes of the corresponding relations. These lexical constraints also allow for predictions about the syntactic structure of SVCs: for example, by constraining the first verb of the *shared-object-svc* to verbs of creation and acquisition, we account for the fact that the construction cannot be formed with ditransitive verbs in VP1. On the other hand, it has been shown in Section 4.2 that the restrictions on possible verbs in SVCs correlate in interesting ways with other syntactic constructions such as the *ba3*-construction and the double-object structure with *gei3*.

The analysis of shared object SVCs presented here uses only machinery that was independently motivated. It therefore differs from the analysis of serial verbs in Ga that was suggested by Kropp Dakubu, Hellan and Beermann (2007). Serial verbs in Ga exhibit analogous argument sharing structures. The authors introduce the use of grammatical functions reminiscent of LFG and project information about arguments inside the feature QVAL. As grammatical functions are usually not assumed in HPSG work, we do not follow this approach but employ the non-cancellation technique that was independently motivated for the analysis of case assignment and partial verb phrase fronting and depictives.

Discussing the RELS feature in the previous section, we pointed out the conceptual advantage of having it at the outermost level of the feature structure rather than under SYNSEM. This feature geometry makes it impossible for a head to select via valence features the internal semantic contribution of a phrase (for instance the relation that is contributed by a verb inside VP). However, the non-cancellation account to valence makes available large parts of the syntactic structure at the mother node of a phrase. We would prefer to have a strictly local theory of selection, that is, a combination of strict locality in semantics as argued for by Sailer (2004) and of syntax as argued for by Sag (2007), but since the sharing of the object comes with a constructional semantic effect, the analysis should be related to a form meaning pair and the identification of the object referents should not be left to pronoun binding or similar devices. If this general approach is correct, we have evidence that information about VP internal objects has to be available at the VP level and hence that a non-cancellation approach to valence or an approach of the kind suggested by Kropp Dakubu et al. (2007) that projects information about the respective dependents is necessary for the analysis of languages like Mandarin Chinese and Ga.

6 Conclusion

In this paper, we provided a description and an analysis of SVCs in Chinese. After a general consideration of the SVC in a typological context and a description of its basic properties, we discussed the issues related to the syntactic underspecification and semantic ambiguity of SVCs. It has been shown that the interpretation of SVCs involves a number of meaning elements which are not contributed by the parts of the construction but rather by the whole configuration. We proposed an analysis of the Chinese SVC in HPSG, using two syntactic constraints for SVCs with unshared and shared objects, as well as complex implicational constraints for the representation of interactions between aspect markers and the subordinative relations in SVCs. The analysis has been implemented in the TRALE system (Meurers, Penn and Richter, 2002; Penn, 2004; Müller, 2007a) as part of a grammar fragment of Mandarin Chinese which uses a core grammar for German, Persian, Danish and Maltese. The respective grammars can be downloaded at http://hpsg.fu-berlin.de/Software/.

References

Baker, M. C. 1989. Object sharing and projection in serial verb constructions. *Linguistic Inquiry* 20, 513–553.

Bender, E. M. 2008. Radical Non-Configurationality without Shuffle Operators: An Analysis of Wambaya. In S. Müller (ed.), *Proceedings of the 15th International Conference on Head-Driven Phrase Structure Grammar, NICT, Keihanna, Japan*, pp 6–24, Stanford, CA: CSLI Publications.

Bodomo, A. B. 1993. Complex Predicates and Event Structure: An Integrated Analysis of Serial Verb Constructions in the Mabia Languages of West Africa. Working Papers in Linguistics 20, University of Trondheim, Norway.

Chang, C. H.-H. 1990. On Serial Verbs in Mandarin Chinese: VV Compounds and Coverbial Phrases. In Z. A. M. Joseph Brian D. (ed.), *When Verbs Collide: Papers from the Ohio State Mini-Conference on Serial Verbs*, pp 288–315, Ohio State University.

Copestake, A., Flickinger, D. P., Pollard, C. J. and Sag, I. A. 2005. Minimal Recursion Semantics: an Introduction. *Research on Language and Computation* 4(3), 281–332.

Dai, J. X.-D. 1990. Syntactic Constructions in Serial Verb Expressions in Chinese. In Z. A. M. Joseph Brian D. (ed.), *When Verbs Collide: Papers from the Ohio State Mini-Conference on Serial Verbs*, pp 316–339, Ohio State University.

Durie, M. 1997. Grammatical Structures in Verb Serialization. In P. S. Alex Alsina, Joan Bresnan (ed.), *Complex Predicates*, pp 289–354, Stanford: CSLI Publications.

Goldberg, A. E. 1995. *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago/London: University of Chicago Press.

Kroeger, P. R. 2004. *Analyzing Syntax: A Lexical-functional Approach*. Cambridge: Cambridge University Press.

Kropp Dakubu, M. E., Hellan, L. and Beermann, D. 2007. Verb Sequencing Constraints in Ga: Serial Verb Constructions and the Extended Verb Complex. In Müller (2007b), pp 99–119.

Li, C. N. and Thompson, S. A. 1981. *Mandarin Chinese. A Functional Reference Grammar*. Berkeley and Los Angeles: University of California Press.

Lipenkova, J. 2009. *Serienverbkonstruktionen im Chinesischen und ihre Analyse im Rahmen von HPSG*. Masters Thesis, Freie Univeristät Berlin.

LIU, G. 1997. *Eine unifikations-basierte Grammatik für das moderne Chinesisch – dargestellt in der HPSG*. Ph. D.thesis, University Constance, SFB 471, FG Sprachwissenschaft, Universität Konstanz, Germany.

Liu, H. 2009. Hanyu tongbin jiegou de jufa diwei [The syntactic status of the "shared object" construction in Chinese]. *Zhongguo yuwen [Chinese language]* 330.

Meurers, W. D. 1999. Raising Spirits (and Assigning Them Case). *Groninger Arbeiten zur Germanistischen Linguistik (GAGL)* 43, 173–226.

Meurers, W. D., Penn, G. and Richter, F. 2002. A Web-Based Instructional Platform for Constraint-Based Grammar Formalisms and Parsing. In D. Radev and C. Brew (eds.), *Effective Tools and Methodologies for Teaching NLP and CL*, pp 18–25, proceedings of the Workshop held at 40th Annual Meeting of the ACL. Philadelphia, PA.

Müller, S. 2007a. The Grammix CD Rom. A Software Collection for Developing Typed Feature Structure Grammars. In T. H. King and E. M. Bender (eds.), *Grammar Engineering across Frameworks 2007*, Stanford, CA: CSLI Publications.

Müller, S. (ed.). 2007b. *Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar, Stanford Department of Linguistics and CSLI's LinGO Lab*, Stanford, CA, CSLI Publications.

Müller, S. 2008. Depictive Secondary Predicates in German and English. In C. Schroeder, G. Hentschel and W. Boeder (eds.), *Secondary Predicates in Eastern European Languages and Beyond*, pp 255–273, Oldenburg: BIS-Verlag.

Paul, W. 2005. The "Serial Verb Construction" in Chinese: A Gordian Knot. *The Linguistic Review* 25, 367–411.

Penn, G. 2004. Balancing Clarity and Efficiency in Typed Feature Logic Through Delaying. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume*, pp 239–246, Barcelona, Spain.

Przepiórkowski, A. 1999. On Case Assignment and "Adjuncts as Complements". In G. Webelhuth, J.-P. Koenig and A. Kathol (eds.), *Lexical and Constructional Aspects of Linguistic Explanation*, pp 231–245, Stanford, CA: CSLI Publications.

Ross, J. R. 1967. *Constraints on Variables in Syntax*. Ph. D.thesis, MIT, reproduced by the Indiana University Linguistics Club.

Sag, I. A. 2007. Remarks on Locality. In Müller (2007b).

Sailer, M. 2004. Local Semantics in Head-Driven Phrase Structure Grammar. In O. Bonami and P. C. Hofherr (eds.), *Empirical Issues in Formal Syntax and Semantics*, pp 197–214, Online.

Schachter, P. 1974. A non-transformational account of serial verbs. *Studies in African Linguistics* Supplement 5, 153–271.

Schiller, E. 1990. On the definition and distribution fo serial verb constructions. In Z. A. M. Joseph Brian D. (ed.), *When Verbs Collide: Papers from the Ohio State Mini-Conference on Serial Verbs*, pp 34–64, Ohio State University.

Sebba, M. 1987. *The Syntax of Serial Verbs*. Amsterdam: John Benjamins.

Seuren, P. A. M. 1990. Serial Verb Constructions. In Z. A. M. Joseph Brian D. (ed.), *When Verbs Collide: Papers from the Ohio State Mini-Conference on Serial Verbs*, pp 14–32, Ohio State University.

Stahlke, H. 1970. Serial Verbs. *Studies in African Linguistics* 1, 60–99.

Tai, J. H.-Y. 1988. Temporal Sequence and Chinese Word Order. In J. Haiman (ed.), *Iconicity in Syntax*, pp 49–72, Amsterdam: John Benjamins.

Tai, J. H.-Y. 1989. Toward a Cognition-Based Functional Grammar of Chinese. In F. F. S. H. James H.-Y. Tai (ed.), *Functionalism and Chinese Grammar*, pp 187–226, Ohio State University: Chinese Language Teachers Association.

Wippermann, D. 1993. *Liandongshi: Der Begriff der Verbalserie in der chinesischen Linguistik*. Heidelberg: J. Groos.

Preposed Negation in Danish

Bjarne Ørsnes

Freie Universität Berlin

Proceedings of the HPSG09 Conference

Georg-August-Universität Göttingen, Germany

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

In Danish the base position of the negation and negated quantifier phrases is between the subject and the finite verb in embedded clauses. However, in embedded clauses introduced by a non-veridical complementizer such as *hvis* ('if') or *om* ('whether') the negation and negated quantifier phrases can also appear between the complementizer and the subject. This phenomenon is referred to as preposed negation. The paper investigates the structure and semantics of this construction. It is argued that preposed negation is no adjunction structure, but a special construction where the negation element is a sister of the complementizer and the filler of a filler-gap-structure. It is further argued that preposed negation is associated with negated verum-focus of a clause lacking an (aboutness-) TOPIC. The negation of a verum predicate explains why preposed negation fails to license strong negative polarity items and to rule out positive ones. The lack of a TOPIC explains why preposed negation is preferred with non-referential subjects and with weak readings of indefinite subjects and why preposed negation is incompatible with TOPIC-binding particles.The final section presents an HPSG-analysis of preposed negation using Minimal Recursion Semantics (MRS).

1 Introduction

In Danish non-V1/V2-clauses[1] sentential negation (and other sentential adverbs) appears between the subject and the finite verb thus marking the left-edge of the VP. Even non-subject negative quantifier phrases appear in the position of the sentential negation even though complements of the verb canonically follow the verbal head, cf. (1) and (2) below. I will refer to this as *ordinary negation*. Cf. the examples below.[2]

(1)	fordi	det	ny	system	ikke	tillader	ansøgere	under	15	år	(DK)
	because	the	new	system	not	allows	applicants	under	15	years	
	'because the new system does not allow applicants under 25 years'										

[†]I am especially indepted to Stefan Müller for numerous discussions and help with the analysis. Furthermore I wish to thank Jørg Asmussen, Philippa Cook, Felix Bildhauer, Jacob Maché, Line Mikkelsen, Patrizia Paggio, Roland Schäfer as well as the audience and reviewers of HPSG09 for discussion and comments. All remaining errors are my responsibility. This research is supported by the *Deutsche Forschungsgemeinschaft* under the grant nr. DFG (MU 2822/2-1).

[1]Here I use the term V1/V2-clauses for clauses where the finite verb precedes sentence adverbials, and the term non-V1/V2-clauses for clauses where the finite verb follows sentence adverbials. Here I will primarily be concerned with non-V1/V2-clauses as exemplified in (i).

(i)	fordi	Peter	ikke	synger
	Because	Peter	not	sings

[2](DK) marks an example from KorpusDK (http://ordnet.dk/korpusdk), (I) an example from the Internet. Other examples are constructed. The authentic examples have been abridged and sometimes slightly modified for reasons of space.

(2) hvis hun ingen erstatning fik, fordi motorcyklisten ikke
if she no compensation became, because motor.cyclist.DEF not
havde forsikret sig (DK)
had insured himself

'if she did not get any compensation, because the motor cyclist had no insurance'

However, in certain non-V1/V2-clauses there is a further possibility: sentential negation and non-subject negative quantifier phrases can also appear between the complementizer and the subject, as shown below. I will refer to this pattern as *preposed negation*.

(3) og hvis ikke kunsten magter at vise det, er det ikke kunst (DK)
and if not art.DEF is.capable.of to show this, is it not art

(4) hvis ingen arvinger der er, [. . .] (I)
if no heirs there are, [. . .]

Preposed negation is also observed in Norwegian and Swedish (Johannessen, 2000; Jensen, 2001), but with (slightly) different properties. In this paper, however, I will only discuss preposed negation in Danish.

Despite the extensive literature on negation preposed negation appears to have received little attention. It is often mentioned as a further possibility of negation-placement in Danish, but apart from the descriptive investigation in Skafte-Jensen (1995) it does not seem to have been subject to detailed study. The paper thus addresses two fundamental questions: what is the structure and what is the semantics of preposed negation.

In line with previous analyses of finite negation in English (Kim and Sag, 2002), I will suggest that the preposed element is a sister of the complementizer and that the preposed negation is the filler of a filler-gap dependency. I will further suggest that preposed negation is associated with special discourse semantic properties. Preposed negation is associated with negation of polarity focus ("verum"-focus) of a proposition lacking a topic. This account explains the peculiar behaviour of positive and negative polarity items with preposed negation. Though being sentential negation, preposed negation does not license strong negative polarity items and it licenses strong positive polarity items. Ordinary negation on the other hand licenses strong negative polarity items and rules out strong positive polarity items, when it is not associated with polarity focus of the clause. Thus while ordinary negation can be associated with both polarity focus and VP focus, preposed negation is only associated with polarity focus and may be seen as a structural means of signaling polarity focus. At the same time the subject of a clause with preposed negation obeys certain interpretative constraints: preposed negation is preferred with non-referential subjects and with weak readings of indefinite subjects. Furthermore topic-binding particles as investigated for German in Breindl (2008) are impossible with preposed negation. The constraints on the subject of a clause with preposed negation point to the conclusion that these clauses lack a topic, the subject being within the scope of the negation, i.e. the focal information (Ambridge and Goldberg, 2008). To account for the specific semantics of preposed negation and for the fact that only complementizers with a specific semantics and a specific phonological shape license preposing I will suggest that preposed negation is a construction, i.e. a specific pairing of syntax and semantics.

The paper is organized as follows. In Section 2 the basic properties of preposed negation are discussed. Negation will be shown to be part of a larger picture of preposing sentential adverbs and the construction will be shown to be subject to semantic as well phonological restrictions on the licensing complementizers. Section 3 deals with the structure of preposed negation. The construction is shown to be a syntactic structure and not a lexical structure or an adjunction structure as otherwise expected. In Section 4 the semantics and pragmatics of the construction are discussed. The construction is shown to be associated with negation of the polarity of a *topic-less* clause. Section 5 finally provides an analysis of the construction within the frame-work of HPSG using Minimal Recursion Semantics (MRS).

2 Preposed negation

2.1 Preposing in Non-Veridical Contexts

Preposing of the negation is only possible in embedded sentences containing a complementizer. It is most often observed in conditional clauses, but it is not restricted to conditional clauses. Preposing is possible with different kinds of non-veridical complementizers, i.e. operators that do not entail the truth of their proposition (Giannakidou, 1999; Skafte-Jensen, 1995).[3] Cf.

(5) jeg spekulerer på om ikke det er for sent
I wonder PREP whether not it is too late

(6) mon ikke det er for sent
MON not it is too late
'don't you think it is too late'

(7) bare ikke han kommer
BARE not he comes
'I hope he doesn't come'

In (5) preposing appears in an embedded polar question, in (6) in a deliberative question where the addressee is not supposed to know the answer to the question

[3]Skafte-Jensen (1995), however, gives (constructed) examples of preposing in temporal (veridical) clauses.

(Erteschik-Shir, 2009) and in (7) in an optative clause.[4] The complementizer *at* ('that') is especially telling, since it allows both a veridical (assertive) reading and a non-veridical (intentional) reading. Preposing is only possible in the latter reading.

(8) a. [. . .] og lagt albuen på pergamentet, at ikke vinden
[. . .] and placed elbow.DEF on pergament.DEF that not wind.DEF
skulle spille med det (I)
should play with it
'and placed the elbow on the pergament so that the wind should not play with it'

b. * [. . .] og sagde, at ikke barnet skulle lege med det
[. . .] and said, that not child.DEF should play with it
'and said that the child should not play with it'

Preposing is not restricted to negation or negative quantifier phrases either. It is also observed with a wide range of (polarity-) adverbs, even with adverb phrases where a preposed adverb is further modified by other adverbs (11) (cf. also Skafte-Jensen (1995)).

(9) hvis alligevel du deltager [. . .]
if anyway you participate [. . .]

(10) hvis godt du vil deltage [. . .]
if AFFIRM you want to participate [. . .]

(11) hvis [$_{\text{ADVP}}$ altså alligevel ikke] du deltager [. . .]
if that.is anyway not you participate [. . .]
'if you don't participate anyway, that is'

Since preposing is only possible in complementizer clauses, it is not observed in embedded constituent questions with the possible exception of *hvorfor* ('why'), where occasional examples of preposing are found, cf. (12).

(12) [. . .] hvori han ligefrem spørger hvorfor ikke Musikerne
[. . .] wherein he actually asks why not musicians.DEF
benytter andre Konsonanter end Octaven [. . .] (I)
use other consonants than octave.DEF

2.2 The Lexical Restriction on Preposed Negation

The fact that preposing occurs with many kinds of adverbs in all kinds of non-veridical contexts casts doubt on the claim that preposed negation is motivated by the close bond between conditional clauses and negation as claimed by Jespersen

[4] Note that *bare* ('I hope') and *mon* ('I wonder') may also occur as adverbs. Erteschik-Shir (2009) actually claims that *mon* ('I wonder') is always an adverb. I will not discuss this possibility further her, but I assume that it may be both a complementizer and an adverb.

(1917) (p. 62). But also other properties of preposed negation argue against a purely semantic account of the phenomenon. Conditional semantics is no sufficient criterion for preposing. Conditional V1-clauses do not allow preposing (contrary to e.g. Norwegian as shown in Johannessen (2000)).

(13) får (*ikke) vi (ikke) pengene [. . .] (DK)
get (*not) we (not) the.money [. . .]

Furthermore not even all conditional complementizers allow preposing - despite their semantics. The complementizers *såfremt* ('provided that') and *ifald* ('in case') do not allow preposing, while the complementizers *hvis* ('if') and *dersom* ('if') do.

(14) hvis / dersom ikke du vil deltage
if / if not you want to participate

(15) * såfremt / ifald ikke du vil deltage
provided that / in case not you want to participate

The relevant generalization appears to be a phonological restriction on the complementizers that allow preposing. Only mono-syllabic complementizers and complementizers with an unstressed final syllable (*'dersom* ('if')) allow preposing. The complementizers *så'fremt* and *i'fald* in (15) have a stressed final syllable. Given that preposed negation is obligatorily stressed this restriction may again be seen as a general restriction against having two adjacent stressed syllables.

2.2.1 Sentential or Constituent Negation

Complementizer clauses with the word order C-Neg-Subj are (in most cases) structurally ambiguous. The negation element may either be a preposed adverbial phrase or it may be a modifier of the Subject-DP, i.e. constituent negation of the subject. Cf. the following structural bracketing (the structural representation of (17) is motivated in Section 3).

(16) hvis [$_{NP}$ ikke regeringen] griber ind (DK)
if not government.DEF intervenes

(17) hvis [$_{ADVP}$ ikke] [$_{NP}$ regeringen] griber ind
if [[$_{ADVP}$] not government.DEF intervenes

However, the two structures are prosodically distinguished. Preposed negation is always stressed (Skafte-Jensen, 1995),[5] while constituent negation is unstressed.[6]

[5] Actually Skafte-Jensen (1995) note that only adverbs capable of being stressed can participate in preposing. This excludes modal adverbs/particles like *jo* ('you know') *vist* ('presumably').

[6] Jensen (2001) (p. 132) fails to distinguish preposed negation from constituent negation. She claims that the subject is obligatorily stressed in the order C-Neg-Subj. But preposed negation is

(18) a. hvis [$_{NP}$ $_{o}$ikke reGEringen] griber ind
if not government.DEF intervenes

b. hvis [$_{ADVP}$ IKKE] [$_{NP}$ regeringen] griber ind
if not the.government intervenes

Another difference between between the two structures in (16) is that preposed negation scopes over the whole subordinate clause and not just the subject. For that reason preposed negation cancels out ordinary negation in post-subject position. Thus preposing does indeed behave as sentential negation.[7]

(19) hvis ikke seerne ikke var advaret → hvis seerne VAR advaret
if not viewers.DEF not were warned → if viewers.DEF WERE warned

As expected, preposed negation like ordinary sentential negation licenses the presuppositional negative polarity adverb *heller* ('either') in the second clause.

(20) hvis du ikke forsøger at sikre dit netværk og Peter heller ikke gør
if you not try to secure your network and Peter either not does

Also preposed negation occurs in neg-raising environments, i.e. environments where a matrix negation scopes over an embedded clause (Horn, 1975, 1989; Sailer, 2006). Neg-raising only applies to sentential negation and not to constituent negation.

(21) hvis ikke du tror du kan klare det → hvis du tror, du ikke kan klare det (I)
if not you think you can manage it → if you think, you not can manage it

Thus there is very clear evidence that the word order C-Neg-Subj is structurally ambiguous and that preposed negation is different from constituent negation. Preposed negation behaves as sentential negation in crucial respects (if not in all respects as will be shown in Section 4).

---

also possible with DPs that cannot be stressed at all and that do not allow constituent negation since these subjects fail to meet the semantic condition of providing a contrastive reading of (contextually salient) alternative referents (Brandtler, 2006). Examples are expletives as in (i) and the pronoun *man* ('one').

(i) a. hvis IKKE det regner
if NOT it rains

b. * hvis [ikke DET] regner
if NOT it rains

[7] Also the occurrence of preposed negative quantifier phrases as in (4) above shows that we are dealing with sentential negation. Negative quantifier phrases cannot occur as constituent negation.

2.3 Negation-Preposing or Subject Lowering?

The particular word order C-Neg-Subj may arise in two ways: the negation is preposed as has been tacitly assumed in the previous discussion, or the subject is not in its canonical position outside the VP, but rather inside the VP. In both cases the negation element will precede the subject as illustrated in the figure below.

(22) hvis/'if' _ Peter ikke/'not' _ kommer/'comes'

To determine whether the negation is preposed or the subject is "lowered" we have to look at the distribution of other adverbs and other determiners.

As mentioned in Section 1, adverbs delimit the left-edge of the VP in embedded clauses. If the subject were inside the VP in the construction under discussion, we should expect adverbs left-adjoined to VP to precede the subject, but they do not. Adverbs occur between the subject and the finite verb also when the negation follows the complementizer, showing that the subject is still in its canonical position outside the VP. Cf.

(23) hvis ikke radiatorer og rør [alligevel] skal renoveres [...](I)
if not radiators and pipes anyway have.to be.renovated
'if radiators and pipes don't have to be renovated anyway'

Further evidence that negation is indeed preposed comes from the interaction with the pleonastic complementizer *at* ('that'). In colloquial Danish *hvis* ('if') may co-occur with the complementizer *at* ('that').

(24) hvis at jeg ikke gjorde det, ville de tvinge en overdosis i mig (I)
if that I not did it would they force an overdose into me

If we were dealing with subject-"lowering" rather than preposing of the negation, we should expect the negation *ikke* to occur after the pleonastic complementizer *at* ('that') as in (25) below.

(25) * hvis at t_i ikke $[_{Vp}$ jeg$_i$ gjorde det], ville de [...]
if that t_i not I$_i$ did it would they [...]

However, as noted by Jespersen (1917) (p. 62) and also in Pedersen (2009) (p. 327) the negation element obligatorily occurs to the left of the pleonastic complementizer *at* ('that') as expected if the negation element is indeed preposed and the subject is in its canonical position outside the VP.

(26) og hvis ikke at Folketinget kan stole på de oplysninger, (I)
and if not that parliament.DEF can trust PREP the information,

To sum up the basic properties of preposed negation so far: this section has established that the construction under discussion is indeed preposing of sentential negation which is lexically restricted to non-veridical complementizers with a certain phonological shape. They must be mono-syllabic or contain an unstressed final syllable. The next section will investigate the syntactic structure of preposed negation.

3 The Structure of Preposed Negation

In this section I turn to the structural analysis of preposed negation. Preposed negation appears adjacent to the complementizer (Pedersen, 2009) and it is semantically and lexically licensed by the complementizer as shown in Section 2. This pattern may imply three things: the complementizer and the preposed negation form a kind of composite complementizer, the negation cliticizes to the complementizer (Johannessen, 2000) or the complementizer and the negation is a lexicalized collocation as suggested by Pedersen (2009). Support for these structural possibilities comes from the fact that negation in some languages surfaces as a lexical element in the syntax (a non-projection word), i.e. the negation does not project a syntactic phrase as claimed for Swedish in Toivonen (2003). I will, however, conclude that preposed negation can indeed be syntactically complex and that a lexical analysis or an analysis as a clitic is untenable. Secondly I show that preposed negation cannot be analysed as either adjunction to C or the following S. Instead I will argue that preposed negation is a daughter of CP and that the negation element or the negative quantifier phrase is extracted from the following S. This allows for two possible analyses of preposed negation as either a complement of the complementizer (as claimed for finite English negation in Kim and Sag (2002)) or as a special construction. Given the particular semantics of preposed negation discussed in Section 4. I will argue that it constitutes a special construction.

3.1 Preposed Negation as a Lexical Structure

A first hypothesis is that preposed negation is part of a lexical structure, i.e. that the negation and the complementizer form a kind of composite complementizer even though complementizers are traditionally assumed to form a closed word class. But if preposed negation is the result of a lexical process we should expect it to be an operation on lexical items and we should expect it to obey blocking-constraints such that existing words block the formation of words with the same semantics.

Preposed negation cannot be the output of a lexical process given that the negation element can also be a syntactic phrase not available for further lexical processes. The negation element may contain (negative polarity) degree words such as *slet* ('at all') (cf. (27)) and it can also be a negative quantifier phrase (a DP or an NP) with prenominal modification, cf. (28). Thus the negation in Danish is a projecting word as opposed to the analysis of negation in Swedish in Toivonen (2003).

(27)	Hvis	slet	ikke	der	står	noget
	if	at.all	not	it	says	anything

(28)	hvis	[ingen	(direkte)	arvinger]	der	er
	if	no	(direct)	heirs	there	are

The possibility of preposed quantifier phrases also argue against a purely collocational analysis as suggested in (Pedersen, 2009) since such quantifier phrases are productively formed and hardly count as collocational constructs.

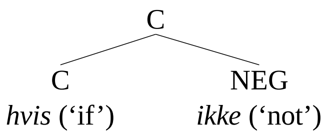
Furthermore a composite complementizer consisting of *hvis* (‘if’) and *ikke* (‘not’) ought to be blocked by the presence of the complementizer *medmindre* (‘unless’) which lexicalizes conditional semantics taking scope over negation. The fact that it is not blocked suggests that preposed negation is a syntactic formation. Thus I conclude that preposed negation is indeed a phenomenon to be dealt with in the syntax.

3.2 The Syntax of Preposed Negation

Preposed negation is a syntactic phrase but where does it attach structurally? Is it a modifier of the following S or is it a modifier of the preceding C? I will discuss both possibities in turn and conclude that the data argue against both possibilities.

Johannessen (2000) (p. 14) suggests that preposed negation in Norwegian is adjoined to C as shown in (29) below.

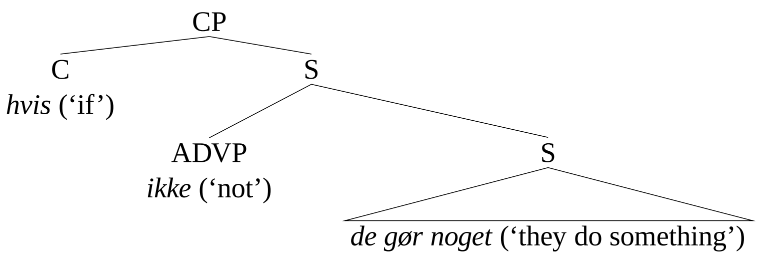
(29)



In fact Johannessen (2000) suggests that preposed negation cliticizes to C, but as already shown in (27) and (4) above, preposed negation in Danish can be syntactically complex and hence cannot be a clitic. Alternatively the negation phrase is a modifier of the complementizer so that the structure in (29) is a modificational adjunction structure. The problem with this analysis is that the negation is within the scope of the complementizer. Conditional semantics always takes scope over the negation element giving the following interpretation: IF(NOT(p)). This is unexpected if the negation is a modifier of the complementizer, since the modifier is otherwise assumed to take scope over the modified head in modificational structures. Thus an analysis as modificational adjunction to C is at odds with the semantic composition of the structure.

Another possibility is that preposed negation left-adjoins to the following S yielding the structure shown in (30) below.

(30)



A first problem is that the structure in (30) obscures the fact that there is a close dependency between the complementizer and the preposed negation: they must be adjacent and preposing is lexically restricted (cf. Section 2.2). If the negation adjoins to the following S it is difficult to state that adjunction to S is only possible if the negation is preceded by a complementizer[8] with a particular semantics and a particular phonological shape. But there is also other evidence that (30) cannot be the right structure. If the negation is allowed to left-adjoin to S, we should expect it also to be able to left-adjoin to the second conjunct of two coordinated Ss occurring with the right kind of complementizer. But this appears to be marginal at best. Cf.

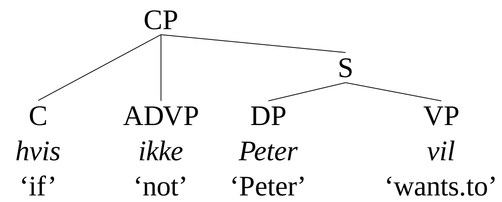
(31) ??/* hvis [$_S$ ikke Peter vil] og [$_S$ ikke Louise er syg]
if not Peter will and not Louise is ill

In addition preposed negation may be stranded in ellipsis. This is unexpected under the adjunction analysis since there is no S for the negation element to adjoin to as also noted for English in Kim and Sag (2002).

(32) [Hvis ikke], er det ikke ulovligt at have dem stående (DK)
If not is it not illegal to have them around

The ellipsis data in (32) and the fact that the negation only marginally can show up before the second conjunct of a coordination as in (31) is expected if the negation element does not adjoin to the following S but if it is a daughter of CP. Thus I conclude that preposed negation is a daughter of CP as shown in (33) below.

(33)



However, this analysis makes preposed negation remarkably different from ordinary negation. Ordinary negation is adjoined to VP and does not occur as a daughter of CP. Ordinary negation occurs in adjunction position to the left of the verbal head, it can be separated from the verbal head by other adjuncts and it may occur adjoined to the second VP-conjunct of a coordination.

(34) fordi han [ikke *ser* filmen]
because he not sees movie.DEF

[8]Negation adjoining to an S is otherwise only possible in so-called metanegation (Horn, 1989; Christensen, 2005). Negation adjoins to an (initial or parenthetical) unembedded complementizer clause with the complementizer *at* ('that') or *fordi* ('because') and serves to deny an otherwise invited (conversational) implicature.

(i) ikke at jeg frygter for hun bliver sur, men jeg er bange for ...
not that I am afraid PREP she gets angry, but I am afraid PREP...

(35)	fordi	de	[ikke	{som	raske	mennesker}	{hurtigt}	*er*
	because	they	not	such as	healthy	people	quickly	are
	i stand til	at slukke	ilden	]				
	capable of	extinguishing	fire.DEF (DK)					

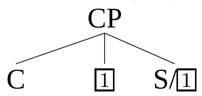
(36)	fordi	han	ikke	læser	avis	og	[ikke	ser	fjernsyn]
	because	he	not	reads	newspaper	and	not	watches	television

Thus it appears that we would have very different analyses of ordinary negation and preposed negation. In addition we have not yet accounted for preposing of negative quantifier phrases. Negative quantifier phrases are part of a filler-gap dependency given that the preposed phrase must be identified with a complement gap to the right of the main verb. Cf.

(37)	hvis	du	ingen	børn$_i$	har	$\_{i}$
	if	you	no	children	have	

(38)	hvis	ingen	børn$_i$	du	$\_{i}$	har	$\_{i}$
	if	no	children	you		have	

The preposed phrase *ingen børn* (‘no kids’) is not just an adjunct but must be associated with the object of the verb *har* (‘has’). Pursuing a unified analysis of preposing, it thus appears to be the case that not only preposed quantifier phrases but also the preposed negation is a filler. The advantage of this analysis is that ordinary negation as well as negative quantifier phrases adjoin to the VP and both kinds of negative constituents may be dislocated to the left of the complementizer given the right kind of complementizer. The analysis of preposing as a filler-gap dependency allows for a unified analysis of negation and negative quantifier phrases.[9] Cf. the following representation.



Thus the conclusion of this section is that preposing is a filler-gap dependency where a complementizer selects an S with a slashed constituent and allows this element to surface as a kind of complement of the complementizer.[10]

---

[9]The ellipsis data shown in (32) may, however, be problematic for this analysis of the preposed negation element as extraction. In elliptical structures the gap of the negation is elided while the filler is still there. While elision of a clause from which an argument has been extracted appears to be marginal, elision is much better if the extracted element is an adjunct.

(i)	??/*	Peter	tror	Poul	bliver	løsladt	i morgen.	Hvem	tror	han?
		Peter	thinks	book.DEF	is	released	tomorrow.	Who	thinks	he?

(ii)	Peter	tror	Poul	bliver	løsladt	i morgen.	Hvornå	tror	han?
	Peter	thinks	Poul	is	released	tomorrow.	When	thinks	he?

Thus it appears that the extraction site of adjuncts can be elided and preposed negation and preposed quantifier phrases positionally behave as adjuncts.

[10]A problem for the analysis as extraction is that preposed negation does not seem to obey “Across-the-Board”-constraints otherwise observed in coordination from which a constituent is extracted.

4 Preposed Negation as *verum*-Negation

Having discussed the basic properties and the syntax of preposed negation, a second question arises: why does the syntax of Danish allow for this additional placement of the negation element? Two factors appear to be crucial to the understanding of preposing: the behaviour of (strong) negative polarity (NPI) items and the interpretation of (indefinite) subjects with preposed negation. In this section I will show that the behaviour of polarity items (PI) point to the conclusion that preposed negation is associated with VERUM-negation (in the sense of Höhle (1992)). Furthermore I will show that preposed negation is associated with an *all_comment* information structure, i.e. a clause lacking an (aboutness-) TOPIC.

4.1 The Behaviour of Strong Polarity Items

As observed in Section 2 preposed negation behaves like (ordinary) sentential negation in crucial respects. However, preposed negation shows a totally different behaviour wrt. strong polarity items. Strong polarity items (either positive or negative) are sensitive to *antiveridical* contexts (Giannakidou, 1999): strong negative PIs are licensed by negation (or negative elements), strong positive PIs are ruled out by negation. Weak PIs on the other hand are licensed in *non-veridical* contexts (Giannakidou, 1999) and may thus occur independently in conditional clauses. Weak PIs are therefore expected to occur with preposed negation, given that also preposed negation is licensed in non-veridical contexts (cf. Section 2). Example (39) shows that the weak PI *nogensinde* ('ever') can also occur in an unnegated conditional clause.

(39) Hvis (ikke) du nogensinde har oplevet mursten, stålplader og
if not you ever have seen bricks steel plates and
jernstænger blive slået igennem med panden (I)
iron sticks be cut through with forehead.DEF

Strong NPIs, however, are licensed in conditional clauses by ordinary negation, but they are marginal at best with preposed negation. In (40a) ordinary negation licenses the polarity item *en rød øre* ('a red cent'). Example (40b) is marginal. As one informant put it: it sounds as if you expect the users to pay a red cent, which is nonsense. Thus it seems that preposed negation is too weak to license strong NPIs.

Given ATB-constraints on coordination a preposed negation ought to have scope over both conjuncts always. While this is indeed possible, preposed negation does not have to have scope over the second conjunct. Thus the following examples allows for two readings: $\neg(p \vee q)$ and $\neg p \vee q$.

(i) hvis ikke CDU går tilbage og FDP gå frem
if not CDU goes back and FDP goes foward
'if note CDU loses votes and FDP gains votes'

(40) a. hvis brugerne ikke skal lægge *en rød øre*, når de stiger på(I)
if users.DEF not must pay a red cent, when they enter PREP

b. ?? *hvis ikke* brugerne skal lægge *en rød øre*, når de stiger på
if not users.DEF must pay a red cent, when they enter PREP

In a similar vein preposed negation is also too weak to rule out strong positive PIs. The underlined strong positive PI in (41a) – a somewhat outdated expression meaning “to be a top-professional” – cannot occur in a conditional clause with ordinary negation, However it is much better with preposed negation as in (41b).

(41) a. ??/* hvis du bare ikke *kan det pis*, skal du lade være
if you just not can that stuff, don’t do it
‘if you are not a top-professional, then don’t do it’

b. hvis ikke du bare *kan det pis*, skal du lade være
if not you just can that stuff, don’t do it
‘if you are not a top-professional, then don’t do it’

On the account of PIs in Giannakidou (1999), NPIs are licensed when they are in the immediate scope of an anti-veridical operator such as *ikke* (‘not’). Thus it appears that NPIs in clauses with preposed negation are not in the immediate scope of the negation. This failure to license strong polarity items is also observed with negated VERUM-focus, i.e. when a finite verb within the scope of ordinary negation is stressed (Höhle, 1992).[11]

(42) ??/* brugerne GIVER ikke en rød øre
users.DEF give not a red cent
‘it isn’t the case, that the users give a red cent’

Negation focus, on the other hand, i.e. stress on the negation element, does license negative polarity items, arguing against an analysis of preposed negation as involving negation focus, despite the fact, that the negation is stressed.

(43) brugerne giver IKKE en rød øre
users.DEF give ikke a red cent
‘the users really don’t give a red cent’

Following this reasoning it appears that preposed negation is associated with negated VERUM-focus. The additional VERUM-predicate *it is the case that* (Höhle, 1992) thus may explain the peculiar behaviour of the strong PIs. Negation of the predicate *it is the case* does not license NPIs (Gajewski, 2007; Horn, 1989; van der Wouden, 1997). Thus, it appears that ordinary negation (without VERUM-focus)

[11] The verb in (42) is within the scope of negation given that the verb in V1/V2-clauses is associated with its canonical position to the right of the negation as observed in non-V1/V2-clauses. I assume that it is associated with a trace as in the analysis of V1 and V2 in German in Müller (2008) (chap. 9).

gives rise to the paraphrase in (44), while preposed negation gives rise to the paraphrase in (45).

(44) hvis brugerne ikke skal lægge en rød øre ...
if users.DEF not have.to pay a red cent ...
⇒ if the users do not have to pay a red cent...

(45) ?? hvis ikke brugerne skal lægge en rød øre ...
if not users.DEF have.to pay a red cent ...
⇒ if *it is not the case* that the users have to pay a red cent...

As the paraphrases make clear, VERUM embeds a positive proposition, thus explaining the impossibility of negative PIs and the possibility of positive PIs. Preposed negation introduces a VERUM-predicate within its scope. With preposed negation the polarity of the conditional clause is negated, not the proposition as such.

4.2 The Information Structure of Preposing

But what distinguishes ordinary negation with VERUM-focus from preposed negation, if preposed negation is also associated with VERUM-focus? Preposed negation is associated with an embedded clause with a particular information structure. Where embedded clauses with ordinary negation are associated with a basic *topic_comment*-articulation, clauses with preposed negation are characterized by the absence of a TOPIC. Clauses with preposed negation do not have an (aboutness-) TOPIC in the sense of Krifka (2007). Evidence comes from the use of non-referential subjects, the interpretation of indefinite subjects and the use of TOPIC-binding particles.

Preposed negation is preferred with non-referential subjects such as *enhver* ('everybody') and *alle* ('everyone').[12] Cf.

(46) Men hvis ikke enhver skulle blive depri af denne elendige
but if not everyone should get depressed by this horrible
sommer [...] (I)
summer [...]

(47) ?? Men hvis enhver ikke skulle blive depri af denne elendige
but if everyone not should get depressed by this horrible
sommer [...]
summer [...]

Indefinite pronouns like *enhver* ('everybody') are non-referential and since an (aboutness-) TOPIC presupposes referentiality, indefinite pronouns are degraded as TOPICS (Pittner, 2004; Frey, 2004). The preference of preposed negation with

[12] This observation is due to Line Mikkelsen (p.c.).

non-referential pronouns thus receives a straight-forward explanation, if preposing is associated with the lack of a topic.

Also the interpretation of indefinite subjects point to the conclusion that clauses with preposed negation have no TOPIC. Following Diesing (1992) indefinite NPs exhibit either a weak (existential) reading or a strong (generic or proportional) reading. The weak reading is typical of non-topicality, while the strong reading is typical of topicality (Diesing, 1992). Preposed negation does indeed favour the weak reading of indefinites again suggesting that the subject is no TOPIC. Cf. (48) where the indefinite subject has an existential reading.

(48) han ville uden tvivl have slået sig ihjel, hvis ikke [en rotte] i det
he would beyond doubt have been killed if not a rat in that
samme var kommet løbende hen over gulvet (DK)
moment had come running across floor.DEF

Ordinary negation in turn favours a strong reading of indefinites as expected if the subject is a TOPIC. In (49) the indefinite is associated with a generic reading.

(49) Hvis [en atlet] ikke vil eller glemmer at fortælle Anti-Doping
If [an athlete] not will or forgets to tell Anti-Doping
Danmark, [. . .] (DK)
Denmark [. . .]

This analysis of the information structure of preposing is further reinforced by the behaviour of TOPIC-binding particles. TOPIC-binding particles are particles indicating TOPIC-shift or TOPIC-continuation (Breindl, 2008). A particle such as *derimod* ('in contrast') can attach to a subject NP of either a V2-clause or a non-V2-clause[13] to indicate a TOPIC-shift as shown in (50). In conditional clauses TOPIC-binding particles are fine with ordinary negation (51), but highly degraded with preposed negation (52), since there is no TOPIC to bind.

(50) men at derimod stress ser ud til at være synderen (I)
but that in contrast stress appears to be sinner.DEF

(51) Hvis derimod lønstigningerne ikke tager af (I)
if in contrast wages rising.DEF not reduce

(52) ?? Hvis ikke derimod lønstigningerne tager af
if not in contrast wages rising.DEF reduce

On the evidence presented in this section preposed negation is used to negate the VERUM of an *all_comment* clause, i.e. a clause lacking a TOPIC.

[13] In V2-clauses the particle occurs to the right of the subject in the so-called *Nacherstposition* (Breindl, 2008).

(i) Regeringen derimod prøver at få danskerne til at arbejde mere (I)
government.DEF in contrast tries to make danes.DEF PREP to work more

5 An HPSG Approach

The crucial argument for positing a construction for preposed negation is that it is associated with a particular semantics. Preposed negation is associated with negated *verum* focus, thus the construction itself introduces a *verum* predicate which in turn is within the scope of negation. Cf. the examples below.

(53) a. hvis Peter ikke vinder
if Peter not wins
CONDITIONAL > NEGATION > PROPOSITION

b. hvis ikke Peter vinder
if not Peter wins
CONDITIONAL > NEGATION > VERUM > PROPOSITION

In Minimal Recursion Semantics the semantic representation is given as a bag of basic relations (RELS) which in turn are connected by means of labels giving the functor-argument relationships holding between the individual predicates (LBL and ARG*n*). Scopal relationships between the individual relations are indicated by so-called *qeq*-constraints (*equality modulo quantifiers*) in the feature H_CONS. An argument position which is *qeq*-related to a label does not have to be filled by that label. The argument position can be filled by another label which in turn has the first label as an argument. Thus other scopal elements can intervene between two elements, where the first outscopes the other (Copestake et al., 2005) (p. 297). The lexical entry for the complementizer *hvis* (‘if’) is given below.

$$\left[\begin{array}{l} \text{SYNSEM} \mid \text{LOC} \left[\begin{array}{l} \text{CAT} \left[\begin{array}{l} \text{HEAD} \left[\begin{array}{l} \textit{compl} \\ \text{MOD} \left\langle \text{S} \right\rangle \end{array}\right] \\ \text{SUBCAT} \left\langle \text{S}\left[\text{LOC} \mid \text{CONT} \mid \text{LTOP} \ \boxed{1}\right] \right\rangle \end{array}\right] \\ \text{CONT} \mid \text{LTOP} \ \boxed{2} \end{array}\right] \\ \text{RELS} \left\langle \left[\begin{array}{l} \textit{if\_rel} \\ \text{LBL} \ \boxed{2} \\ \text{ARG} \ \boxed{3} \end{array}\right] \right\rangle \\ \text{H\_CONS} \left\langle \left[\begin{array}{l} \textit{qeq} \\ \text{HARG} \ \boxed{3} \\ \text{LARG} \ \boxed{1} \end{array}\right] \right\rangle \end{array}\right]$$

The complementizer selects its clause through the feature SUBCAT. The complementizer introduces the basic predicate *if_rel* and the conditional semantics takes as its argument the subcategorized S or a quantifier outscoping the subcategorized S as guaranteed by the *qeq*-constraint in H_CONS. This is crucial in accouting for preposed negative quantifier phrases. The entry for the negation is given below.

The negation selects its modified VP through the feature MOD. The negation introduces the basic relation *neg_rel* taking as its argument the modified VP modulo intervening quantifiers.

$$\begin{bmatrix} \text{SYNSEM}\,|\,\text{LOC} & \begin{bmatrix} \text{CAT}\,|\,\text{HEAD}\,|\,\text{MOD}\left\langle \text{VP}\big[\text{LOC}\,|\,\text{CONT}\,|\,\text{LTOP}\ \boxed{1}\big] \right\rangle \\ \text{CONT}\,|\,\text{LTOP}\ \boxed{2} \end{bmatrix} \\ \text{RELS} & \left\langle \begin{bmatrix} \textit{neg\_rel} \\ \text{LBL}\ \boxed{2} \\ \text{ARG}\ \boxed{3} \end{bmatrix} \right\rangle \\ \text{H\_CONS} & \left\langle \begin{bmatrix} \textit{qeq} \\ \text{HARG}\ \boxed{3} \\ \text{LARG}\ \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix}$$

Consider next the construction for preposed negation.

$$\begin{bmatrix} \text{SYNSEM} & \begin{bmatrix} \text{LOC}\,|\,\text{CAT} & \begin{bmatrix} \text{HEAD} & \boxed{1} \\ \text{SPR} & \langle\rangle \\ \text{SUBCAT} & \langle\rangle \end{bmatrix} \\ \text{NONLOC}\,|\,\text{INHER} & \begin{bmatrix} \text{REL} & \langle\rangle \\ \text{SLASH} & \langle\rangle \end{bmatrix} \end{bmatrix} \\ \text{C\_CONT} & \begin{bmatrix} \text{HOOK}\ \boxed{2} \\ \text{RELS}\left\langle \begin{bmatrix} \textit{verum\_rel} \\ \text{LBL}\ \boxed{3} \\ \text{ARG}\ \boxed{4} \end{bmatrix} \right\rangle \\ \text{H\_CONS}\left\langle \begin{bmatrix} \textit{qeq} \\ \text{HARG}\ \boxed{5} \\ \text{LARG}\ \boxed{6} \end{bmatrix}, \begin{bmatrix} \textit{qeq} \\ \text{HARG}\ \boxed{7} \\ \text{LARG}\ \boxed{3} \end{bmatrix}, \begin{bmatrix} \textit{qeq} \\ \text{HARG}\ \boxed{4} \\ \text{LARG}\ \boxed{8} \end{bmatrix} \right\rangle \end{bmatrix} \\ \text{NH-DTRS} & \left\langle \begin{bmatrix} \text{SYNSEM}\,|\,\text{LOC} & \begin{bmatrix} \text{CAT}\begin{bmatrix} \text{HEAD}\ \boxed{1}\ \textit{compl} \\ \text{SUBCAT}\left\langle \boxed{9} \right\rangle \end{bmatrix} \\ \text{CONT}\ \boxed{2} \end{bmatrix} \\ \text{RELS} & \left\langle \begin{bmatrix} \textit{non\_veridical} \\ \text{ARG1}\ \boxed{5} \end{bmatrix} \right\rangle \end{bmatrix}, \begin{bmatrix} \text{SYNSEM}\,|\,\text{LOC}\ \boxed{10} \\ \text{RELS}\left\langle \begin{bmatrix} \textit{neg\_rel} \\ \text{LBL}\ \boxed{6} \\ \text{ARG}\ \boxed{7} \end{bmatrix} \right\rangle \oplus \textit{list} \end{bmatrix}, \right. \\ & \left. \begin{bmatrix} \text{SYNSEM}\ \boxed{9} \begin{bmatrix} \text{LOC}\,|\,\text{CONT}\,|\,\text{KEY}\,|\,\text{LBL}\ \boxed{8} \\ \text{NONLOC}\,|\,\text{INHER}\,|\,\text{SLASH}\left\langle \boxed{10} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

The construction for preposed negation defines three daughters: the complementizer, the negation and the clause. The first daughter is the head of the construction and it is constrained to be a non-veridical complementizer subcategorizing for the third daughter (the clause). The second daughter is constrained to be negated (it contains the negation relation as the first of its basic relations). This semantic constraint ensures that not only the negation *ikke* ('not') but also negated quantifier phrases can be preposed. The second daughter is the filler of the gap associated with the third daugher (the LOC(al) value of the second daughter is structure-shared with the SLASH-value of the third daughter), ensuring that negative preposed quantifier phrases are analyzed as complements of the verb. The motivation for positing a separate construction is given in the constructional content (C_CONT). The construction introduces the basic *verum*-relation which has the proposition in its scope. The scoping constraints in H_CONS state that the complementizer out_scopes the negation, that the negation outscopes the *verum*-relation and that the *verum*-relation outscopes the proposition. These constraints give the scoping relationships shown in (53b). The semantic representation for the whole construction is constrained by an independent semantics principle to be the union of the RELS and H_CONS of the daughters.

6 Conclusion

The paper has provided an analysis of preposed negation in Danish uncovering a host of properties that appear to have gone unnoticed in the literature. It is proposed that preposed negation is associated with negated *verum*-focus of a propostion lacking a topic and it has been argued that this should be analyzed as a construction given that this semantics is not associated with a particular lexical entry but with a specific ordering of existing lexical entries. The analysis has been formalized in construction-based HPSG and it has been implemented[14] in the TRALE system (Meurers et al., 2002; Penn, 2004; Müller, 2007) as part of a grammar fragment of Danish which uses a core grammar for German, Persian, Mandarin Chinese and Maltese. The respective grammars can be downloaded at URL: http://hpsg.fu-berlin.de/Software/.

References

Ambridge, Ben and Goldberg, Adele E. 2008. The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics* 19(3), 357–389.

Brandtler, Johan. 2006. On Aristotle and Baldness - Topic, Reference, Presupposition of Existence, and Negation. *Working Papers in Scandinavian Syntax* 77, 177–204.

[14]The implementation is due to Stefan Müller.

Breindl, Eva. 2008. Die Brigitte nun kann der Hans nicht ausstehen. Gebundene Topiks im Deutschen. *Deutsche Sprache* 01, 27–49.

Christensen, Ken Ramshøj. 2005. *Interfaces: Negation - Syntax - Brain*. Ph. D.thesis, University of Aarhus.

Copestake, Ann, Flickinger, Dan, Pollard, Carl and Sag, Ivan A. 2005. Minimal Recursion Semantics. *Research on Language and Computation* 3, 281–332.

Diesing, Molly. 1992. *Indefinites*, volume 20 of *Linguistic Inquiries Monograph*. MIT Press.

Erteschik-Shir, Nomi. 2009. The Phonology of adverb placement, Object Shift and V-2; the case of Danish 'MON'.

Frey, Werner. 2004. A Medial Topic Position for German. *Linguistische Berichte* 198.

Gajewski, Jan. 2007. Neg-Raising and Polarity. *Linguistics and Philosphy* 30, 289–328.

Giannakidou, Anastasia. 1999. Affective Dependencies. *Linguistics and Philosophy* 22(4), 367–421.

Höhle, Tilman N. 1992. Über Verum-Fokus im Deutschen. In Joachim Jacobs (ed.), *Informationsstruktur und Grammatik*, volume 4 of *Linguistische Berichte*, pages 112–142, Westdeutscher Verlag.

Horn, Laurence R. 1989. *A Natural History of Negation*, volume 20. The University of Chicago Press.

Horn, L.R. 1975. Neg-raising predicates: Toward an explanation. In *CLS*, volume 11, pages 279–294.

Jensen, Britta. 2001. On sentential negation in the Mainland Scandinavian languages. In Maria Liakata, Britta Jensen and Didier Maillat (eds.), *Oxford University Working Papers in Linguistics, Philology and Phonetics*, volume 6, pages 115–137.

Jespersen, Otto. 1917. *Negation in English and other languages*, volume I,5 of *Det Kgl. Danske Videnskabernes Selskab. Historisk-filologiske Meddelelser*. Høst & Søn.

Johannessen, Janne Bondi. 2000. Negasjonen *ikke*: kategori og syntaktisk posisjon, tekstlaboratoriet, Universitetet i Oslo.

Kim, Jong-Bok and Sag, Ivan A. 2002. Negation without Head-Movement. *Natural Language and Linguistic Theory* 20(2), 339–412.

Krifka, Manfred. 2007. Basic notions of Information Structure. In C. Féry et al. (ed.), *The notions of information structure. Interdisciplinary Studies on Information Structure*, volume 6 of *ISIS*, Potsdam: Universitätsverlag Potsdam.

Meurers, Walt Detmar, Penn, Gerald and Richter, Frank. 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*, pages 18–25, proceedings of the Workshop held at 40th Annual Meeting of the Association for Computational Linguistics. Philadelphia, PA.

Müller, Stefan. 2007. The Grammix CD Rom. A Software Collection for Developing Typed Feature Structure Grammars. In Tracy Holloway King and Emily M. Bender (eds.), *Grammar Engineering across Frameworks 2007*, Studies in Computational Linguistics ONLINE.

Müller, Stefan. 2008. *Head-Driven Phrase Structure Grammar – Eine Einführung*. Stauffenberg, second edition.

Pedersen, Karen Margrethe. 2009. Indre *at* i danske dialekter. In Rita Therkelsen and Eva Skafte Jensen (eds.), *Dramatikken i Grammatikken*, pages 321–333, Roskilde Universitet.

Penn, Gerald. 2004. Balancing Clarity and Efficiency in Typed Feature Logic Through Delaying. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume*, pages 239–246, Barcelona, Spain.

Pittner, Karin. 2004. Where syntax and semantics meet: adverbial positions in the German middle field. In Jennifer R. Sutin, Stefan Engelberg and Gisa Rauh (eds.), *Adverbials: The interplay between meaning, context and syntactic structure*, volume 70 of *Linguistics Aktuell*, pages 253–288, Amsterdam/Philadelphia: John Benjamins.

Sailer, Manfred. 2006. *Don't believe* in Underspecified Semantics. In O. Bonami and P. Cabredo Hofherr (eds.), *Empirical Issues in Syntax and Semantics*, volume 6, pages 375–403.

Skafte-Jensen, Eva. 1995. Neksusadverbialets placering i ledsætninger. *NyS* 20, 77–90.

Toivonen, Ida. 2003. *Non-Projecting Words. A Case Study of Swedish Particles*. Dordrecht: Kluwer.

van der Wouden, Ton. 1997. *Negative Contexts - Collocation, polarity and multiple negation*. Routledge Studies in Germanic Linguistics, Routledge: London, New York.

Hindi Aspectual Complex Predicates*

Shakthi Poornima and Jean-Pierre Koenig

State University of New York at Buffalo

Proceedings of the HPSG09 Conference

Georg-August-Universität Göttingen, Germany

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/

*This is a revised and expanded version of Poornima and Koenig (2008).

Abstract

This paper discusses ergative case assignment in Hindi and its interaction with aspectual verb complexes or complex predicate constructions. It is shown that ergative case is assigned by the last head in the aspectual verb complex and that ergative case on the subject of intransitive verbs denoting bodily-functions is associated with a counter-to-expectation meaning. It is then shown that aspect complex predicates in Hindi involve two distinct syntactic structures, which have similar semantics. While one syntactic structure involves argument composition, the other involves a head-modifier structure. It is argued that the existence of two structures favor approaches to the interface between syntax and semantics which do not require a uniform isomorphism between the semantics and syntax of aspect.

1 Introduction

Determining variation between languages allows linguists to hypothesize about how much natural languages can actually vary. The syntax of aspect is a fertile ground for comparing approaches that explain variation in the interface between syntax and semantics, given the varied surface realization of aspectual functors (e.g., verbal affixes, auxiliaries, ordinary verbs, see Bybee et al. (1994) for details). Koenig and Muansuwan (2005) compared two class of hypotheses regarding the mapping between aspectual functors and syntactic structure. One class of hypotheses, dubbed the UNIFORMITY HYPOTHESIS, holds that at a particular level of representation, one can establish an almost isomorphic, cross-linguistically uniform, correspondence between the syntax and semantics of aspect. This is best exemplified by Cinque (1999), who posits that the geometry of verbal functional projections (head-complement relations, in particular) corresponds for the most part to the geometry of semantic functor-argument relations. Another class of hypotheses, dubbed REPRESENTATIONAL MODULARITY, holds that syntactic and semantic structures are independent levels of representations related by correspondence rules and constraints which do not require a one-to-one relation either within or across languages. As a consequence, Koenig and Muansuwan (2005) argue, the correspondence between the syntax and semantics of aspect is weaker and cross-linguistic variation in the surface expression of aspectual distinctions might reflect the true extent of the non-correspondence between syntactic and semantic structure. Koenig and Muansuwan present data from Thai that support the Representational Modularity hypotheses. In this paper, we present corroborating data from Hindi which show that the same (or, at least, identical in all relevant respects) aspectual notions can be expressed in Hindi in two distinct ways. Aspect markers can be verbs that take main verbs as complements to form complex predicates or they can be verbs that modify main verbs. Although Hindi aspect markers have been described in the previous literature (see (Hook, 1975; Kachru, 1980; Butt, 1994)), a critical interaction between the order of verbs in the complex predicate structure and case assignment and verb-subject agreement has not. This interaction provides compelling evidence, we suggest, that the syntactic structures involved in these two kinds of aspectual complex predicates are truly distinct and cannot be reduced to the same syntactic structure "deep down". Hindi thus parallels the split in the syntax of aspect that Koenig and Muansuwan (2005) argue exists in Thai.

2 Hindi Aspectual Complex Predicates

In Hindi, aspectual complex predicates or verb complexes (we will use the two expressions interchangeably) are formed by the combination of a verb that denotes a situation-type (hereafter, the MAIN verb) and a a finite LIGHT verb, an aspectual functor which semantically modifies the main verb's meaning. Light verbs are homophonous with form-identical lexical verbs that do not carry aspectual meanings. We use the term *light* to suggest that their meaning is more abstract than their non-aspectual counterpart meanings. A list of the most common Hindi light verbs is presented in Table 1. The combination of the main verb and light verb involve two types of structures. In what is standard for a head-final language, the non-finite main verb can be followed by a finite light verb (1) to form a *standard aspectual complex predicate construction*. The order of the main and light verbs can also be reversed to form a *reverse aspectual complex predicate construction*, where the finite light verb precedes the non-finite main verb (2).[1]

(1) *Ram=ne Leela=ko tamaachaa maar di-yaa*
Ram=Erg Leela=Dat slap.M.Sg hit:MV give-M.Sg:LV
'Ram slapped Leela (hit Leela with a slap).'

(2) *Ram=ne Leela=ko tamaachaa de maar-aa*
Ram=Erg Leela slap.M.Sg give:LV hit-M.Sg:MV
'Ram slapped Leela (hit Leela with a slap).'

Note that the inflection is carried by the light verb in the *standard*, but by the main verb in the *reverse* aspectual complex predicate construction (hereafter standard and reverse CP construction). As we will show in more detail below, the two constructions differ in more than just linear ordering. More generally, we will argue that the two constructions differ in terms of which verb is the construction's head: the light verb in the standard CP construction, and the main verb in the reverse CP construction.

[1]The gloss used for a light verb refers to its meaning as a full verb. Abbreviations are as follows: MV = main verb, LV = light verb, F = feminine, M = masculine; Erg = ergative, Nom = nominative, Gen = genitive, Dat = dative, Acc = accusative, Inst = instrumental, Loc = locative; Inf = infinitive; Pfv = perfective, Impfv = imperfective; Pres = present; Pron = pronoun; Sg = singular, Pl = plural. The marker '-' indicates a morpheme boundary, '=' separates a clitic from a lexical item. Following ':' we indicate whether the verb is a main verb or a light verb. Most examples in this paper were created by the author and cross-verified by 3 native speakers from northern India.

Transitive light verbs	**Intransitive light verbs**
baith (sit)	*aa* (come)
Daal (put)	*jaa* (go)
de (give)	*paD* (fall)
le (take)	*nikal* (leave)
maar (hit)	*uth* (rise)
nikaal (remove)	

Table 1: Aspectual Light Verbs

3 Constituent Structure of Aspectual Complex Predicates

This section analyzes the constituent structure of the standard and reverse CP constructions. We show that the two verbs form a constituent in both constructions. They differ in that only the standard CP constuction allows certain particles to intervene between the two verbs and that the range of auxiliaries that can follow the light verb-main verb combination is more restricted in the reverse CP construction.

Butt (1994) shows that Hindi aspectual complex predicate constructions are monoclausal and that, furthermore, the main and light verbs form a constituent. We briefly summarize Butt's arguments here (expanding her arguments when needed to the reverse construction, which Butt does not discuss). For instance, although the ordering of subjects and objects is fairly free in Hindi, the main verb and the light verb in an aspectual complex predicate must be reordered with other clausal constituents *as a unit*, as demonstrated for the reverse construction in (3) (see But, op.cit. for similar data on the standard CP construction).

(3) a. *[Leela=ne] [Shyam=ko] [ciTThii] [maar likh-ii]*
[Leela.F=Erg] [Shyam.M=Dat] [letter.F.Sg] [hit:LV write-Perfv.F.Sg:MV]
'Leela wrote a letter to Shyam.'

b. *[Shyam=ko] [Leela=ne] [ciTThii] [maar likhii]*

c. *[Leela=ne] [maar likhii] [ciTThii] [Shyam=ko]*

d. *[maar likhii] [Leela=ne] [Shyam=ko] [ciTThii]*

e. *[maar likhii] [ciTThii] [Shyam=ko] [Leela=ne]*

f. *[ciTThii] [maar likhii] [Leela=ne] [Shyam=ko]*

g. *[ciTThii] [maar likhii] [Shyam=ko] [Leela=ne]*

h. **[ciTThii] [likhii] [Shyam=ko] [Leela=ne] [maar]*

i. **[ciTThii] [likhii] [Shyam=ko] [maar] [Leela=ne]*

The scrambling possibilities in (3a)-(3g) show that the light verb and the main verb *can* be reordered with other clausal constituents as a unit, and the ungrammaticality of (3h) and (3i) shows that they *must* be re-ordered with other clausal constituents as a unit. The data in (3) indicates that the main verb and the light verb in a Hindi CP construction behave as a constituent with respect to scrambling.

Butt (op.cit.) presents two additional kinds of data that suggest that the combination of a main verb and a light verb behaves a single predicate. First, the complement of the light verb cannot be coordinated with main verbs in the standard CP construction, as shown in (4a). Similarly, coordinated main verbs cannot follow light verbs in the reverse CP construction (see (4b)).

(4) a. **Leela=ne Shyam=ko ciTThii likh aur de*
Leela.F=Erg Shyam.M=Dat letter.F.Sg write:MV and give:MV
maar-ii
hit-Perfv.F.Sg:LV
'Leela wrote and gave a letter to Mohan.'

b. **Leela=ne Shyam=ko ciTThii maar*
Leela.F=Erg Shyam.M=Dat letter.F.Sg hit:LV
likh-ii aur di-i
write.Perfv-F.Sg:MV and give.Perfv.F.Sg:MV
'Leela wrote and gave a letter to Mohan.'

The impossibility of coordinating main verbs is not specific to the aspectual CP construction (standard or reverse). It also applies to main verbs (or light verbs) that are followed by (passive, imperfective, or tense) auxiliaries.[2]

(5) a. *nadyaa haar banaa rah-ii*
Nadya.F=Nom necklace.M=Nom make Stat-Perf.F.Sg
t^h-ii aur us-ii vakt pahan rah-ii
be.Past.F.Sg and that-Emph time wear Stat-Perf.F.Sg
t_h-ii
be.Past.F.Sg
'Nadya was making a necklace and wearing it at the same time.'

b. **nadyaa haar [[banaa aur pahin]*
Nadya.F=Nom necklace.M=Nom make and wear
rah-ii t^h-ii]
Stat-Perf.F.Sg be.Past.F.Sg
'Nadya was making a necklace and wearing it at the same time.'

c. **nadyaa [haar [banaa aur haar*
Nadya.F=Nom necklace.M=Nom make and necklace.M=Nom
pahin] rah-ii t^h-ii]
wear Stat-Perf.F.Sg be.Past.F.Sg

[2]Auxiliaries and light verbs show distinct syntactic behaviors with regard to case marking, word order, reduplication, and topicalization.

'Nadya was making a necklace and wearing it at the same time.'

Second, temporal adverbial modifiers such as *kal* (yesterday/tomorrow) can appear in various positions to the left of the reverse CP, as indicated in (6a) and (6b), but not between the main verb and the light verb (6c). Butt (1994:99) provides examples that show that the same to be true of the standard CP construction.

(6) a. *Leela=ne* **kal** *saaraa* *din* *gappõ* *mein* *[maar*
Leela.F=Erg yesterday all day.M chats.M.Pl in hit:LV
bitaay-aa]
spend-Perfv.M.Sg:MV
'Leela spent all day yesterday chatting.'

b. *Leela=ne saaraa din gappõ mein* **kal** *[maar bitaay-aa]*

c. **Leela=ne saaraa din gappõ mein [maar* **kal** *bitaay-aa]*

The fact that main verbs cannot be coordinated when precede or followed by a light verb and no adverbial modifiers can intervene between the light and main verbs is analyzed by Butt (1994) as showing that the two verbs behave as a single predicate. We would rather analyze it as meaning that the combination of a light and main verb is *lite* in the sense of Abeillé and Godard (2002). For reasons of space, we simply outline our analysis of the coordination and adverbial modification data, here:

- Adverbial modifiers like *kal* 'yesterday/tomorrow' are non-*lite* and the combination of a *lite* and non-*lite* constituent is non-*lite*;
- Coordination of *lite* constituents is non-*lite* in Hindi;
- Some phrase-structure constructions in Hindi, in particular the two informally stated in (7) and (8) are sensitive to the "liteness" of their daughters.

(7) $S \rightarrow XP^* \begin{bmatrix}\textsc{weight} & \textit{non-lite}\end{bmatrix} V \begin{bmatrix}\textsc{weight} & \textit{lite}\end{bmatrix}$

(8) $V \begin{bmatrix}\textsc{weight} & \textit{lite} \\ \textsc{head} & \boxed{1}\end{bmatrix} \rightarrow V{*} \begin{bmatrix}\textsc{weight} & \textit{lite}\end{bmatrix} V \begin{bmatrix}\textsc{weight} & \textit{lite} \\ \textsc{head} & \boxed{1}\end{bmatrix}$

The phrase-structure construction informally stated in (7) is almost identical to the constituency assumed by Butt (op. cit.) for Hindi clauses, namely a string of phrases followed by a verbal constituent that consists of a sequence of verbs (main verb followed, optionally, by a *lite* verb and a sequence of auxiliaries). We merely add constraints that require the XPs to be non-lite and verbal constituent to be *lite*. The construction in (8), in turn, licenses a sequence of lite verbs construct to consist of a string of *lite* verbs. The phrase-structure constructions in (7) and (8) together with the first two assumptions we listed above explain the restrictions on coordination and temporal modification presented in Butt (op.cit.) which we just discussed. Coordination and temporal modification make the light verb-main verb combination or the main verb(s) non-*lite*, and therefore unable to participate in the sequence of *lite* verbs licensed by construction (8).

Although the data presented so far suggest that the main and light verb form a *lite* constituent, an alternative hypothesis is that the two verbs combine in the morphology and form some kind of compound. Butt (1994) provides evidence against that hypothesis for the standard CP construction. Discourse clitics such as *hii* (exclusive focus particle 'only') and *bhii* (inclusive focus particle 'also') can be inserted between the verbs in a standard complex predicate construction (pp. 91-93). In the standard CP, in order to take narrow scope over the verb, the emphatic particle must appear between the main verb and the light verb (9b). It cannot appear after the verbal complex, either before (9c) or after an auxiliary (9d).

(9) a. *us=ne* *ciTThii* **bhii** *bhej* *di-yaa*
Pron.3.Sg=Erg letter.F.Sg also send:MV give-Perfv.M.Sg:LV
(t^h-aa)
(be.Past.3.Sg)
'He sent a letter also (along with other things).'

b. *us=ne* *ciTThii* *bhej* **bhii** *di-yaa*
Pron.3.Sg=Erg letter.F.Sg send:MV also give-Perfv.M.Sg:LV
(t^h-aa)
(be.Past.3.Sg)
'He sent a letter (in addition to doing other things).'

c. **us=ne ciTThii bhej di-yaa* **bhii** (t^h-aa)

d. **us=ne ciTThii bhej di-yaa (t^h-aa)* **bhii**

The same pattern that Butt observed for the focus particle *bhii* holds true of a particular negative question construction exemplified below. In the standard CP construction, *wh-* + *neg* marker ('why not') can appear between the main and light verb (10a) but not at the end of the clause (10b).

(10) a. *tum apne beimaan naukar=ko nikaal* ***kyõ nahii***
you self rogue servant=Dat remove:MV why neg
de-te?
give-Impf.M.Sg:LV
'Why don't you remove your rogue servant?' (Nespital 1997:2)

b. **tum apne beimaan naukar=ko nikaal de-te kyõ nahii*?

The restriction on focus particles in the reverse CP is different. Here, *bhii* can only precede the complex predicate (11a) but cannot be inserted between the two verbs (11b) or, as indicated previously, appear at the end of the clause.

(11) a.

us=ne	*ciTThii*	**bhii**	*de*	*bhej-aa*
Pron.3.Sg=Erg	letter.F.Sg	also	give:LV	send-Perfv.M.Sg:MV

‘He also sent off a letter (in addition to doing other things).’

b. **us=ne ciTThii de* **bhii** *bhej-aa*

Since the first predicate in the reverse construction is a light verb, the ungrammaticality of (11) may be semantic, namely the light verb cannot be the scope of the focus particle. Therefore, the fact that *bhii* cannot appear between the two verbs in the reverse construction does not provide evidence for or against the claim that the reverse CP construction involves some kind of compounding.

Finally, while the standard construction can appear with the full range of Hindi auxiliaries (12), the reverse construction is more restricted. Neither the progressive nor the passive auxiliary can appear in a reverse construction, as shown in (13a) and (13b) respectively.[3]

(12)

Shyam=ka	*ghar*	*beech*	*di-yaa*	***jaa***
Shyam.M=Gen	house.M.Sg	sell:MV	give-M.Sg:LV	**go**
rah-aa	*hai*			
stay-Imperfv.M.Sg	be.Pres.3.Sg			

‘Shyam’s house is being sold off.’

(13) a.

**Shyam*	*kitaab*	*jor=se*	*de*	*phekh*
Shyam.M	book.M.Sg	force=Inst	give:LV	throw:MV
rah-aa	*th-aa*			
stay-Imperfv.M.Sg	be.Past-M.3.Sg			

*‘Shyam threw the book forcefully.’

b.

**Kitaab*	*jor=se*	*de*	*phekh-aa*	***ga-yaa***
book.M.Sg	force=Inst	give:LV	throw:MV	**go-M.Sg**
th-aa				
be.Past-M.3.Sg				

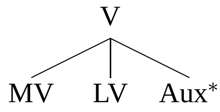
‘The book was thrown forcefully.’

To summarize, constituency tests show that the main and light verbs in the standard and the reverse CP construction form a single V-V constituent (with or without following auxiliaries). The two structures differ in that the reverse construction does not allow the insertion of any element between the two verbs and does not co-occur with the passive or progressive auxiliaries. The two trees below (informally) represent the constituent structure we will hereafter assume for the standard and reverse CP constructs, respectively.

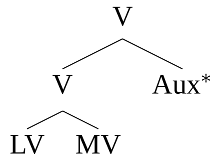
[3] We currently have no cogent explanation for the fact that the reverse complex predicate construction cannot be followed by the passive or the passive + progressive auxiliaries.

(14)

a.



b.



4 Case-marking and subject-verb agreement

The previous section has shown that both the standard and the reverse complex predicate constructions form a V-V constituent. We now present case assignment and subject-verb agreement data that is critical to comparing the Uniformity and Representation Modularity hypotheses. We suggest, based on the government of subject case assignment, that the light verb is the head of that constituent in the standard CP construction (at least when no auxiliary follows) and the main verb is the head of that constituent in the reverse CP construction. We show that the same case assignment constraints that are operative for simple predicate constructions can model case assignment facts for the standard and the reverse construction as well, *but only if* the light verb is the head of the V-V constituent in the standard CP construction, and the main verb in the reverse CP construction.

In this paper, we focus on the alternation between the unmarked and the ergative case on the subject.[4] Hindi is generally considered to have a *split-ergative* case system; the ergative case is aspectually driven . Hindi ergative case can also be assigned to the subject of a semantically defined class of intransitive verbs (Butt and King, 2005; De Hoop and Narasimhan, 2008).

Ergative subject case assignment in *transitive* or *ditransitive* verbs is straightforward. When the verb is in the perfective aspect (marked by the suffix *-(y)aa/ii*), their subjects bear ergative case marking, as illustrated in example (15).[5] In contrast, when the verb is imperfective i.e. either in the habitual aspect (16a) or the future (16b), the subject cannot bear ergative case and is unmarked.

(15) *Shyam=ne ghar=ko banaa-yaa*
Shyam=Erg house=Dat make-Perfv.M.Sg
'Shyam made the house.'

(16) a. *Shyam ghar=ko banaa-taa hai*
Shyam house=Dat make-Impfv be
'Shyam makes the house.'

[4]The unmarked case in Hindi is phonologically null and has been labeled as Nominative by some scholars (Kachru, 1980; Butt, 1994; Butt and King, 2005). However, both proto-agent and proto-patient roles can be unmarked for case and we therefore call it *unmarked*.

[5]In infinitive clauses, the subject is typically assigned dative case, but see Butt and King (2005) for data from the Lahori dialect of Urdu where the subject of infinitive clauses alternates between the ergative and dative case.

b. *Shyam ghar=ko banaa-yeg-aa*
Shyam house=Dat make-Fut-M.Sg
'Shyam will make the house.'

As Kachru (1980:52) points out, volitionality does not play a role in the assignment of ergative case to the subject of transitive verbs in Hindi. Non-volitional verbs such as *bhool* (forget), *kho* (lose), or *jaan* (know) can also select for ergative subjects. Only the verb's aspect marking (perfective) matters.

The assignment of ergative case to the subjects of *intransitive* verbs is more complex. The subject of most intransitive verbs are unmarked for case, as shown by the verb *fisal* (slip) (see (17)); even verbs like *bhaag* (run), *uchal* (jump) or *baith* (sit), where the agent must employ some volition, take only an unmarked and not an ergative subject, as (18) shows. But, some intransitive verbs (called intransitive unergative verbs by Butt and King (2005)) can select either an ergative or an unmarked subject, as (19) illustrates.

(17) *Shyam(*=ne) fisl-aa*
Shyam.M(=Erg) slip-M.Sg
'Shyam slipped.'

(18) *Shyam(*=ne) bhaag-aa*
Shyam.M(=Erg) run-Perfv.M.Sg
'Shyam ran.'

(19) *Shyam(=ne) khaans-aa*
Shyam(=Erg) cough-Perfv.M.Sg
'Shyam coughed (without meaning to).'

Intransitive verbs that can optionally select for an ergative subject are primarily bodily function verbs (including sound emission) verbs such as *khaas* (cough), *chiikh* (sneeze), *bhauk* (bark), *ciik* (scream), *cillaa* (yell), *muut* (urinate), and *thuuk* (spit) (De Hoop and Narasimhan, 2008). But the intransitive verb *nahaa* 'bathe', one of the few Hindi verbs denoting grooming actions (most other grooming actions are expressed via a N+light verb complex predicate), can also take ergative subjects as the attested examples in (20)-(21) show.

(20) ***kissi=ne*** *nahaa-yaa nahii th-aa*
any=Erg bathe-M.Sg neg be.Past-3.Sg
'Nobody had bathed.'

(21) *ghar aa-kar nal=ke niichee saabun=se malmal-kar*
home come-do tap=Gen below soap=Inst scrub.scrub-do
ek-ek=ne *nahaa-yaa*
one-one=Erg bathe-M.Sg
'Upon coming home, each one bathed under the tap by scrubbing (hard) with soap.'

One frequent analysis of ergative case assignment to intransitive verbs is that ergative case indicates conscious control or choice that the subject's referent exerts over the action (see Mohanan, 1994; Butt and King, 2002). Under this analysis, ergative case on the subject of intransitive verbs indicates that the action is within the internal control of the subject's referent. Several attested corpus examples (cross-checked with consultants) suggest that this analysis is incorrect. Consider the following example, where it is very doubtful that the dog made a conscious choice not to bark.

(22)

court	*mein*	*bahut*	*log*	*moujuud*	*th-ee*	*phir bhii*	*kiisii*	*par*
court	in	many	people	present	be-Past.3.Pl	still	any	on

bhii	***kuttee=ne***	*bhauunk-aa*	*tak*	*nahii*
also	**dog=Erg**	bark-M.Sg	even	neg

'Many people were present in court but still the dog did not even bark at anyone.'

Example (22) and similar corpus examples suggest an alternative hypothesis, which for lack of space we state here without further justification. Ergative marking on intransitive verbs describing bodily functions (including sound emission verbs) indicates that the property expressed by the sentence minus its subject runs counter to expectations given the subject's denotation. For example, it is unexpected for a dog not to bark in the situational context of (22).

The above facts show that the assignment of ergative case to the subject can be captured by the following constraints:

(23) *Default Unmarked Constraint*: By default, the subject is unmarked.

(24) *Transitive Perfective Constraint*: If the verb is transitive and perfective, then the subject is assigned ergative case.

(25) *Contrary to Expectation Constraint*: If the verb is intransitive and perfective, denotes a bodily function, and the subject is assigned ergative case, then the action is unexpected given the actor.[6]

Let us now turn to case assignment in the CP constructions. As indicated previously, the same case assignment constraints that operate on single predicates can model the case assignment facts in the CP constructions. Previous research on the standard CP construction has argued that the light verb always assigns case to the subject (Butt, 1994): The subject must be ergative if the light verb is transitive, and nominative (unmarked in our terminology) if the light verb is intransitive. For instance, although the main verb *gaa* (sing) is transitive in both (26) and (27), the subject is only assigned ergative case in (26). This is because the light

[6]In Sinhala, another Indo-Aryan language, the selection of ergative case for the subjects of involitive verbs is correlated with whether or not the event was supposed to be intentional Inman (1994). Also see Malchukov (2008) for similar data from unrelated languages.

verb *Daal* ‘put’ is transitive whereas the light verb *paD* ‘fall’ is intransitive. (The (in)transitivity of the light verb itself is an idiosyncratic property of light verbs that is a carry-over from their main verb usage, as, semantically, both *Daal* ‘put’ and *paD* ‘fall’ are (monadic) aspectual functors.) A similar pattern is illustrated in the contrast between (28) and (29) for the main verb *ciikh* (scream). The subject is unmarked if the light verb is intransitive (28) and is assigned ergative case if the light verb is transitive (29). Finally, note that among intransitive verbs, only verbs denoting bodily function can appear in the standard CP construction (a restriction we explain below). That the assignment of ergative case depends on the transitivity of the light verb in the standard CP construction is explained by the *Transitive Perfective Constraint*, provided the light verb governs case assignment in that construction.

(26) *Ram=ne* *gaanaa* *gaa* *Daal-aa*
Ram.M=Erg song sing:MV put-M.Sg:LV
‘Ram sang a song (had to).’

(27) *Ram* *gaanaa* *gaa* *paD-aa*
Ram.M song sing:MV fall-M.Sg:LV
‘Ram sang a song (without wanting to).’

(28) *Ram* *ciikh* *paD-aa*
Ram.M scream:MV fall-M.Sg:LV
‘Ram screamed suddenly.’

(29) *Ram=ne* *ciikh* *Daal-aa*
Ram=Erg scream:MV put-M.Sg:LV
‘Ram screamed violently.’

Different conditions on the assignment of ergative case apply to the reverse construction. Here it is properties of the main verb that governs assignment of ergative case. For instance, even though the light verb *de* ‘give’ is transitive, the subject in (30) is unmarked for case, because the main verb *bhaag* (run) is intransitive. Conversely, when the intransitive light verb *jaa* ‘go’ in (31) combines with the transitive main verb *beech* ‘sell’ to form a reverse CP construction, the complex predicate selects for an ergative subject. In both (30) and (31), then, the transitivity of the main verb, not the transitivity of the light verb, determines the assignment of ergative (vs. unmarked) case to the subject.[7]

(30) *Ram* *de* *bhaag-aa*
Ram.M give:LV run-M.Sg:MV
‘Ram ran (rapidly).’

[7]Note that bodily function verbs do not seem to be able to appear in the reverse CP construction; we have no explanation for this restriction.

(31) *Ram=ne* *apnaa* *makaan* *jaa* *beech-aa*
Ram.M=Erg self house go:LV sell-M.Sg:MV
'Ram sold his house.'

The summary of case assignment patterns in Hindi aspectual CP constructions is as follows. While the transitivity of the light verb determines the presence of ergative case on the subject in the standard CP construction, it is the transitivity of the main verb that determines the presence of ergative case on the subject in the reverse CP construction. Case assignment in Hindi complex predicate constructions is therefore position-dependent, i.e. it is determined by the transitivity of the last verb of the complex predicate.

Subject-verb agreement data provide additional support for the claim that the main verb is the head of the construction in the reverse CP construction and the light verb in the standard CP construction. Hindi verbs agree with the highest unmarked argument in number and gender. In a single predicate construction, the finite verb agrees with the subject if it is unmarked (32a). If the subject is marked for case, the verb instead agrees with the object if it is unmarked, as shown in (32b) and (32c). When there is no unmarked argument in the clause, the verb receives a default masculine singular inflection (32d).

(32) a. *Leela* *ghar* *aa-t-**ii*** *hai*
Leela.**F** home.M.Sg come-Pres-**F.Sg** be.Pres.3.Sg
'Leela comes home.'

b. *Leela=ne* *ghar* *khariid-**aa***
Leela.F=Erg house.**M.Sg** buy-**M.Sg**
'Leela bought a house.'

c. *Leela=ne* *gaadii* *khariid-**ii***
Leela.F=Erg vehicle.**F.Sg** buy-**F.Sg**
'Leela bought a vehicle.'

d. *Leela=ne* *gaadii=ko* *beech-**aa***
Leela.F=Erg vehicle.F.Sg=Dat sell-**M.Sg**
'Leela sold the vehicle.'

In the standard and reverse aspectual CP constructions as well, the finite verb agrees with the unmarked argument. As shown below, the light verb in the standard construction agrees with the subject if the subject is unmarked (33a) or with the object if the subject is overtly marked for case, as shown in (33b) and (33c).

(33) a. *baaz* *parinde=par* *jhapaT* *gay-**aa***
eagle.**M.Sg** bird.M.Sg=Loc swoop:MV go-**M.Sg**:LV
'The eagle swooped on the bird.'

b. *Leela=ne* *Shyam=ko* *xat* *likh*
Leela.F=Erg Shyam.M.Sg=Dat letter.**M.Sg** write:MV
*maar-**aa***
hit-**F.Sg**:LV
'Leela wrote a letter to Shyam (hurriedly).'

c. *Leela=ne* *Shyam=ko* *ciTThii* *likh* *maar-**ii***
Leela.F=Erg Shyam.M.Sg=Dat letter.**F.Sg** write:MV hit-**F.Sg**:LV
'Leela wrote a letter to Shyam (hurriedly).'

The unmarked subject NP in (33a) is masculine and therefore, the light verb is assigned masculine gender *ga-yaa* (go) instead of feminine *ga-yii*. When the subject is marked for case, the verb *maar* (hit) agrees with the unmarked direct object in (33b) and (33c). In (33b), the finite verb is inflected for masculine gender since the direct object *xat* (letter) is masculine and similarly, the finite verb in (33c) is inflected for feminine gender since *ciTThii* (letter) is feminine.

In the reverse CP construction, it is the main verb that agrees with the highest unmarked argument, the subject in (34a) and the object in (34b) and (34c). In (34b), the main verb *likh* (write) is inflected for masculine gender since the highest unmarked NP, the object *xat* (letter), is masculine and similarly in (34c) *likh* is inflected for feminine gender since the object *ciTThi* (letter) is feminine. Overall, the examples in (33) and (34) show that the last verb in the complex predicate, irrespective of whether it is the light verb or the main verb, agrees with the subject.

(34) a. ***baaz*** *parinde=par* *de* *jhapt-**aa***
eagle.M.Sg bird.M.Sg=Loc give:LV swoop-**M.Sg**:MV
'The eagle swooped on the bird (forcefully).'

b. *Leela=ne* *Shyam=ko* ***xat*** *maar* *likh-**aa***
Leela.F=Erg Shyam.M=Dat letter.**M.Sg** hit:LV write-**M.Sg**:MV
'Leela wrote a letter to Shyam (hurriedly).'

c. *Leela=ne* *Shyam=ko* ***citthii*** *maar* *likh-**ii***
Leela.F=Erg Shyam.M=Dat letter.**F.Sg** hit:LV write-**F.Sg**:MV
'Leela wrote a letter to Shyam (hurriedly).'

5 Hindi CP constructions and the Uniformity vs. Representational Modularity hypotheses

Let us now come back to the issue we started with, namely how uniformally isomorphic the semantic and syntactic structures of Hindi aspectual markers truly are. The properties of heads are a critical determinant of case across languages; similarly, agreement is another relation between heads and their dependents. Therefore, the fact that the assignment of ergative case or subject-verb agreement is determined by the properties of the main verb in the reverse CP construction and the light verb in the standard CP construction indicate a difference in headedness between the two constructions. The light verb is the head in the standard CP construction and the main verb is the head in the reverse CP construction. It is generally assumed in the kind of syntactic approach Cinque proposes that agreement is a relation between heads and their specifiers and, in the Minimalist framework of Chomsky (1995), checking of case features is also predicated on the presence of a head-specifier relation. The difference in case assignment and agreement between the standard and reverse CP construction therefore strongly supports the hypothesis that the light verb is the head of the standard CP construction and the main verb is the head of the reverse CP construction. But such an hypothesis is hard to reconcile with the Uniformity Hypothesis, which posits that there is a uniform set of aspectual functional heads across languages and within languages. If the light verb is an aspectual functional head in the standard CP that takes the main verb as its complement^{+} (i.e., as a complement of a complement of a complement . . .), as Cinque's Uniformity Hypothesis would predict, it should also be an aspectual functional head that takes the main verb as its complement^{+} in the reverse CP. After all, both constructions express the same perfective semantics. There are some minor, hard to pin down subtle semantic differences between the standard and reverse CP constructions, but none that would affect the respective geometry of the relevant functional heads and main verbs.

At this point, we can imagine two possible solutions to this quandary. First, one could explore the possibility that, even though the light verb is still a functional aspectual head higher than the main verb in the reverse CP, it is the main verb that "counts" as a head for ergative case assignment and subject-verb agreement. We do not presently know of any independent motivation for such a claim (which, of course, could reflect our lack of imagination). Leftward movement of a verb, for example, does not typically affect the head status of the functional heads it moves to the left of. Second, one could treat the light verb-main verb combination in the reverse CP construction as being an instance of compounding (since we do not know of any marker than can appear between the light verb and the main verb in the reverse CP construction) and exempt compounding from the purview of the Uniformity Hypothesis. This line of inquiry seems even less appealing to us, as the relative productivity of the reverse CP construction makes it hard to see how one would distinguish the kind of compounding purportedly present in reverse CP constructs from true VV syntactic combinations. More importantly, exempting compounding from the purview of the Uniformity Hypothesis greatly weakens it, and would run counter to its current scope, as it is standardly assumed that suffixal tense is the expression of a higher functional T head. We take this admittedly cursory discussion to suggest that the Hindi facts present challenges to the Uniformity hypothesis, although a firm conclusion must await a more thorough discussion. In what follows, we show that the Representational Modularity Hypothesis and the approach taken in Koenig and Muansuwan (2005) for Thai provide a straightforward model of the two Hindi aspectual CP constructions.

The ergative/unmarked alternation is captured by the rules in (35)-(40). As discussed previously, the default case value is unmarked.

(35) *Default Unmarked Constraint*: $\left[\text{CASE} \quad /\textit{unmarked}\right]$

The default in (35) is overriden when either of the other two case assignment constraints apply. The *Transitive Perfective constraint* requires us to define transitivity, which we define here not in terms of properties of the ARG-ST list (its inclusion of two NP *synsem* descriptions), but rather in terms of the attribute/value pair [TRANS +]. We have two reasons to define transitivity in terms of such an attribute/value pair rather than directly in terms of ARG-ST membership. First, as we mentioned above, the constraint must apply to "transitive" light verbs whose ARG-ST need not include two NP descriptions (as when a "transitive" light verb combines with an intransitive main verb, but can still be "transitive" as an idiosyncratic property left over from their main verb uses. Second, treating transitivity as a feature is useful to model the positional nature of ergative assignment within the sequence of main verb, light verb, and auxiliaries. We have suggested above that in the standard CP, the transitivity of the light verb determines the assignment of ergative case to the subject. This is true when no auxiliary follows the light verb (as in 26)-(29). But, matters are more complex when auxiliaries follow the light verb. When the passive (36), or passive and imperfective auxiliaries together (37) follow a transitive light verb, the subject remains unmarked. In contrast, when the tense auxiliary follows a transitive light verb, the subject bears ergative case, just as when no auxiliary is present, as shown in (38).

(36)	*Shyam=ka*	*ghar*	*beech*	*di-yaa*	***ga-yaa***
	Shyam.M=Gen	house.M.Sg	sell:MV	give-M.Sg:LV	go-M.Sg

'Shyam's house is being sold off.'

(37)	*Shyam=ka*	*ghar*	*beech*	*di-yaa*	***jaa***
	Shyam.M=Gen	house.M.Sg	sell:MV	give-M.Sg:LV	**go**
	rah-aa	*hai*			
	stay-Imperfv.M.Sg	be.Pres.3.Sg			

'Shyam's house is being sold off.'

(38)	*Shyam=**ne***	*ghar*	*beech*	*di-yaa*	***hai***
	Shyam.M=Erg	house.M.Sg	sell:MV	give-M.Sg:LV	be.Pres.3.Sg

'Shyam has sold the house.'

It is rather straightforward to explain why the passive and imperfective do not license ergative case assignment as these auxiliaries are not transitive and perfective. The behavior of the tense auxiliary is more complex, as it seems "transparent" to the transitivity and perfectivity of the auxiliary that precedes it. When the tense auxiliary follows a transitive light verb, the clause's subject bears ergative case, but when it follows the passive or progressive auxiliaries, it does not. To model this rather complex set of facts, we make the following assumptions:

- Ergative case assignment to the subject of "transitive" verbs applies to all verbs that bear the head properties $\begin{bmatrix}\text{TRANS} & + \\ \text{ASP} & \textit{perf}\end{bmatrix}$;

- The value of the TRANS and PERF attributes of the tense auxiliary are identical to the values of its verbal complement;

- Each verb in the verb complex sequence licensed by construction (8) include the argument structure of the preceding verb in its argument structure, i.e., induces argument composition. This constraint does not apply to the combination of the light verb and main verb in the reverse CP construction, as such combinations are not licensed by the construction in (8), but by a modifier-head construction (see below);

Based on the above discussion, the *Transitive Perfective Constraint* in (24) is modeled as follows. (We use the relational constraint *last-member* to select the last daughter of the sequence of verbs licensed by the construction (informally) represented in (8).) Note that the aspectual value of the verb is treated as a head feature since it affects verbal morphology.

(39) *Transitive Perfective Constraint*

$$\begin{bmatrix}\textit{verb-complex-cx} \\ \text{DTRS} \quad \boxed{1}\end{bmatrix} \wedge \text{LAST-MEMBER}(\boxed{1}, \boxed{2}) \wedge \begin{bmatrix}\text{HEAD} & \begin{bmatrix}\text{TRANS} & + \\ \text{ASP} & \textit{perf}\end{bmatrix}\end{bmatrix}$$

$$\Rightarrow \begin{bmatrix}\text{ARG-ST} & \left\langle \text{NP}\begin{bmatrix}\text{CASE} & \textit{erg}\end{bmatrix}, \ldots \right\rangle\end{bmatrix}$$

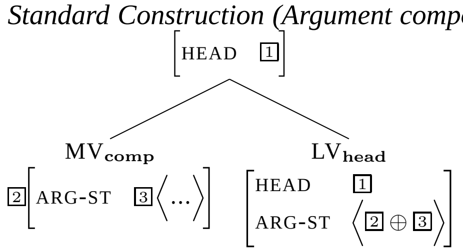
The assignment of ergative case to the subject of intransitive verbs i.e., the *Counter to Expectation Constraint* in (25) is more complex. It applies only to a small semantically-defined subset of intransitive verbs and requires that the conversational background support the contention that the bodily function is counter-to-expectation for the subject's referent.

(40) *Counter to Expectation Constraint*

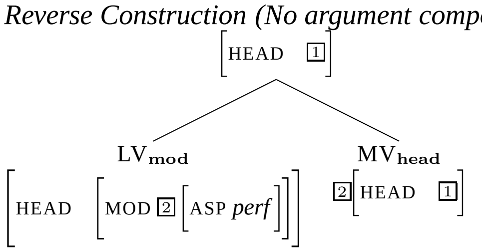
$$\begin{bmatrix}\textit{iv-lxm} & \\ \text{HEAD} & \begin{bmatrix}\text{ASP} & \begin{bmatrix}\textit{perf}\end{bmatrix}\end{bmatrix} \\ \text{ARG-ST} & \left\langle \text{NP}_{\boxed{1}}\begin{bmatrix}\text{CASE} & \textit{erg}\end{bmatrix} \right\rangle \\ \text{SEM} & \begin{bmatrix}\text{RELS} & \boxed{3}\begin{bmatrix}\textit{bodily-function-rel} & \\ \text{EVENT} & \boxed{2} \\ \text{ARG} & \boxed{1}\end{bmatrix}\end{bmatrix}\end{bmatrix} \Rightarrow \begin{bmatrix}\text{BGRND} & \left\{\begin{bmatrix}\textit{counter-expect-rel} & \\ \text{EVENT} & y \\ \text{ARG1} & \boxed{1} \\ \text{ARG2} & \boxed{2}\end{bmatrix}\right\}\end{bmatrix}$$

We have now implemented the basic ergative case assignment constraints for Hindi. Crucially, the same rules model the assignment of subject case in single verb clauses as well as (standard and reverse) complex predicate constructions. To model the difference in headedness between the standard and the reverse CP constructions, we propose that only the standard complex predicate construction involves argument composition; the reverse complex predicate construction involves a head-modifier structure. Within HPSG, constructions similar to the standard CP have been analyzed as involving an operation of *argument composition* wherein the light verb is considered an operator that subcategorizes for the main verb, and its argument structure also includes what its complement verb subcategorizes for (cf. Hinrichs and Nakasawa (1994) for German, or Abeillé and Godard (2002) for Romance complex predicates). We suggest that an argument composition analysis is also appropriate for the standard aspectual CP construction in Hindi. This is illustrated in the abbreviated phrase structure tree in (41).

(41) *Standard Construction (Argument composition)*



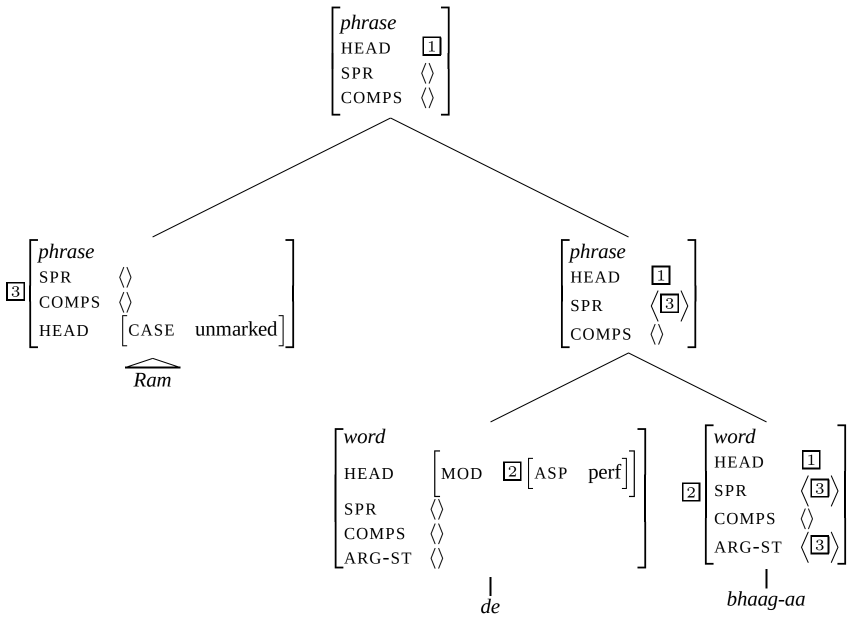
(42) *Reverse Construction (No argument composition)*



Note that our argument-composition analysis of the standard CP construction accounts for the fact that main verbs that do not denote bodily functions cannot combine with transitive light verbs in the standard CP construction. We assume that only verb whose subject can alternate between ergative and unmarked case do not lexically specify their case value. Since the subject of intransitive verbs that do not denote bodily functions never alternate, their case value is *strictly unmarked*. Since light verbs compose their argument structure with that of their verbal complement, the unmarked case value of this intransitive verb would clash with the ergative value that a transitive light verb would require.

In the reverse CP construction, on the other hand, the main verb is the syntactic head because it assigns case to the subject and agrees with the highest unmarked argument. Furthermore, argument selection in Hindi, a head-final language, takes place from right to left as shown in (41); i.e., the light verb would be expected to follow the main verb if it were the head of the reverse construction). We therefore need a different mechanism to account for the light verb and main verb combination. We analyze light verbs in the reverse construction as modifiers that take what they modify as arguments, since modifiers (e.g., adjectives or adverbs) in Hindi typically precede the expressions that they modify (Kachru, 1980). We model the modifier status of the light verbs in the reverse construction as shown in (42). The reverse CP construction exemplified in (30) is modeled in (43). Here the subject Ram appears only on the specifier and argument-structure list of the main verb, as there is no argument composition in the reverse construction. The light verb *de* 'give' modifies the head of the phrase, the main verb *bhaag* 'run', which determines the subject's case. Crucially, the non-null value of the MOD feature indicates that the light verb cannot be the head of the construction thus ensuring that it cannot assign case to the subject in spite of being the clause's semantic head.

(43)



Treating the light verb-main verb combination in the reverse CP as an instance of modifier/head combination makes for an interesting parallel between Hindi and Thai. Both languages involve the same two possible structures for the expression of aspect (aspectual verbs heading a head-complement structure and aspectual verbs modifying a main verb). The difference between the two languages reduces to whether the complement or modified verb is a VP (Thai) or a V (Hindi) and parallels the difference between serial verb constructions that involve sequences of VPs or sequences of V discussed in Andrews and Manning (1999).[8]

[8]It should be noted that, although our analysis of the reverse complex predicate construction accounts for all the data we are aware of, other analysis are possible, as reviewers pointed out to us. One may analyze the reverse CP construction via the kind of type-raising analysis proposed in Kim and Sag (2002) for French postverbal negation *pas* (and other similar functors). In a nutshell,

6 Conclusion

This paper has made several contributions. First, we described a complex set of ergative case assignment constraints in Hindi and their interaction with aspectual complex predicate constructions. We suggested that conscious control is not the appropriate information contributed by ergative case for verbs denoting bodily functions, and provided evidence that the last verb in the sequence of *lite* verbs assigns case (ergative, in particular), and, finally, we showed that it is the main verb, not the light verb, that governs ergative case assignment in the reverse CP construction. Second, we argued that this last fact, as well as corroborating subject-verb agreement data support the claim that the head of the standard CP construction is the light verb, but the head of the reverse CP construction is the main verb. Third, we argued that the fact that case-marking is "positional", supports the conclusion that the mapping between aspectual semantics and syntactic structure need not be uniform within a language, an argument similar to the one presented in Koenig and Muansuwan (2005) for Thai. Such data present a challenge to the hypothesis (such as in Cinque (1999)) that the semantic structure of aspectual functors is almost isomorphic to the syntactic structures that express them. On the other hand, a framework such as HPSG that distinguishes between syntactic and semantic heads and allows for semantic and syntactic information to be partially dissociated can easily model these facts. Finally, we presented an HSPG analysis of the Hindi ergative case assignment constraints as well as of the standard and reverse CP constructions. Clearly, more work is needed, but the intriguing parallels between the syntax of aspect in Hindi and Thai suggest that aspectual verbs can be either heads or modifiers and this split can occur within the same language.

References

Abeillé, A. and D. Godard (2002). The Syntactic Structure of French Auxiliaries. *Language 78*(3), 404–452.

Andrews, A. and C. Manning (1999). *Complex Predicates and Information Spreading in LFG* (First ed.). Stanford, CA: Stanford Monographs in Linguistics.

such an analysis would treat the light verb in the reverse construction as a syntactic complement of the main verb despite the fact that it semantically takes scope over the main verb. Our main reason for not exploring this kind of analysis is motivated by the parallel between the syntax of aspect in Hindi and Thai, given that the corresponding Thai structure cannot involve type-raising. A head (main verb)/marker (light verb) structure is also in principle possible. The fact that the light verb is semantically a functor that selects the index of the main verb as argument in a way similar to what verbal modifiers do and the fact that light verbs in the reverse CP construction do not select for a particular form of the main verb makes it quite different from the kind of words classified as markers in, say, Pollard and Sag (1994). None of our arguments against these other possible analysis are unsurmountable, probably a testimony to the fact that deciding which head-non-head construction is involved outside of standard head-complement/head-subject structure is not easy, and possibly not critical within HPSG.

Butt, M. and T. H. King (2005). The Status of Case. In V. Dayal and A. Mahajan (Eds.), *Clause Structure in South Asian Languages* (First ed.). Kluwer Academic Publishers.

Butt, M. J. (1994). *The Structure of Complex Predicates in Urdu*. Ph. D. thesis, Stanford University, Stanford, CA.

Bybee, J., R. Perkins, and W. Pagliuca (1994). *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World* (2 ed.). Chicago: University of Chicago Press.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Cinque, G. (1999). *Adverbs and Functional Heads: A Cross-linguistics Perspective* (1 ed.). Oxford, UK: Oxford University Press.

Copestake, A., D. Flickinger, C. Pollard, and I. Sag (2005). Minimal Recursion Semantics: An Introduction. *Research on Language and Computation 3*, 281332.

De Hoop, H. and B. Narasimhan (2008). Ergative case-marking in Hindi. In H. de Hoop and P. de Swart (Eds.), *Differential subject marking*, pp. 63–78. Springer.

Hinrichs, E. and T. Nakasawa (1994). Linearizing AUX's in German Verbal Complexes. In *German in Head-Driven Phrase Structure Grammar*, Number 46, pp. 11–37. Stanford, CA: CSLI Lecture Notes.

Hook, P. (1975). *Compound Verbs in Hindi*. Ph. D. thesis, University of Michigan.

Inman, M. V. (1994). *Semantics and Pragmatics of Colloquial Sinhala involitive verbs*. Ph. D. thesis, Stanford University.

Kachru, Y. (1980). *Aspects of Hindi Grammar*. New Delhi, India: Manohar Publishers and Distributors.

Kim, J.-B. and I. A. Sag (2002). Negation without Head-Movement. *Natural Language and Linguistic Theory 20*(2), 339–412.

Koenig, J.-P. and N. Muansuwan (2005). The Syntax of Aspect in Thai. *Natural Language and Linguistic Theory 23*(2), 335–380.

Malchukov, A. L. (2008). Animacy and asymmetries in differential case marking. *Lingua 118*, 203–221.

Pollard, C. and I. Sag (1994). *Head-driven Phrase Structure Grammar*. Stanford: CSLI Publications.

Poornima, S. and J.-P. Koenig (2008). Reverse Complex Predicates in Hindi. In S. Moran, D. Tanner, and M. Scanlon (Eds.), *Proceedings of the 24th Northwest Linguistics Conference*, Volume 27, Seattle, WA, pp. 17–26.

Phraseological Clauses in Constructional HPSG

Frank Richter and Manfred Sailer

University of Tübingen and University of Göttingen

Proceedings of the HPSG09 Conference

Georg-August-Universität Göttingen, Germany

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

In this paper we investigate German idioms which contain phraseologically fixed clauses (PCl). To provide a comprehensive HPSG theory of PCls we extend the idiom theory of Soehn (2006) in such a way that it can distinguish different degrees of regularity in idiomatic expressions. An in-depth analysis of two characteristic PCls shows how our two-dimensional theory of idiomatic expressions can be applied and illustrates the scope of the theory.

1 Introduction

The literature on idioms often focuses on VP idioms such as *kick the bucket* or *spill the beans*, where a particular verbal lexeme combines with a particular NP or PP complement. These combinations show different degrees of flexibility. Hardly any attention has been paid to idioms which comprise complete clauses. Idioms with phraseological clauses are mentioned in passim in phraseological studies such as Fleischer (1997) but have never been in the focus of empirical studies, or of detailed theoretical discussions. As clausal parts of idioms are structurally more complex than NPs or PPs, they are ideally suited for investigating a greater range of structural and semantic variation in idiomatic expressions.

In this paper we will look at phraseologically fixed clauses (PCl) in German. The discussion of PCls is particularly interesting in light of attempts to combine aspects of Construction Grammar with HPSG. One of the important insights of Construction Grammar is that constructions may span more than a local tree. This contrasts with the lexical nature of HPSG and its historical ties to context-free phrase structure grammars. Another result of phraseological research is the insight that idioms are not all of the same kind. In the context of PCls we will want to distinguish between decomposable and non-decomposable idioms (Wasow et al., 1983), and between grammatical and extra-grammatical idioms (Fillmore et al., 1988).

We start with a presentation of the empirical properties of German PCls (Section 2). In order to get the necessary theoretical tools for their analysis, we extend an existing approach to idioms in HPSG to be able to distinguish in our theory between different types of idioms (Section 3). In Section 4 the theory is applied to the PCl data. Section 5 contains a short comparison to an alternative attempt to formalize basic ideas of Construction Grammar in an HPSG-inspired framework. A short summary and conclusion are given at the end of the paper.

2 Data

In (1) and (2) we list idioms with phraseological clauses (PCl). Each PCl in (1) combines with a particular verb or a small group of verbs. Their behavior resembles

†We would like to thank Ivan Sag for the lively and inspiring discussions at HPSG’09 in Göttingen, which led to numerous improvements of our theory. Thank you to Janina Radó for proofreading.

the behavior of the word *headway* in the English expression *make headway*, i.e. they act as a complement in a VP idiom where both the verb and the complement are part of the idiom. The PCls are declarative clauses ((1-c), (1-f)), interrogative clauses ((1-a), (1-b), (1-d)), and a free relative clause in (1-e). The PCls in (2) are adjunct clauses.

(1) PCl is a complement clause to one or a small group of verbs

a. wissen, wo Barthel den Most holt
know where Barthel the young wine gets
('know every trick in the book')

b. (nicht) wissen, wo X_dat der Kopf steht
not know where X the head stands
('have a lot of stress')

c. glauben, X_acc tritt ein Pferd
believe X kicks a horse
('be very surprised')

d. wissen, wo (X_acc/dat) der Schuh drückt
know where X the shoe presses
('know what is worrying X')

e. hingehen/ bleiben (sollen), wo der Pfeffer wächst
go/ stay (should) where the pepper grows
('go/ stay away')

f. glauben, X's Schwein pfeift
believe X's pig whistles
('be very surprised')

(2) PCl is an adjunct

a. bis der Arzt kommt
until the doctor arrives
('ad nauseam')

b. wenn Ostern und Pfingsten auf einen/ denselben Tag fallen
when Eastern and Pentecost on one/ the same day fall
('never')

c. aussehen, als hätten X_dat die Hühner das Brot weggefressen
look as if had X the chicken the bread eaten away
('look stupefied')

d. wie Gott X_acc geschaffen hat
as god X created has
('naked')

Apart from their idiomatic semantics, the PCls have the structural properties of regular German sentences. On closer scrutiny, they display an interesting continuum of grammatical and lexical fixedness and flexibility.

In (1-b), (1-c), (1-f), (2-c), and (2-d) the constituent marked with X is anaphoric to the matrix subject. In (1-d) the constituent marked with X is optional and need not be anaphoric to the matrix subject.

(3) Ich möchte wissen, wo (dich/dir) der Schuh drückt.
(lit.: I want to know where the shoe presses you)

PCls permit a certain degree of grammatical variation. Speakers of some German dialects prefer to use proper nouns with definite articles. These speakers use a variant of (1-a) with a PCl subject of the form *der Barthel* (*the Barthel*). Similarly, *until*-clauses in German may optionally contain an overt complementizer *dass* (*that*). Indeed, a variant of (2-a) with an overt complementizer is attested, i.e. *bis dass der Arzt kommt* (*until that the doctor arrives*).

However, not just any grammatical variation is permitted. Let us consider the idiom in (1-b). Outside of idiomatic phrases a combination of a possessive dative NP and a definite NP can be freely replaced with a construction with the same dative NP and a definite NP that contains a possessive determiner. The possessor is then coreferential with the dative NP. The pattern is illustrated in (4-a). This otherwise systematic variation is not possible with the idiom. We use "#" to indicate the non-availability of an idiomatic interpretation. The same alternation is also excluded for (2-c).

(4) a. Ich habe Peter den/seinen Kopf verbunden.
(lit: 'I bandaged Peter the/his head')
b. Peter weiß nicht, wo ihm der/#sein Kopf steht.

Another systematic variation is the active-passive alternation. None of the PCls with a transitive verb in (1) allow a passive in their idiomatic meaning.

(5) a. #wissen, wo vom Barthel der Most geholt wird (passive of (1-a))
b. #wissen, wo X vom Schuh gedrückt wird (passive of (1-d))
c. #glauben, X wird von einem Pferd getreten (passive of (1-c))

Finally, the PCl in (1-c) is a verb-second clause. In free uses, we can find two kinds of alternation. First, verb-second complement clauses alternate with verb-final complement clauses. Second, any constituent of the clause can occur as the first constituent (in the *Vorfeld*) in verb-second clauses without a change in meaning. Both types of grammatical alternation are excluded in (1-c).

(6) a. #Ich glaube, dass mich ein Pferd tritt. (*dass*-clause)
b. #Ich glaube, ein Pferd tritt mich. (different first constituent)

All alternations discussed in (4)–(6) are neutral with respect to the truth-conditional semantics of the literal reading of the PCls, but some of the alternations influence the information structure. Among the latter kind are valence alternations, clause type alternation, and constituent fronting. This subclass of alternations is impossible with PCls. The permitted alternations, viz. *dass* insertion and the insertion of a definite determiner in front of a proper name, do not affect information structure.

Let us, next, turn to variation at the level of changing or adding lexical material. Some lexical variation is clearly permitted. In (2-b) the holidays can be changed, the subject may be any combination of Easter, Pentecost, and Christmas. However, some form of the verb *(zusammen-) fallen* is obligatory.

(7) #wenn Ostern und Pfingsten auf demselben Tag liegen/ zu liegen kommen/ am selben Tag sind.

We also find some variation with respect to the matrix predicate that the PCl occurs with. The expression in (2-c) may combine with any matrix predicate that describes someone's facial expression or someone's appearance. The same variation is also found with other PCls that express a similar content. They all are comments on the way the referent of the matrix subject looks.

(8) aussehen/ ein Gesicht machen/ dastehen,
look/ a face make/ appear ...
  a. als hätten X die Hühner das Brot weggefressen. (=(2-c))
  b. als hätte es X die Ernte verhagelt
     lit. look as if X's harvest was destroyed by hail.
  c. als hätte X ein Lineal verschluckt.
     lit. as if X had swallowed a ruler
  d. wie eine Kuh, wenn's donnert.
     lit. like a cow when it thunders

There is also some systematic variation in the matrix predicates that express the idea of "thinking". All PCls that occur with *glauben* (*believe*) in the present tense also marginally accept *denken* (*think*) in the present tense. In past tense, however, they systematically prefer *denken*.

(9) Ich glaube/ ?denke/ ?*glaubte/ dachte ...
I believe.PRES/ think.PRES/ believed.PAST/ thought.PAST
  a. mich tritt ein Pferd. (=(1-c))
  b. mein Hamster bohnert. (lit.: my hamster is polishing the floor)
  c. ich steh im Wald. (lit.: I am standing in the woods)

In addition to the variation of obligatory material, some PCls may host more lexical or semantic material. For example, there is variation in the tense form of some but not all PCls.

(10) temporally flexible idioms
  a. Ich hab damals Tetris gespielt, bis der Arzt gekommen ist.
     (pres. perfect of (2-a))
     ('I used to play Tetris ad nausam')
     (www.stern.de/digital/computer/scheibe/
     scheibes-kolumne-pc-nostalgie-619141.html, 14.10.2009.)
  b. Er wusste nicht, wo ihm der Kopf stand. (simple past of (1-b))

(11) temporally fixed idioms

a. #Sie hat nicht gewusst, wo Barthel den Most geholt hat. (pres. perf. of (1-a))

b. #Ich glaube, mein Schwein hat gepfiffen. (pres. perf. of (1-f))

Similarly, modals are allowed in some but not all of the PCls.

(12) modally flexible idioms

a. Hudezeck versteht sich auf die Kunst, die Lachmuskeln so zu strapazieren, bis der Arzt kommen muss. ((2-a) with *must*)
(www.fnp.de/fnp/mobil/rmn01.c.5440589.de.htm, 14.10.2009.)

b. Als Reiseleiter ist Terje ein Mann der Praxis und weiß, wann und wo auf Reisen der Schuh drücken könnte. ((1-d) with *could*)
(www.skantur.de/allgemein/award.htm, 14.10.2009.)

(13) modally fixed idioms

a. #Peter soll bleiben, wo der Pfeffer wachsen kann. ((1-e) with *should*)

b. #Ich glaube, mein Schwein könnte pfeifen. ((1-f) with *could*)

PCls do not tolerate negation (see (14)), but non-truth-conditional modifiers such as *eigentlich* (*actually*), *sprichwörtlich* (*proverbial*) can usually be added (see (15)).

(14) a. #Peter weiß, wo ihn der Schuh nicht drückt. ((1-d) with negation)

b. #Peter weiß, wo Barthel den Most nicht holt. ((1-a) with negation)

c. #wenn Ostern und Pfingsten nicht auf einen Tag fallen ((2-b) with negation)

(15) a. Peter weiß nicht mehr, wo ihm eigentlich der Kopf steht. ((1-b) with *actually*)

b. Martha weiß, wo Barthel den sprichwörtlichen Most holt. ((1-a) with *proverbial*)

Focus sensitive particles such as *auch* (*as well*) and *selbst* (*even*) are allowed if they combine with lexically free slots of the PCls but not when combining with lexically fixed constituents. This is illustrated in (16).

(16) a. Peter weiß, wo [selbst ihn]/ [auch ihn] der Schuh drückt.
Peter knows where even him/ him as well him the shoe presses
(‘Peter knows what is worrying even him/ him as well.’)

b. Peter weiß, wo ihn #[auch/ selbst der Schuh] drückt.

It is possible to add a negation to the expression in (1-d) if the negation is inside the embedded free slot.

(17) Peter weiß, wo [nicht ihn sondern den Hans] der Schuh drückt.
Peter knows where not him but the Hans the shoe presses

('Peter knows what is worrying not him but Hans.')

The following picture emerges from our inspection of the properties of PCls: First, the information structure of the literal meaning must not be changed. This implies that the application of valence alternating operations is excluded (passive, possessive dative shift) and so are changes with respect to the topicalized constituent. Second, the propositional semantic core of the literal meaning must remain constant. In some restricted cases it may be modified by modal and temporal expressions. Third, syntactic and semantic alternations are possible if they concern free slots in the PCl or if they do not affect the information structure or the core propositional semantics of the PCl. Fourth, obligatory anaphoric relations may hold between elements inside the PCl and matrix elements.

The properties of PCls show that they cannot be treated as big "words with spaces". Instead, they are inherently complex syntactic units with different degrees of flexibility. This is parallel to what was observed for other idioms in Wasow et al. (1983) and elsewhere, and clearly sets PCls apart from fully fixed forms such as proverbs.

3 The Two-Dimensional Theory of Idioms

In this section we propose an extension of the theory of irregularity developed in Richter and Sailer (2003), Sailer (2003), and Soehn (2006). After summarizing the most important parts of that theory in Section 3.1, we will significantly extend it in Section 3.2 to capture the different degrees of regularity found in idiomatic constructions in a straightforward way. As PCls are characterized by a high degree of syntactic regularity, this extension is particularly important for a systematic analysis of PCls.

3.1 Internal and External Idiosyncrasies

The two-dimensional theory of idioms builds on the distinction between *decomposable* and *non-decomposable* idioms in Wasow et al. (1983). In our terminology, decomposable idioms will be treated as combinations of words with *external* irregularities, non-decomposable idioms are analyzed as phrases whose structures are *internally* irregular.

Decomposable idioms comprise expressions such as *make waves* (*cause trouble*) and *spill the beans* (*divulge information*). They show a considerable degree of syntactic and semantic flexibility. Following Wasow et al. (1983) and Gazdar et al. (1985) we treat them as syntactically free combinations of words with an idiom-specific meaning: In its idiomatic use the word *spill* is synonymous to *divulge* and the word *beans* means *information*. Our theory forces the two idiom-specific meaning variants of *spill* and *beans* to co-occur obligatorily. Under this perspective the expression *spill the beans* is not idiosyncratic with respect to the way it is put together from its syntactic and semantic components. What is special about it is that it contains two words with a highly restricted distribution, the idiomatic variants of *spill* and *beans*. For this reason we regard the idiomatic variants of these words as *distributionally* or *externally* idiosyncratic.

To implement the idea of distributional restrictions formally, two new attributes have been introduced into the grammar architecture. First, Soehn (2004) proposes the feature LISTEME.[1] Its value is a unique atomic identifier for each item that is listed in the lexicon. This feature allows us to distinguish between the word *spill* in the meaning *divulge*, used in the idiom *spill the beans*, and the non-idiomatic word *spill*. Second, the list-valued feature COLL (context of lexical licensing) is defined on the sort *sign*. For every lexical item, the value of COLL is a non-empty list of objects of sort *barrier*. These barrier objects specify two things: (i) a syntactic domain (in which they apply), and (ii) a requested licensing property. A general principle guarantees that in each structure, for each lexical item in this structure, and for each *barrier* object on the COLL list of each lexical item, the licensing property holds in the syntactic domain specified in the *barrier* object. To take a concrete example, the lexical specification of the idiomatic word *beans* introduces a COLL list with one *barrier* object on it. The barrier demands that the idiomatic word *beans* co-occur with a verb whose LISTEME value is *divulge-spill* and that the word *beans* act as the theme argument of that verb. Since the idiom *spill the beans* is very flexible syntactically, the licensing domain of our example is the entire utterance in which the idiomatic word *beans* occurs.[2]

The external dimension of idiosyncrasy is complemented by a dimension of *internal* idiosyncrasy. Non-decomposable idioms, such as *saw logs*, are syntactically and semantically frozen. This type of idioms may even have an otherwise unattested syntactic structure. Wasow et al. (1983) mention, among others, *kingdom come* and *trip the light fantastic* as extreme cases. An analysis in terms of fixed phrasal expressions appears to be most appropriate for these expressions. An expression such as *saw logs* is stored in the lexicon as a complex phrase with fixed syntactic structure and the idiosyncratic meaning *snore*.

The theory of idioms in Sailer (2003) and Soehn (2004, 2006) offers an encoding of such internally idiosyncratic phrases in HPSG that directly exploits the COLL feature. If a phrase has the specification [COLL *elist*], it is regular, or non-idiomatic. In this case the rules of regular syntactic and semantic combinatorics apply to it. An internally irregular phrase has the specification [COLL *nelist*] and is exempt. In this architecture, the lexicon contains the descriptions of words and of internally idiosyncratic phrases. The lexical entries of the latter are called *phrasal lexical entries* (PLE). The PLE for the expression *kingdom come* in (18) provides a simple example.

[1]For most purposes, this feature corresponds to the feature LEXICAL-ID (LID) in Sag (2007b). In contrast to Soehn (and Sag), we assume that LISTEME is not a head feature. See Section 3.2 for our motivation.

[2]The exact formalization of the collocational mechanism is not relevant in our present discussion. See Soehn (2004, 2006) for details.

(18) The phrasal lexical entry of *kingdom come*:

$$\left[\begin{array}{l} \textit{phrase} \\ \text{PHON } \boxed{1} \oplus \boxed{2} \\ \text{SS L} \left[\begin{array}{ll} \text{CAT LISTEME} & \textit{kingdom-come} \\ \text{CONT MAIN} & \textit{paradise} \end{array}\right] \\ \text{DTRS} \left[\text{N-DTRS} \left\langle \left[\text{PHON } \boxed{1} \langle \textit{kingdom} \rangle\right], \left[\text{PHON } \boxed{2} \langle \textit{come} \rangle\right] \right\rangle \right] \\ \text{COLL } \textit{nelist} \end{array}\right]$$

The phrase in (18) cannot be a regular phrase of English since its semantics is not derived regularly from the semantics of its constituents. The irregularity is possible because the specification [COLL *nelist*] exempts the phrase from the principles of combinatorial semantics. With respect to the internal syntactic structure, there is no identifiable syntactic head.[3] For this reason we specify it as consisting of two non-head daughters, but leave the details of the syntactic combination underspecified. This accounts for the fact that it is not clear what exactly the syntactic relation is between the two words *kingdom* and *come*.

The analysis accounts for the irregularity of the expression, and it also captures the idea that the two words in it are regular members of the English lexicon. This is a direct consequence of merely mentioning the phonological values of the daughters without restricting their syntactic or semantic structure in any other way. The grammar must independently license signs with the required phonological properties. This requirement can be met by the regular words *kingdom* and *come*. This mechanism may seem trivial at first, but it illustrates a principled locality restriction on idiosyncratic phrases: A phrasal lexical entry can only locally license idiosyncratic properties of a phrasal node, but it cannot introduce idiosyncratic properties at its daughters or at any deeper level of syntactic embedding.[4]

Thus far it may seem that the central attribute, COLL, is used for two unrelated and independent purposes: It encodes idiosyncratic distributional requirements, and it specifies whether a phrase is internally irregular. However, these two dimensions of irregularity are in fact related. Inspired by the idea that all lexical elements may enter into collocational relations (Sinclair, 1991), Sailer (2003) formulates the *Predictability Hypothesis.*

(19) Predictability Hypothesis (Sailer, 2003, p. 366):
For every sign whose internal properties are fully predictable, the distributional behavior of this sign is fully predictable as well.

Since simple lexical items such as basic words or nonderived lexemes are internally idiosyncratic, they may also display idiosyncratic distributional properties.

[3]Historically the expression stems from the phrase *thy kingdom come* in the Lord's Prayer. There the noun *kingdom* is used with a determiner, and the entire expression is a clause, not a noun phrase.

[4]As discussed in Sailer (2003), this sets apart our HPSG treatment of internally idiosyncratic phrases from an analysis that relies on *en bloc* insertion or the analysis in Tree Adjoining Grammar in Abeillé (1995), which treats non-decomposable idioms as idiosyncratic trees of arbitrary depth.

Similarly, internally idiosyncratic phrases may also show distributional irregularities. Again, the phrase *kingdom come* is a good example. It is almost exclusively restricted to the combinations with one of the three prepositions in *until/till/to kingdom come*.[5] Based on this observation we may assume that a more complete description of the COLL list of the PLE (18) should mention a *barriers* object whose purpose it is to require that *kingdom come* be the complement of one of these prepositions.

3.2 Partial Regularity in Irregular Phrases

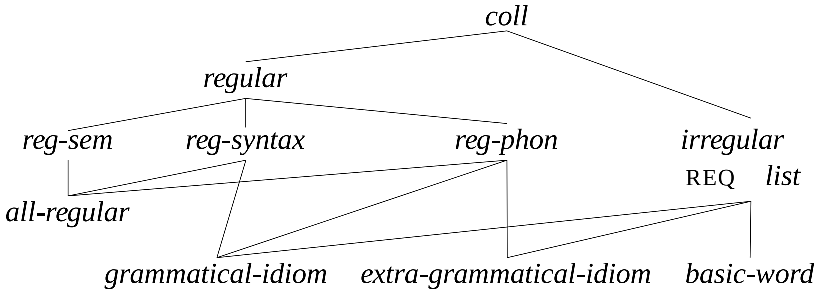
While the idiom *kingdom come* is syntactically irregular, the majority of idioms is not. Idiomatic expressions are characterized by a particular, 'idiomatic' meaning and by their frozen syntactic structure, but apart from their unusual fixedness they have regular internal syntactic properties. Typical examples are expressions such as *kick the bucket* (*die*) and *saw logs* (*snore*). In both cases, we assume that the syntactic structure is that of a transitive verb that combines with its direct object. In the version of our theory in Section 3.1 we had to encode all regular aspects of the structure of idioms explicitly in each phrasal lexical entry. This introduces an unwanted descriptive overhead in grammars which already contain all relevant syntactic well-formedness conditions on phrases. In the present section we propose an extension of the theory that captures the insight that many irregular phrases are merely irregular with respect to the syntax-semantics mapping but not with respect to their syntax.

The core innovation is a distinction between constraints that apply to different modules of the grammar. We tentatively assume that there are phonological, syntactic, and semantic principles. The SUBCAT PRINCIPLE is a syntactic constraint, the SEMANTICS PRINCIPLE is one of the semantic ones, and linearization principles count as phonological constraints. Since the COLL value of signs is the place where we mark their idiosyncrasies, this is also where we specify to which degree a sign exhibits idiosyncratic behavior. For that purpose, we enrich the structure of COLL values. From now on they are of sort *coll*. The subsorts of *coll* specify the degree of regularity of an expression. For irregular items an attribute REQ(UIREMENT) is defined, whose value is a list of *barrier* objects. This list corresponds exactly to the earlier COLL value. The details are provided in FIGURE 1.

The names of the maximally specific subsorts of *coll* are inspired by the corresponding classification of idioms in Fillmore et al. (1988). In that system the idiom *kingdom come* is classified as extra-grammatical because it only shows regularity with respect to its phonological properties. In our hierarchy, it is classified as an irregular construction which obeys the restrictions of phonologically regular constructions but not those of syntactically or semantically regular constructions. It receives a phrasal lexical entry with the COLL value *extra-grammatical-idiom*.

[5] The British National Corpus contains 35 occurrences of *kingdom come* out of which 5 are irrelevant (band name or coincidental co-occurrence of the two words), 14 reflect the biblical usage, 13 are with one of the above-mentioned prepositions, 3 others are uses of *kingdom come* as a noun.

Figure 1: Sort hierarchy below the sort *coll*



The idiom *saw logs* is a grammatical idiom according to Fillmore et al. (1988), and is therefore specified as [COLL *grammatical-idiom*]. The maximally specific sort *grammatical-idiom* is a subsort of both *regular-phonology* and *regular-syntax*. For that reason it is subject to all syntactic and phonological principles, but not to the principles of regular semantic composition. Finally, regular phrases have the COLL value *all-regular* and obey all principles of syntax, semantics and phonology.

In Section 3.1 we said that the principles of grammar only apply to signs with an empty COLL list. This theory must now be revised and made sensitive to the subsorts of *coll*. Syntactic principles such as the IMMEDIATE DOMINANCE PRINCIPLE (ID PRINCIPLE), the HEAD FEATURE PRINCIPLE, the NONLOCAL FEATURE PRINCIPLE etc. must be modified. All syntactic principles defined on the sort *sign* are relativized to apply only to signs with the COLL specification *regular-syntax*. This can be achieved by simply adding a further antecedent to the original formulation of the principles. The relativized version of the ID PRINCIPLE in (20) illustrates this technique.

(20) Relativized ID PRINCIPLE:

$$\left[\textsc{coll}\ \textit{regular-syntax}\right] \\ \quad \Rightarrow \left(\textit{phrase} \Rightarrow (\textsc{Head-Specifier-Schema}\ \textbf{or}\ \textsc{Head-Complement-Schema}\ \textbf{or}\ \ldots)\right)$$

Idioms that are not extra-grammatical are still subject to all syntactic well-formedness conditions according to the *coll* hierarchy, because *grammatical-idiom* is subsort of *regular-syntax*. It follows that each grammatical idiom must obey one of the ID SCHEMATA. For example, the VP *saw logs* is specified as a regular head-complement construction, and as such it is licensed by the HEAD-COMPLEMENT SCHEMA. No special steps need to be taken to guarantee this result.

Let us finally turn to the principles of semantic composition. While many idioms are syntactically regular, they all show semantic idiosyncrasy. To capture this behavior, the principles of semantic composition need to be relativized parallel to what we did in (20). The relativized SEMANTICS PRINCIPLE is given in (21), where SP is the description of the original SEMANTICS PRINCIPLE.

(21) Relativized SEMANTICS PRINCIPLE:

$\left[\text{COLL } \textit{regular-semantics}\right] \Rightarrow \text{SP}$

The sort hierarchy in FIGURE 1 is constructed in such a way that we exclude the existence of irregular phrases that are semantically regular but syntactically or phonologically irregular. On the other hand, as soon as a phrase is syntactically irregular, we expect it to be semantically irregular as well. In this respect we agree with the assumptions in Fillmore et al. (1988).[6]

So far we have focussed on the principles of grammar. Let us now consider the question of what constitutes the lexicon. All properties of signs with the COLL specification *regular* follow from the properties of their constituents and from the general combinatorial principles of the grammar. Signs with the COLL value *irregular*, however, require a further specification of those of their properties that are not predictable from general grammar rules. This specification is given in the lexicon. In our sort hierarchy below *coll* we distinguish three subsorts of *irregular*. The sorts *grammatical-idiom* and *extra-grammatical-idiom* are confined to irregular signs that have non-trivial internal syntactic structure. In the context of the present discussion, these are phrasal signs. The sort *basic-word* is reserved for signs without internal structure. In the present discussion, this means that it is the COLL value of words. Words (which we view here as non-recursive signs) always display an unpredictable form-meaning combination, which qualifies them as irregular. The idea that basic words are necessarily irregular and that phrases cannot have the COLL value *basic-word* is captured in the principle in (22).

(22) BASE-LEXICON PRINCIPLE:

$\left[\textit{word}\right] \Leftrightarrow \left[\text{COLL } \textit{basic-word}\right]$

In the preceding discussion we deliberately ignored the fact that words may have internal structure as well. As soon as a more elaborate view on morphological structure (such as the one presented in Sag et al. (2003)) is adopted, the BASE-LEXICON PRINCIPLE needs to be refined in such a way that the most basic, non-recursive subsort of *sign* replaces *word* on the lefthand side of the principle. Furthermore, the type hierarchy below *coll* will need some extension as well in response to additional principles of the morphological combinatorics.

The lexicon is defined by means of a WORD PRINCIPLE. This principle provides lexical entries for all irregular signs. In (23) LE refers to lexical entries of basic words, PLE refers to phrasal lexical entries of grammatical or extra-grammatical idioms.

[6] All idioms considered in this paper are phonologically regular. Nonetheless we include the type *regular-phon* to allow for a relativization of the principles of phonological combinatorics such as the CONSTITUENT ORDER PRINCIPLE.

(23) WORD PRINCIPLE:

$$\begin{bmatrix} \textit{sign} \\ \text{COLL} \ \ \textit{irregular} \end{bmatrix} \Rightarrow (\ \text{LE}_1 \vee \ldots \text{LE}_n \vee \text{PLE}_1 \vee \ldots \vee \text{PLE}_{n'}\ )$$

In Section 3.1 we emphasized the importance of the LISTEME attribute for our theory of idioms. The name "listeme" is chosen very deliberately in Soehn (2006) because Soehn assumes that all listed expressions contribute their own unique LISTEME value. This means that every lexical entry, phrasal or not, may have its own LISTEME value. An internally irregular phrase such as *kick the bucket* has a LISTEME value, say *kick-the-bucket-idiom*, which differs from the LISTEME values of all of its daughters. While the HEAD FEATURE PRINCIPLE guarantees that all head features such as VFORM, AUX are shared between the phrasal mother and the head daughter, the idiomatic phrase *kick the bucket* and its head daughter do not share the LISTEME value. It follows that Soehn's assumption that LISTEME is a head feature is not compatible with the present architecture. For this reason we treat LISTEME as a *category* feature instead. With the new position in the feature geometry, it is necessary to introduce a principle for the percolation of LISTEME values in regular phrases. The principle that takes care of that is given in (24). It is among those principles that apply only to non-idiomatic phrases.

(24) The LISTEME PRINCIPLE:

$$\begin{bmatrix} \text{DTRS} \ \ \textit{headed-phrase} \\ \text{COLL} \ \ \textit{all-regular} \end{bmatrix} \Rightarrow \begin{bmatrix} \text{SYNS LOC CAT LISTEME} \ \boxed{1} \\ \text{DTRS} \left[\text{H-DTR} \ \left[\text{SYNS LOC CAT LISTEME} \ \boxed{1}\right]\right] \end{bmatrix}$$

As a consequence of our classification of grammar principles into syntactic, semantic and phonological, the structure of the *coll* hierarchy, and the fundamental lexical principles (22) and (23), there are four possibilities for a well-formed sign: Basic words always exhibit some degree of idiosyncrasy and are singled out as having their own irregular collocation type, *basic-word*. Phrases come in three flavors. A phrase may be completely regular, in which case it has the COLL *all-regular* and is subject to all principles of grammar. If it is irregular, it must be licensed by one of the phrasal lexical entries in the WORD PRINCIPLE. If a phrase is of COLL TYPE *grammatical-idiom*, it is subject to the regular syntax principles. If a phrase is an *extra-grammatical-idiom*, it is irregular to a degree that it is exempt from the principles of syntax.

Before closing this section let us briefly return to the role of the Predictability Hypothesis (19). This hypothesis establishes a link between internal irregularity and the potential of specifying external idiosyncrasy. In the version of the theory in Section 3.1, all and only regular signs have an empty COLL list. In the modified architecture of the present section, only signs with a COLL value of sort *irregular* have an REQ attribute, and only they can be specified for externally idiosyncratic behavior. The idea of the Predictability Hypothesis is directly encoded in the signature of the grammar module that handles irregularity.

Note that our new architecture foresees cases in which an internally irregular sign is not distributionally constrained. While this possibility is denied in most of the work on collocations, it seems to be the standard assumption in formal approaches to grammar. It is also compatible with the original formulation of the Predictability Hypothesis. So far, we do not see compelling reasons for claiming that idioms such as *kick the bucket* are distributionally constrained. In previous versions of the theory we were forced by the architecture to assume that there was some *barrier* object inside the COLL value, although no such object was explicitly specified. With the new version of the theory, it is possible to combine the COLL value *grammatical-idiom* of a phrase with an empty COLL REQ list.

4 Modeling Phraseological Clauses as Phrasal Lexical Entries

In FIGURE 2 we sketch the PLE for the idiom in (2-a). The PLE specifies that the overall clause is a modifier with the semantics *ad nauseam*. The phrase is a head-complement combination, where the head daughter is the preposition *bis* (*until*). The non-head daughter is a finite clause. Inside the complement, there must be a verbal word with the LISTEME value *kommen* whose subject is a definite singular NP with the word *Arzt* as its lexical head. The PLE specifies the COLL value as *grammatical-idiom*. Consequently, all principles of syntax apply, which means that we do not need to specify the HEAD value of the clause nor the effect of the SUBCAT PRINCIPLE or of the HEAD-COMPLEMENT SCHEMA. The REQ value of the clause is empty, which expresses the observation that there are no further constraints on the distribution of the PCl.

The data section showed that there are some restrictions on the structure of this PCl: While tense and modality may vary, negation is not permitted. This can be expressed by requiring that there be no negation in the content of the PCl. For other PCls we must also ban modal operators from the semantic representations. Since modalities can be contributed by modal verbs and by adverbials, the restriction must be imposed on operators in the CONTENT value of the PCls.

In Section 2 we saw that all PCls we considered disallow alternations that change the information structure of their literal meaning. Since the constituents of PCls are non-idiomatic in our theory, the literal meaning of their combination is in principle available. As there are various proposals to model information structure in HPSG, it should in principle be possible to formulate an appropriate constraint on information structure. For reasons of space, we will not pursue this direction here. Instead, we exclude valence alternations by other types of restrictions in the PLEs. To exclude the passive and the dative-possessive alternation in (1-a) and (1-b), we impose syntactic restrictions on the ARG-ST or the VAL value of the words in the expressions, which are all available in PLEs. To keep the analysis simple and as complete as possible, we will stick with this strategy for the rest of the paper.

Figure 2: Sketch of the phrasal lexical entry of *bis der Arzt kommt*:

$$\begin{bmatrix} \textit{phrase} \\ \text{SS L} \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{HEAD MOD} \begin{bmatrix}\text{L CONT } e\end{bmatrix} \\ \text{LISTEME} \quad \textit{ad-nauseam}\end{bmatrix} \\ \text{CONT MAIN } \mathsf{ad\text{-}nauseam}'(e) \end{bmatrix} \\ \text{DTRS} \begin{bmatrix} \textit{head-complement-struc} \\ \text{H-DTR} \begin{bmatrix}\text{SS L CAT}\begin{bmatrix}\text{HEAD} & \textit{prep}\\ \text{LISTEME} & \textit{bis} \\ \text{VAL COMPS} & \left\langle \begin{bmatrix}\text{L CAT HEAD}\begin{bmatrix}\textit{verb}\\ \text{VFORM } \textit{fin}\end{bmatrix}\end{bmatrix}\right\rangle\end{bmatrix}\end{bmatrix} \\ \text{C-DTRS} \left\langle \begin{bmatrix}\text{DTRS}\begin{bmatrix}\ldots\text{H-DTR}\begin{bmatrix}\textit{word}\\ \text{SS L CAT}\begin{bmatrix}\text{LISTEME } \textit{kommen}\\ \text{VAL SUBJ}\left\langle \text{NP}\begin{bmatrix}\text{LISTEME } \textit{arzt},\\ \text{DEF } +, \textit{sg}\end{bmatrix}\right\rangle\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}\right\rangle \end{bmatrix} \\ \text{COLL}\begin{bmatrix}\textit{grammatical-idiom}\\ \text{REQ } \textit{elist}\end{bmatrix}\end{bmatrix}$$

Let us now turn to a more intricate example. The data (25)–(28) reveal details about the frozenness of the PCl in (1-c). We compare the PCl in the (a) sentences with a parallel non-idiomatic construction in the (b) sentences. (25) shows that the PCl requires an anaphoric relation between the matrix subject and the accusative argument in the PCl. As we remarked earlier, neither an overt complementizer nor a change in the constituent that occupies the *vorfeld* are permitted (see (26) and (27)). The PCl may not occur in the *vorfeld* of the matrix sentence (28).

(25) a. Ich glaub, mich/#dich tritt ein Pferd.
I believe me/you kicks a horse
'I am very surprised.'
b. Ich glaub, mich/dich jagt eine Kuh.
I believe me/you chases a cow
'I believe a cow is chasing me/you.'

(26) a. #Ich glaub, dass mich ein Pferd tritt.
b. Ich glaub, dass dich eine Kuh jagt.

(27) a. #Ich glaub, ein Pferd tritt mich.
b. Ich glaub, eine Kuh jagt dich.

(28) a. #[Mich tritt ein Pferd], glaub ich.
b. [Dich jagt eine Kuh], glaub ich.

In FIGURE 3 we sketch the relevant PLE. The COLL value *grammatical-idiom* accounts for the syntactically regular internal structure. The PCl is specified as a verb-second clause whose lexical head is the verb *treten*. This verb must take two arguments. The first one is an indefinite NP headed by *Pferd*. The second one is an accusative NP which is fronted. This condition follows from the LOC value identity, $\boxed{1}$, between the element on the ARG-ST list and the highest non-head daughter, which is a filler daughter. The VP might be modified by adjuncts, which is accounted for by only requiring that *treten* be the syntactic head of the construction. The appeal to the regular expression notation ((DTRS HDTR)$^{+}$) is only meant as a more readable abbreviation of a (technically more accurate) relational expression that relates the head daughter of the head-filler structure to its head daughters.

Next we turn to the COLL REQ value. As in the other sentences in (1), the combination of the matrix verb and the PCl behaves like in decomposable idioms of the type *spill the beans*, except that the complement is now a clause instead of an NP. According to the theory in Soehn (2004), this means that the matrix verb selects a complement with a particular LISTEME value. The complement clause, in turn, has a non-empty REQ list. The element on its REQ specifies that the PCl must co-occur with a particular matrix verb, the listeme *surprise-glauben*. The PCl must be the complement clause of this matrix verb. Furthermore, the sort specification indicates the syntactic domain within which the co-occurrence must hold. In Soehn (2004) the sort *vp_ne* is used to specify that the relevant domain is the LOCAL value of the smallest projection of the matrix verb that dominates both the matrix verb and the complement clause. In other words, the PCl must occur as a sister to (the trace of) the matrix verb. What is most important for our purposes is that information about the matrix verb is available in the formulation of the PLE of the clausal complement. This is necessary to encode that the INDEX value of the embedded direct object, $\boxed{2}$, is identical with that of the matrix subject.

To sum up, the PLE in FIGURE 3 excludes passive alternation (see (5-c)), because it specifies that the verb *treten* occurs with a transitive argument structure. It also requires that the PCl be a verb-second clause (see (6-a)) by specifying that it is a head-filler structure, and it determines the first constituent (see (6-b)) by specifying it in the PLE. The anaphoric relationship between the embedded accusative and the matrix subject is also encoded directly.

There is in fact further evidence that a special relationship holds between the embedded PCl and the matrix predicate, and that the matrix predicate is not the free form of the verb *glauben*. We already saw in (9) that, depending on the tense form, the matrix verb is either a form of *glauben* or of *denken*. The relevant judgments are shown in (29).

(29) a. Ich dachte, mich tritt ein Pferd. (past of *denken*, present in the PCl)
b. ??Ich glaubte, mich tritt ein Pferd. (past of *glauben*, present in the PCl)
c. #Ich glaubte, mich trat ein Pferd. (past of *glauben*, past in the PCl)
d. ?Ich denke, mich tritt ein Pferd. (present of *denken*, present in the PCl)

The combination of *glauben* and the PCl is a decomposable idiom, because the very same variation of the matrix predicate can be observed with other PCls (see (9)). German has a special listeme (which we call *surprise-glauben*) which comprises forms of *glauben* and *denken* in its paradigm and combines with complement

Figure 3: Sketch of the PLE for the idiom *glauben, X_acc tritt ein Pferd*:

$$
\left[\begin{array}{ll}
\textit{phrase} & \\
\text{SS LOC} & \left[\begin{array}{ll}\text{CAT LISTEME} & \textit{very-surprised}\\ \text{CONT MAIN} & \text{surprised}'(x_{\boxed{2}})\end{array}\right]\\
\text{DTRS} & \left[\begin{array}{ll}
\text{F-DTR} & \left[\begin{array}{l}\textit{word}\\ \text{SS LOC } \boxed{1}\end{array}\right]\\
\text{HDTR} & \left[\begin{array}{ll}
\text{LF EXC} & \text{'a horse kicks } x_{\boxed{2}}\text{'}\\
(\text{DTRS HDTR})^{+} & \left[\begin{array}{ll}
\textit{word} & \\
\text{SS LOC CAT} & \left[\begin{array}{ll}\text{HEAD} & \left[\text{TENSE } \textit{pres}\right]\\ \text{LISTEME} & \textit{treten}\end{array}\right]\\
\text{ARG-ST} & \left\langle \begin{array}{l}\text{NP}\left[\text{LISTEME } \textit{pferd}, \text{ DEF -}, \textit{sg}\right],\\ \left[\text{LOC } \boxed{1}\left[\begin{array}{ll}\text{CAT} & \left[\text{HEAD CASE } \textit{acc}\right]\\ \text{CONT} & \left[\begin{array}{l}\textit{ppro}\\ \text{INDEX } \boxed{2}\end{array}\right]\end{array}\right]\right]\end{array}\right\rangle
\end{array}\right]
\end{array}\right]
\end{array}\right]\\
\text{COLL} & \left[\begin{array}{ll}
\textit{grammatical-idiom} & \\
\text{REQ} & \left\langle\left[\begin{array}{ll}\textit{vp\_ne} & \\ \text{LOC-LIC} & \left[\text{CAT}\left[\begin{array}{ll}\text{LISTEME} & \textit{surprise-glauben}\\ \text{VAL SUBJ} & \langle \text{NP}_{\boxed{2}}\rangle\end{array}\right]\right]\end{array}\right]\right\rangle
\end{array}\right]
\end{array}\right]
$$

clauses that express (negative) surprises, astonishment, or annoyance. (30) shows that this listeme can be found with non-idiomatic complement clauses as well.

(30) Ich glaub/ ?denk(e)/ ?*glaubte/ dacht(e)

a. der hat 'nen Vogel.
he has a bird ('... he is crazy')

b. das muss jetzt echt alles nochmal neu gemacht werden.
this must now really all again new made be
('... this must all be redone [annoyed]')

We conclude that even though the matrix predicate is not the free form of *glauben*, it is an instance of a (special) attitude predicate that also occurs outside of idioms. For this reason, the matrix predicate need not be restricted to a particular PCl. However, the PCl in FIGURE 3 must be collocationally bound to this special matrix predicate, and the PCl must impose its context requirement in the lowest dominating VP to exclude its own topicalization.

5 Modelability under Strict Locality Assumptions?

The two-dimensional theory of idioms is capable of capturing the properties of PCls. Being able to refer to deeply embedded parts of a phrase in a PLE is an important ingredient of this theory. It makes HPSG especially well-suited to integrate a fundamental insight of Construction Grammar: Constructions can span more than a local tree (Fillmore et al., 1988; Jackendoff, 1995).

In this section we briefly consider a few interesting aspects of a second approach to construction-like phenomena in HPSG, which offers a possible alternative to our analysis of PCls. However, we do not intend a thorough comparison of the two approaches and only point out a few interesting similarities and differences. In a recent series of papers (Sag (2007a,b) and others) it was shown that various phenomena of apparent non-locality can be encoded using an extension of HPSG's feature geometry and a restructuring of signs. In the framework proposed there, *Sign-Based Construction Grammar* (SBCG), phrasal signs no longer contain their daughters. Instead, *construct* objects are introduced that correspond to local trees. Signs only occur as nodes in these constructions. A sentence consists of a set of constructions, each of which represents a local tree, but these trees do not form a single joined feature structure. With this architectural change the formulation of PLEs like the ones in FIGURE 2 and FIGURE 3 is not possible.

To account for non-locality SBCG uses two head features: the listeme attribute LEXICAL-ID and the attribute XARG whose value is the subject of the sentence. These two attributes are sufficient to describe the construction in (2-a), because the obligatory elements in the embedded clause are the lexical head *kommen* and the subject, *Arzt*, i.e. exactly those parts that are locally available for the overall construction.

(31) A SBCG description of *bis der Arzt kommt*:

$$\begin{bmatrix} \textit{bis-der-arzt-kommt-cxt} & \\ \text{MOTHER} & \begin{bmatrix} \text{MOD} & \begin{bmatrix}\text{SEM} & \boxed{1}\end{bmatrix} \\ \text{SEM} & \textit{ad-nauseam}(\boxed{1}) \end{bmatrix} \\ \text{DTRS} & \left\langle \begin{bmatrix}\text{LID} & \textit{bis}\end{bmatrix}, \text{S}\begin{bmatrix} \text{XARG} & \begin{bmatrix}\text{LID} & \textit{arzt}\end{bmatrix} \\ \text{LID} & \textit{kommen}\end{bmatrix} \right\rangle \end{bmatrix}$$

To allow modal verbs and temporal auxiliaries we can simply assume that the LID value of a verbal complex is identical with that of the most deeply embedded lexical verb in the verbal complex. To exclude modal and temporal variation in other idioms, we could impose the same kind of restrictions as in Section 4, i.e. we could describe which operators may not occur in the content values of the daughters.

Recall that truth-conditionally neutral, grammatical variation occurs in some but not all PCls. In the two-dimensional account we refer to the ARG-ST value of an embedded verb to exclude passive and other valence alternations. Since SBCG allows reference to the highest subject in a PCl, active-passive alternations can similarly be excluded by requiring a particular LID value inside the XARG value. Alternations that do not involve the subject are harder to treat because it is only the subject information that percolates up the tree.

This brings us to an interesting problem. Information about the arguments inside a PCl is not only necessary to restrict valence alternation, it is also important to express the coreference constraints attested with many PCls. The PCl in (1-c) is a good example. The accusative NP inside the PCl must be a pronoun that is anaphorically related to the matrix subject. The accusative object is on the ARG-ST list of the embedded verb. The matrix subject is on the ARG-ST list of the matrix verb, the matrix verb has access to the LID value of the embedded verb and to its XARG value. However, neither of them can be used to establish a link between the embedded accusative NP and the matrix subject. The same problem occurs in other cases where the PCl contains an embedded open slot that must be anaphorically related to the matrix subject: The PCls in (1-b) and (2-c) require such a relation to an embedded dative object. A potential way out within SBCG is the introduction of a percolation mechanism for the entire ARG-ST values instead of the more restricted subject percolation mechanism. While this solution works for the cases of German PCls that we have found so far, the English example in (32) might still be a problem. In this expression the element X must be coreferential with the matrix subject. However, X is embedded in a locative modifier. Unless locative modifiers are on the ARG-ST list, the locality assumptions of SBCG do not seem to leave the necessary kind of structure accessible to enforce coreference between X and the matrix subject.

(32) look as if butter wouldn't melt [in X's mouth] ('look completely innocent')

At the moment, we do not see which kind of solution is most appropriate for the general locality assumptions that underly SBCG. We thus leave this issue to future research.[7]

6 Conclusion

In this paper we drew attention to a largely neglected subclass of idioms: Idioms that contain full clauses, phraseologically fixed clauses (PCl). We investigated properties of German PCls and arrived at new generalizations about their potential fixedness and flexibility. While there is a certain range of syntactic and lexical variation, all PCls we investigated forbid the application of syntactic processes that change the information structure of their literal meaning.

To account for the frozenness of PCls together with their regular internal syntactic structure we substantively modified the two-dimensional theory of idioms developed in Richter and Sailer (2003), Sailer (2003), and Soehn (2006). These earlier versions of the theory had already incorporated the distinction between decomposable and non-decomposable idioms, but only the modified theory lets us express the systematic differences between grammatical and extra-grammatical idioms. The theory captures the empirical properties of German PCls.

[7] See (Müller, 2007, Chapter 12.3) for further critical remarks on SBCG's locality assumptions and fundamental open questions about its architecture.

We very briefly compared our account with a possible alternative analysis in the framework of Sign-Based Construction Grammar (SBCG). Some properties of PCls may be problematic for SBCG's strict locality assumptions. Our theory can be seen as taking a middle position between the SBCG view, which demands that constructions only span local trees, and the traditional Construction Grammar perspective, which holds that constructions can be of arbitrary structural complexity. In our system a construction is licensed by a phrasal lexical entry (PLE). A PLE does two important things: First, an idiosyncratic semantic and/or syntactic combination is licensed in a local tree. Second, restrictions can be imposed on constituents that are embedded inside this combination. The first property is a weak version of a locality assumption: A PLE can only license an idiosyncrasy in an immediate mother-daughter relation. The second property, however, is a weak version of a complexity assumption: We can refer to properties of elements that are deeply embedded in the structure of the phrasal sign licensed by the PLE. In this setting it is crucial to realize that the embedded constituents must be independently well-formed. This means that we can restrict which ones of the well-formed signs may occur inside the overall expression, but a PLE cannot license embedded, idiosyncratically structured signs. In this sense, our approach incorporates the idea of arbitrary depth of constructions, but it also inherits the insight of phrase structure grammars that complex structures are built from local trees.

The two-dimensional theory of idioms that we developed in this paper helps us to reduce the amount of individually specified idiosyncrasy in the description of idiomatic constructions even further than its predecessor. The principles of the regular syntactic combinatorics apply to grammatical but non-decomposable idioms. We obtain a very flexible grammar architecture which covers two apparently contradicting tendencies in the domain of idioms at the same time: The need to allow for irregularity at all levels; and the observation that most idioms are not completely arbitrary in their structure but largely obey regular principles of grammar.

References

Abeillé, Anne. 1995. The Flexibility of French Idioms: A Representation with Lexical Tree Adjoining Grammar. In Martin Everaert, Eric-Jan van der Linden, André Schenk and Rob Schreuder (eds.), *Idioms. Structural and Psychological Perspectives*, pages 15–42, Lawrence Erlbaum Associates, Hillsdale.

Fillmore, Charles, Kay, Paul and O'Connor, M. 1988. Regularity and Idiomaticity in Grammatical Constructions: The Case of *Let Alone*. *Language* 64, 501–538.

Fleischer, Wolfgang. 1997. *Phraseologie der deutschen Gegenwartssprache*. Niemeyer, Tübingen, second edition.

Gazdar, Gerald, Klein, Ewan, Pullum, Geoffrey and Sag, Ivan. 1985. *Generalized Phrase Structure Grammar*. Harvard University Press, Cambridge, Mass.

Jackendoff, Ray. 1995. The Boundaries of the Lexicon. In Martin Everaert, Eric-Jan van der Linden, André Schenk and Rob Schreuder (eds.), *Idioms. Structural and Psychological Perspectives*, pages 133–165, Lawrence Erlbaum Associates, Hillsdale.

Müller, Stefan. 2007. *Head-Driven Phrase Structure Grammar: Eine Einführung*. Tübingen: Stauffenburg Verlag.

Richter, Frank and Sailer, Manfred. 2003. Cranberry Words in Formal Grammar. In Claire Beyssade, Olivier Bonami, Patricia Cabredo Hofherr and Francis Corblin (eds.), *Empirical Issues in Formal Syntax and Semantics*, volume 4, pages 155–171, Paris: Presses Universitaires de Paris-Sorbonne.

Sag, Ivan A. 2007a. Remarks on Locality. In Stefan Müller (ed.), *Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar, Stanford, 2007*, pages 394–414, Stanford: CSLI Publications.

Sag, Ivan A. 2007b. Sign-Based Construction Grammar. An informal synopsis, manuscript, Stanford.

Sag, Ivan A., Wasow, Thomas and Bender, Emily M. 2003. *Syntactic Theory: A Formal Introduction*. Stanford: CSLI, second edition.

Sailer, Manfred. 2003. Combinatorial Semantics and Idiomatic Expressions in Head-Driven Phrase Structure Grammar. Phil. Dissertation (2000). Arbeitspapiere des SFB 340. 161, Universität Tübingen.

Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Soehn, Jan-Philipp. 2004. License to COLL. How to bind bound words and readings to their contexts. In Stefan Müller (ed.), *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar, Center for Computational Linguistics, Katholieke Universiteit Leuven*, pages 261–273, Stanford: CSLI Publications.

Soehn, Jan-Philipp. 2006. *Über Bärendienste und erstaunte Bauklötze. Idiome ohne freie Lesart in der HPSG*. Frankfurt am Main: Peter Lang, Ph.D. thesis, Friedrich-Schiller-Universität Jena.

Wasow, Thomas, Sag, Ivan A. and Nunberg, Geoffrey. 1983. Idioms: An Interim Report. In S. Hattori and K. Inoue (eds.), *Proceedings of the XIIIth International Congress of Linguists*, pages 102–115.

Accounting for Underlying Forms in HPSG Phonology

Filip Skwarski

University of Warsaw

Proceedings of the HPSG09 Conference

Georg-August-Universität Göttingen, Germany

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

The paper aims to present approach to HPSG phonology which would account for underlying forms of phonemes. It shows some of the issues arising in monostratal analyses of phonology, and proposes a solution based on a notion of underlying representations. The approach presented, partly inspired by Optimality Theory, resolves cases of neutralisation and opacity by formulating constraints which either restrict the surface representation or relate it to the underlying form.

1 Why are underlying forms desirable in HPSG phonology?

Since most work in HPSG is focused on syntax or morphology, standard representations of phonology are reduced to supplying a word with a feature of type *phon*, a string of symbols equivalent to the orthographical spelling or pronunciation of the word (and its particular variants). Such strings are afterwards combined in higher-level objects to form phrases. While such a simplified approach is sufficient for solving problems based around syntax, morphology, and semantics, the HPSG does show a lot of potential for expanding the phonology within its framework.

One notable attempt to do so was undertaken by Bird (1995), who introduced constraint-based phonology into HPSG. The framework outlined in *Constraint Based Phonology: A Computational Approach* is essentially a monostratal system, where well-formedness of a particular word or phrase is decided by phonological and morphological constraints. This system allowed for linking phonology and morphology, and resolved issues by ruling out ill-formed segments and word structures.

The framework proposed was based around the principles of COMPOSITIONALITY and a requirement that a framework be MONOSTRATAL. The latter meant, in simplified terms, that any phonological representation has only one level, corresponding to forms actually appearing in the surface representations, and no abstract representation is stored:

> An even stronger constraint than those mentioned above is the requirement that a linguistic framework be MONOSTRATAL. This means that there is only a single level of linguistic description; descriptions pertain to occurring surface forms and not to artificially constructed abstract representations. As we shall see in section 1.5.1, the requirement that a linguistic framework be monostratal is equivalent to

> the true generalisation condition from Natural Generative Phonology. (Bird 1995, 1.4.5, p. 34)

Although such an approach would seem to be desirable in a computational framework, the phonological phenomena in various languages cannot be adequately described without a further reference to an underlying representation of a phoneme (Shoun 2005, 4.4). The cases pointed out in Shoun (2005) include eg. neutralisation phenomena in Bengali, there is no proposal as for the actual implementation of the underlying forms in HPSG phonology, however - which is the aim of this paper.

Evidence for usefulness of underlying representations can be seen in consonant alternations and voicing processes in languages where those phenomena are complicated, even though Bird (1995) seems to disregard events such as final devoicing as purely phonetic processes which need not be described with binary features, based on Port and Crawford (1989):

> The data show that speakers can control the degree of neutralisation depending on pragmatics and that information about the underlying contrast is distributed over much of the word. The data support a scalar valued neutralisation effect in the German voicing rule, and clearly refute a rule using a binary voicing feature. (Port and Crawford 1989, 257f)

Assuming such a position avoids the problem entirely by postulating that no instances of homophony due to devoicing exist, or in general that many alternation-related phenomena can be simplified as phonetic processes, while a substantial amount of evidence points out to the contrary.

In Polish (my native language), for example, the phoneme /ɡ/ exhibits the following alternations:

(1)	księga	a tome (nom.sg.)	[kɕɛŋɡa]	[ɡ]
	ksiąg	of tomes (gen.pl)	[kɕɔŋk]	[k]
	księdze	to a tome (dat.sg.)	[kɕɛndzɛ]	[dz]
	książka	a book (nom.sg.)	[kɕɔ̃w̃ʃka]	[ʃ]
	książek	of books (gen.pl.)	[kɕɔ̃w̃ʒɛk]	[ʒ]

Although these alternations result from historical palatalisation and voice assimilation processes, all of them are fully productive in modern Polish, in specific morphology-related cases, like noun declension patterns.

Likewise, in Polish - unlike eg. German - the process traditionally called "final obstruent devoicing" is intertwined with a process of "voice assimilation". Voiced obstruents are devoiced word-finally and before voiceless obstruents, while voiceless obstruents become voiced before voiced obstruents, including across word boundaries (Rubach 1982, 4.2, 4.3). As a result, /d/ and /t/ can both surface as [t] and [d] accordingly, phonetically identical with the "default" form of their opposite-voiced counterpart. Before sonorants (except, in most cases, across word boundaries), obstruents retain their "underlying" voice values, and so, in a traditional monostratal framework, we would have no way of arriving at this basic form if we simply described sonorants as either alternations of their surface representations or underspecifications (as suggested by Bird (1995), 1.5).

(2a)	kod	code	[kɔt]
	kody	codes	[kɔdɨ]
	kod dostępu	access code	[kɔd dɔstɛmpu]
	kod miasta	city code	[kɔt mʲasta]
	kod pocztowy	postal code	[kɔt pɔtʃtɔvɨ]
(2b)	kot	cat	[kɔt]
	koty	cats	[kɔtɨ]
	kot domowy	house cat	[kɔd dɔmɔvɨ]
	kot mały	small cat	[kɔt mawɨ]
	kod perski	Persian cat	[kɔt pɛrski]

The above data demonstrates that obstruents in Polish can behave in three ways depending on context:

1. assimilate their voice to that of the following segment (before other obstruents, including across word boundaries),
2. retain their "underlying" voice feature (before sonorants, except word-finally),
3. become voiceless regardless of their "underlying" voice feature (word-finally before a pause or before sonorants).

This example (not unlike the neutralisation example in Bengali, in Shoun 2005) will be used as a basis for representing the possibilities of accounting for underlying forms in HPSG phonology, in a further section.

1.1 Underspecification and Surface Constraints

Before presenting an approach utilising the notion of underlying representation to resolve these issues, it is worth looking at some of the views on alternations in HPSG phonology presented so far. One of the possibilities in line with Bird's original idea would be to use purely surface-relevant constraints, providing separate structures for various levels of the sentence if necessary (words, utterances, syllables, etc.). Such a solution is adopted by Bird and Klein (1994) and suggested by Höhle (1999).

There are two possibilities of expressing the phenomenon of Polish final devoicing within such a framework:

1. Restrict the word structure in such a way that no voiced word-final obstruents are allowed, and the phrase structure in such a way that all obstruent segments must agree in voicing.
2. Restrict the word structure in such a way that no voiced word-final obstruents are allowed before a sonorant, and the phrase structure in such a way that all obstruent segments must agree with voicing.

Of these, solution 1. leads to an obvious conflict whereby in a phrase "kod dostępu" the phrase demands a voiced [d] while the word demands a voiceless [t], and therefore no proper form can be generated. Solution 2. leads to an underspecification, where in the cases of "kod", "kot", "kod pocztowy" or "kod dostępu", the voicing value is correctly predicted, but in "kody" or "koty" (word-medially), it is not determined at all (it is [d] v [t]), and we are in fact left with no means to predict it. We simply cannot "consult" it with anything.

Höhle (1999) attempts to tackle Russian obstruent voicing rules, not very different from the Polish ones, and in his approach seems to allow for different phoneme surface representations arising on different levels (Höhle 1999, fig. 7). While it is possible to differentiate between the representations of a particular phoneme present on the word level and the phrase level (by simply not making them identical in HPSG sense, and by arriving at the two by separate means), this leads to problems with coordinating the entire structure - see the notes on principles adopted for this framework in section 2., "Representing the Representations".

1.2 Morphology and Stem Spaces

Apart from Bird's and Höhle's proposals regarding HPSG phonology, an attempt to tackle phonological alternations (including irregular patterns displayed by some forms) is demonstrated in *Deriving Inflectional Irregularity*, Bonami and Boyé (2006). Here, a notion of stem space is introduced: declension patterns are based around a number of stem spaces, accounting for all stem alternations in inflection. For example, instead of (as assumed in transformational phonology) deleting endings of French feminine adjectives to produce masculine ones, the two are assumed to have different basic stem spaces (Bonami and Boyé 2006, 2.2).

But while French is (relatively) simple in terms of phonological processes, adopting such a framework in Polish would be complicated for a number of reasons:

1. In Polish, the voicing phenomena - as demonstrated above - are not only affected by the declension pattern, but also by the context of the following and preceding words. Eg. the lexical item "kod" cannot be simply a part of a class of nouns where [d] occurs in the stem before affixes and [t] in the nominative, unless we further account for the fact that the [t] in the nominative may alternate with [d] anyway.

Moreover, entire clusters may change their voice features:

(3a)	mózg	a brain	[musk]
	mózg był	a brain was	[muzg bɨw]
(3b)	zjadł	he ate	[zjatw̥]
	zjadł go	he ate him	[zjadw gɔ]

2. Polish is further complicated by other phonological phenomena accounting for further alternations:

(4a)	kocha	loves	[kɔxa]
	kochają	(they) love	[kɔxajɔ̃w̃]
	kochając	loving	[kɔxajɔnts]
(4b)	robi	does	[rɔbʲi]
	rób	do (imp.)	[rup]
(4c)	zjedli	they (masc.) ate	[zjɛdli]
	zjadliwe	edible (neut.)	[zjadlivɛ]
	zjadł	he ate	[zjatw̥]
	zjedzony	eaten (masc.)	[zjɛdzɔnɨ]
(4d)	Paryż	Paris	[parɨʃ]
	paryski	Parisian (masc.)	[parɨski]

The above examples demonstrate some of the processes: alternation between nasal vowels followed by a glide and oral vowels followed by a nasal consonant in (4a), alternation between [u] and [ɔ] in (4b), alternation between [l] and [w], [ɛ] and [a], as well as [d] and [dz] in (4c), and disappearance of fricatives in (4d). In addition, all obstruents (and extra-syllabic approximants) are also affected by voicing rules, best exemplified in (4c).

In the more extreme cases like (4c), virtually any segment found in a word can alternate with something else, leading to some situations where representing alternations exhibited by individual phonemes is actually easier than representing alternations of all possible stem forms.

3. Approaches based on morphological tools, unlike those based on global phonological constraints, essentially say nothing about the permissible structures of the words themselves, as long as they are assigned into productive patterns. In an optimal system, aside from handling alternations, it would be desirable to predict which words are well-formed according to the phonotactics of a given language, especially since, as demonstrated before, global constraints are useful in ruling out erroneous forms in situations where the choice of a proper form of a stem, affix, etc., is based purely on phonological background. On the other hand, introducing separate information about phonotactics would in many cases overlap with what is already handled by morphology.

The above remarks should demonstrate the major problems arising when using morphology-based tools and noun classes for adequately representing actual pronunciation of spoken words. While such an approach could be expanded, it would have to become mind-bogglingly complex for some languages, while invoking underlying representation removes the need for merging morphological, phonological and phonostylistic phenomena into one monster of a framework - every process can be dealt with separately by operating at the level at which it occurs and on the phonemes or features it is related to. No distinction between phonotactics and inflection becomes necessary, even though more advanced morphological issues, like the examples back in (1), can be addressed by invoking both morphological classes and the underlying representations.

2 Representing the Representations

This section is concerned with establishing the structural side of the framework which would involve underlying features. A well-functional framework should achieve the following aims:

(a) Allow formulated rules to operate at various levels of the structure (stem, word, syllable, utterance, etc.)
(b) Accurately provide just one surface form for any phoneme in the complete utterance.
(c) Append lower-level representations in higher-level representations (words into phrases, syllables into feet, etc.)
(d) Allow for interactions between the underlying representation and the surface

representation in cases where the underlying representation is directly relevant to the surfacing form.

Principle (a) is dictated by the observation that certain phonological phenomena are limited within syntactic context, such as word boundaries or phrase boundaries, and the constraints have to be formulated in a way accounting for this (Bird 1994, 2.2).

Principles (b) and (c) are related: because of the observation mentioned in (a), various constraints operating solely on one level of the structure (word, phrase, etc.) would predict different criteria of well-formedness. For example, a constraint demanding that the word-final segment be voiceless would apply to the PHON structure of a *word* object, but not to the PHON structure of a *phrase* object. Similarly, constraints operating across word boundaries would not say anything about the PHON structure of a word object.

As a result, for a situation like the exemplary interaction between Polish final devoicing and voice assimilation processes (2a and 2b), we are left with a choice of either predicting different phonological structures for different levels of syntactic and morphological representation, or postulating that all surface representations at all levels have to be the same. Höhle (1999) appears to use (presumably for simplification) the first case scenario, and in his representations, different phoneme sorts (used for contrastive voicing) appear at the level of the word and at the level of an utterance. Applying this to our Polish devoicing example would yield a situation in which the phrase "kod dostępu" would have a PHON listing [kɔd dɔstɛmpu], but its first daughter element would still display a structure ending in a voiceless obstruent: [kɔt].

Such a solution makes it possible to account for predictions made at different levels, but causes problems with principle (c), that is, it requires a separate system for appending daughter elements together (since we cannot simply append [kɔt] and [dɔstɛmpu] to get [kɔd dɔstɛmpu]. Again, introducing underlying representation seems to be an advantage here, as it does not require clearly defined and sorted phonemes (which would be superfluous, Shoun 2005, 4.2), but allows forms to combine precisely because higher level structures are appended based on the underlying structure of their elements, while the surface structure may be separately predicted.

2.1 Summary of Proposals

To summarise - a system I propose is a system where the underlying and the surface forms are stored separately, where the higher level lists are appended separately (from underlying and surface lists of daughter elements), and where the underlying and surface forms can interact through formulated rules and constraints. The following section shall encompass all the major technical details of such an approach.

The approach presented here uses two levels of phonological representation, but is otherwise non-transformational and does not rely on rule ordering. The notion of underlying representation goes way back to transformational phonology, but can be found also as a solid basis in more recent phonological theories, most notably Optimality Theory. Applied in HPSG, it would not produce the surface forms through ordered rules or evaluating a number of universal constraints, but by allowing constraints that relate the surface representation to the underlying one. Constraints could be formed involving either of the representations (underlying and surface), but the surface representation could be restricted to depend on the underlying representation in cases where it cannot be arrived at purely through surface level constraints.

The way the relationship between the underlying representation and the surface representation operates essentially resembles the core ideas of Optimality Theory (Prince and Smolensky 1993/2004, 1.2), where surface representations are selected depending on the criteria of surface well-formedness (markedness) and closeness to the underlying representations (faithfulness). However, in this HPSG-based approach, the relationship is unambiguous and rather than relying on an algorithm selecting the most favourable form according to a universal constraint hierarchy (as is the case in OT), the surface forms are predicted based on language-specific, global constraints.

The notion of the underlying representation adopted here adheres to that in Shoun (2005, 4.2), ie. there is no clearly defined "list" of underlying phonemes for a language. Segments which show no productive alternations within a stem can be represented as identical to the surface form, disregarding baroque historical recreations. Finally, some of the core structural and technical sides of this approach are based on Bird's original proposals (Bird 1995).

2.2 Organising Segments

Although phonology in HPSG is traditionally handled through lists, the solution I propose is to use a new type of object, which I term here *segs* (for "segments"). This seemingly bizarre decision is dictated by the aforementioned principles: in order to coordinate the PHON values of utterances, phrases, words, etc., and at the same time allow constraints to operate at different levels, *segs* can be divided into subtypes, ie. *utterance-segs*, *phrase-segs*, *word-segs*, etc. Furthermore, due to the implementation of underlying representation, *segs* contains list features (UR-LIST and SR-LIST) for coordinating daughter elements, similar to DTR-LIST used by HPSG phrases.

The structure of *segs* would look like the following:

$$(5)\ \begin{bmatrix} \textit{segs} & \\ \text{SR-LIST} & \textit{list} \\ \text{UR-LIST} & \textit{list} \\ \text{FIRST} & \textit{ph-str} \\ \text{REST} & \textit{segs} \vee \textit{e-list} \end{bmatrix}$$

(*ph-str* here stands for "phonetic structure", and corresponds to the structure used to express the relationship between the UR and the SR, ie. one-to-one, one-to-many, or one-to-none)

While the FIRST feature of *segs* always has to be a phonetic structure, the REST can either be another *segs* or an empty list. Such a selection of REST value is not the only option: my original concept was to allow either *segs* or *ph-str*, in a manner resembling how phrases and words (or morphemes) are handled in HPSG syntax. However, such an approach requires us to either formulate numerous constraints twice, or introduce a phonological equivalent of Head Feature Principle. It is thus easier to settle down for ending all final *segs* in an empty list. While this adheres to the conventional way of handling list-like objects, it may cause its own problems with implementation by demanding an object which, in traditional HPSG ontology, belongs to an entirely different class than *segs*. One more possibility would be to replace *e-list* with a new, feature-empty subtype of *ph-str*, but for simplicity's sake I will just use the familiar *e-list* throughout the paper for *segs* and other list-like structures.

(in reality, the detailed structuring of *segs* is not as crucial as it seems, because most phonological structures can be introduced into the lexicon by specifying just UR-LIST, as shall be seen further on)

Last but not least, note that the features FIRST and REST are named after lists. In reality, HPSG ontology would demand these to be named distinctly in order to differentiate *segs* from regular lists: S-FIRST and S-REST are one of the possibilities, but I will use FIRST and REST throughout, again, for the sake of simplicity.

As an example of subtypes, in this model, the PHON feature of the word object will be an object of type *word-segs*, whose non-empty daughter elements will also have to be *word-segs*. But if the *word* object and eg. another *word* object get appended into a higher level *phrase* object, the PHON value of that phrase will be composed of *phrase-segs* objects.

segs → *word-segs* ∨ *phrase-segs* ∨ *utterance-segs* ... etc.

(6a) $\textit{word} \rightarrow \begin{bmatrix} \textit{word} \\ \textsc{phon} \mid \textsc{seg-list} \ \ \textit{word-segs} \end{bmatrix}$

(6b) $\begin{bmatrix} \textit{word-segs} \\ \textsc{rest} \ \ \textit{segs} \end{bmatrix} \rightarrow \begin{bmatrix} \textit{word-segs} \\ \textsc{rest} \ \ \textit{word-segs} \end{bmatrix}$

The structures for every level of the sentence will be, thus, different, but the constraints operating on UR-LIST and SR-LIST will demand that the actual phonetic representations in the daughter elements (stored in eg. *word-segs*: UR-LIST) be carried over and appended into mother elements (eg. *phrase-segs*: UR-LIST).

The correspondences between the underlying and the surface representations are handled through an object of type *ph-str* ("phonetic structure"), of which I propose three subtypes:

$$\text{(8) } \textit{ph-str} \rightarrow \begin{bmatrix} \textit{simple} \\ \textsc{ur} \ \ \textit{rep} \\ \textsc{sr} \ \ \textit{rep} \end{bmatrix} \vee \begin{bmatrix} \textit{complex} & \\ \textsc{ur} \ \ \textit{rep} & \\ \textsc{sr} & \begin{bmatrix} \textit{complex-rep} & \\ \textsc{sr-list} \ \ \textit{list} & \\ \textsc{first} & \textit{rep} \\ \textsc{rest} & \textit{complex-rep} \vee \textit{e-list} \end{bmatrix} \end{bmatrix} \vee \begin{bmatrix} \textit{empty} \\ \textsc{ur} \ \ \textit{rep} \end{bmatrix}$$

(*rep* stands for "representation": the actual phonetic description of features)

The *simple* object corresponds to the casual scenario where one underlying form corresponds to one uttered segment ("phone"). The *complex* object accounts for epenthesis, a case where one underlying phoneme corresponds to a more complex phonetic structure of two or more segments. Finally, the *empty* object accounts for deletion, ie. a situation the underlying segment is not visible on the surface at all. The two latter objects will be seen in action in section 2.2. on opacity in Turkish.

The formulation rules of UR-LIST and SR-LIST are expressed through the following constraints (for all three subtypes respectively):

$$\text{(7a) } \begin{bmatrix} \textit{segs} & \\ \textsc{first} & \textit{simple} \\ \textsc{rest} & \textit{segs} \end{bmatrix} \rightarrow \begin{bmatrix} \textit{segs} & \\ \textsc{ur-list} \ \ \left\langle \boxed{1}, \boxed{3} \right\rangle & \\ \textsc{sr-list} \ \ \left\langle \boxed{2}, \boxed{4} \right\rangle & \\ \textsc{first} & \begin{bmatrix} \textit{simple} \\ \textsc{ur} \ \ \boxed{1} \\ \textsc{sr} \ \ \boxed{2} \end{bmatrix} \\ \textsc{rest} & \begin{bmatrix} \textit{segs} \\ \textsc{ur-list} \ \ \boxed{3} \\ \textsc{sr-list} \ \ \boxed{4} \end{bmatrix} \end{bmatrix}$$

(7b) $$\begin{bmatrix} \textit{segs} & \\ \textsc{first} & \textit{simple} \\ \textsc{rest} & \textit{e-list} \end{bmatrix} \longrightarrow \begin{bmatrix} \textit{segs} & \\ \textsc{ur-list} & \langle \boxed{1} \rangle \\ \textsc{sr-list} & \langle \boxed{2} \rangle \\ \textsc{first} & \begin{bmatrix} \textit{simple} & \\ \textsc{ur} & \boxed{1} \\ \textsc{sr} & \boxed{2} \end{bmatrix} \end{bmatrix}$$

(8a) $$\begin{bmatrix} \textit{segs} & \\ \textsc{first} & \textit{complex} \\ \textsc{rest} & \textit{segs} \end{bmatrix} \longrightarrow \begin{bmatrix} \textit{segs} & \\ \textsc{ur-list} & \langle \boxed{1}, \boxed{3} \rangle \\ \textsc{sr-list} & \langle \boxed{2}, \boxed{4} \rangle \\ \textsc{first} & \begin{bmatrix} \textit{complex} & \\ \textsc{ur} & \boxed{1} \\ \textsc{sr} \mid \textsc{sr-list} & \boxed{2} \end{bmatrix} \\ \textsc{rest} & \begin{bmatrix} \textit{segs} & \\ \textsc{ur-list} & \boxed{3} \\ \textsc{sr-list} & \boxed{4} \end{bmatrix} \end{bmatrix}$$

(8b) $$\begin{bmatrix} \textit{segs} & \\ \textsc{first} & \textit{complex} \\ \textsc{rest} & \textit{e-list} \end{bmatrix} \longrightarrow \begin{bmatrix} \textit{segs} & \\ \textsc{ur-list} & \langle \boxed{1} \rangle \\ \textsc{sr-list} & \langle \boxed{2} \rangle \\ \textsc{first} & \begin{bmatrix} \textit{complex} & \\ \textsc{ur} & \boxed{1} \\ \textsc{sr} \mid \textsc{sr-list} & \boxed{2} \end{bmatrix} \end{bmatrix}$$

(9a) $$\begin{bmatrix} \textit{segs} & \\ \textsc{first} & \textit{empty} \\ \textsc{rest} & \textit{segs} \end{bmatrix} \longrightarrow \begin{bmatrix} \textit{segs} & \\ \textsc{ur-list} & \langle \boxed{1}, \boxed{2} \rangle \\ \textsc{sr-list} & \langle \boxed{3} \rangle \\ \textsc{first} & \begin{bmatrix} \textit{empty} & \\ \textsc{ur} & \boxed{1} \end{bmatrix} \\ \textsc{rest} & \begin{bmatrix} \textit{segs} & \\ \textsc{ur-list} & \boxed{2} \\ \textsc{sr-list} & \boxed{3} \end{bmatrix} \end{bmatrix}$$

(9b) $$\begin{bmatrix} \textit{segs} & \\ \textsc{first} & \textit{empty} \\ \textsc{rest} & \textit{e-list} \end{bmatrix} \longrightarrow \begin{bmatrix} \textit{segs} & \\ \textsc{ur-list} & \langle \boxed{1} \rangle \\ \textsc{sr-list} & \langle \rangle \\ \textsc{first} & \begin{bmatrix} \textit{empty} & \\ \textsc{ur} & \boxed{1} \end{bmatrix} \end{bmatrix}$$

While it would be possible to group UR-LIST and SR-LIST into a single list, thus simplifying the system, dividing them has an advantage visible particularly in computational implementations: the UR-LIST and SR-LIST objects contain the clearest linear phonetic representation of a particular utterance, phrase or word, which can be invoked to generate the entire structure for the word's PHON. For example, specifying a word's SR-LIST is enough to predict structures for all the possible lexical items with that surface representation, in a manner in which specifying PHON is traditionally used.

The UR-LIST and SR-LIST features are among the more important ones in this framework: they are used to coordinate generated structures, most importantly appending daughter phonologies to the PHON of mother objects: words into phrases, etc. Using them, rather than simple concatenation of entire PHON structures, allows for using different subtypes of *segs* for different syntactic objects and restricting rules to various levels of the sentence structure. While the PHON structures of words and phrases can be composed of different objects, the core phonological representations are required to be the same. Such an approach combines principles (a) and (c) mentioned in the beginning of the second section.

With these general foundations of the framework in mind, below is an exemplary PHON structure provided according to my proposals for the English word "cat":

$$
(10)\quad \begin{bmatrix} \textit{word} & \\ \textsc{phon} \mid \textsc{seg-list} & \begin{bmatrix} \textit{word-segs} & \\ \textsc{ur-list} & \langle [\text{k}], [\text{\ae}], [\text{t}] \rangle \\ \textsc{sr-list} & \langle [\text{k}], [\text{\ae}], [\text{t}] \rangle \\ \textsc{first} & \begin{bmatrix} \textit{simple ur} & [\text{k}] \\ \textsc{sr} & [\text{k}] \end{bmatrix} \\ \textsc{rest} & \begin{bmatrix} \textit{word-segs} & \\ \textsc{ur-list} & \langle [\text{\ae}], [\text{t}] \rangle \\ \textsc{sr-list} & \langle [\text{\ae}], [\text{t}] \rangle \\ \textsc{first} & \begin{bmatrix} \textit{simple ur} & [\text{\ae}] \\ \textsc{sr} & [\text{\ae}] \end{bmatrix} \\ \textsc{rest} & \begin{bmatrix} \textit{word-segs} & \\ \textsc{ur-list} & \langle [\text{t}] \rangle \\ \textsc{sr-list} & \langle [\text{t}] \rangle \\ \textsc{first} & \begin{bmatrix} \textit{simple} & \\ \textsc{ur} & [\text{t}] \\ \textsc{sr} & [\text{t}] \end{bmatrix} \\ \textsc{rest} & \textit{e-list} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}
$$

As seen above, the *segs* hierarchy is introduced as a feature of SEG-LIST ("segment list"), and not PHON directly. While SEG-LIST is used for linear phonology (and rules operating on segments), other structures can be introduced into the framework, eg. SYL-LIST used for syllables, similar to the solutions introduced in Bird and Klein (1994). This expansion, though possible, will not be covered in this paper.

As can be also seen, the subtype of a *segs* object used above is *word-segs*. The reason for the division of *segs* into various subtypes is to allow formulating rules pertaining to a particular level of representations, as mentioned before. To provide a short example, in English - in much simplified terms - we can postulate a constraint, working on the level of the word, demanding that all surface obstruent clusters have to agree in voicing (in actuality, that would be true for English only for word-final clusters). We restrict this constraint to the word level by evoking *word-segs*:

(11) $\begin{bmatrix} \textit{word-segs} & \\ \text{FIRST} \mid \text{SR} & \textit{obs} \\ \text{REST} \mid \text{FIRST} \mid \text{SR} & \textit{obs} \end{bmatrix} \rightarrow \begin{bmatrix} \textit{word-segs} & \\ \text{FIRST} \mid \text{SR} \mid \text{VOICE} & \boxed{1} \\ \text{REST} \mid \text{FIRST} \mid \text{SR} \mid \text{VOICE} & \boxed{1} \end{bmatrix}$

With such a constraint formulated, the form [kæts] will be well-formed, while [kætz] will violate the constraint. However, because the context is restricted to the word level, the phrase [kæts, doʊnt, flaɪ] is fine, even though the cluster [sd] occurs across the word boundary in the *phrase* object's PHON:SEG-LIST:SR-LIST.

2.3 Final Remarks

The system proposed here is an approach to HPSG phonology in which segments are described dually in terms of their underlying and surface features, and phonological phenomena are handled through constraints restricting or relating the two. Specifying the underlying representation in lexical items would allow us further to leave the surface representation entirely unspecified, thus the possible transformations of the phonemes (such as the ones in example (1)) can be handled purely through constraints.

The simplest constraint linking the surface representation to the underlying one would demand that the SR be identical with the UR. This would, of course, only work in all contexts for an ideal language with no phonological rules (dream on?):

(12) *ph-str* $\rightarrow \begin{bmatrix} \textit{ph-str} & \\ \text{UR} & \boxed{1} \\ \text{SR} & \boxed{1} \end{bmatrix}$

Actual applications of the UR/SR distinctions will be demonstrated in the following sections using more specific examples, primarily the aforementioned Polish voicing phenomena.

Here I would like to remark that, to focus on the general ideas of this approach, I will not go into the topic of representing individual phonemes in terms of features, and my examples will be only as complex as necessary. For the exhaustive analysis for representing phonemes, consult Bird and Klein (1994), Bird (1995, ch. 4) and Höhle (1999). For example, in my analysis, VOICE will the the feature of *rep* (representation) directly, without introducing divisions such as LARYNGEAL/SUPRALARYNGEAL.

3 Word Final Obstruent Devoicing Meets Obstruent Voice Assimilation

The analysis in this section is based around the data and processes in (2), with the goal of adequately describing Polish obstruent voicing processes through HPSG constraints. As mentioned before, there are three elements of the process:

1. Obstruents before other obstruents, including across word boundaries, assimilate their voice to that of the following obstruent, regressively (obstruent clusters have to agree in voicing).
2. Obstruents before sonorants, but not across word boundaries, retain their underlying, distinctive voice.
3. Word-finally, voiced obstruents become voiceless before sonorants or a pause (all word-final obstruents must be voiceless before sonorants or a pause).

The above is true for mainstream Polish, but in south-western variants, the voicing context may be different (Höhle 1999). This will not be dealt with here, although the provided example may easily be altered to account for different voicing phenomena.

The phenomenon of word-final devoicing (3) seemingly acts at the level of the word. However, because it can be "overridden" by voice assimilation (1), the two processes are intertwined and both have to be dealt with at the level of the utterance (thus, the objects *utterance* and *utterance-segs* will be involved).

Before dealing with the rule on the global level, we can use the examples in (2a) as evidence for this phenomenon. We can establish the underlying structure structure of /d/, in simplified terms, as the following (the voiced form it takes in the intervocalic position):

(13) $\begin{bmatrix} \textit{rep} & \\ \text{UR} & \begin{bmatrix} \textit{obs} & \\ \text{MANNER} & \textit{obstruent} \\ \text{PLACE} & \textit{coronal} \\ \text{VOICE} & + \end{bmatrix} \end{bmatrix}$

While the feature MANNER is specified here as *obstruent*, later on I will use the object *obs* to stand for a group of *rep* objects with MANNER: *obstruent*. Höhle (1999), in fact, uses the manner of articulation as a basis for subtypes of segments, eliminating the feature MANNER, while indeed rendering *obs* an existing object.

As mentioned before, because the surface representation can be generated through constraints, declaring anything about the surface structure of the phoneme would be superfluous. Using the same descriptive structure for the underlying and the surface representation, in terms of features (as opposed to alternating between phonemic sorts, cf. Höhle 1999, 3.25) allows us to work with, for any phonological phenomenon, only the features in question, and also organise phonemes into natural groups by invoking the features defining the group. While /d/ in Polish can undergo alternations of its place and manner of articulation (potentially becoming /dʑ/ or, in its devoiced version, /tɕ/), I am going to focus solely on the phenomenon of voicing here, and so, only the feature VOICE will be relevant.

Also note that in this example, it is assumed that all representations are *simple* objects. While epenthesis and deletion exist in Polish, they are, again, beyond the scope of this example and the related constraints can be easily produced by comparing both the global constraints described in the previous section and the examples from the following one.

In Polish, we are dealing with a situation where sometimes the VOICE values of obstruents are determined by the phonological context (before other obstruents, at the end of the phonological word), and sometimes by faithfulness to the UR (everywhere else).

First, we can attempt to translate the rule of voice assimilation:

$$(14)\ \begin{bmatrix} \textit{utterance-segs} & \\ \textsc{first} \mid \textsc{sr} & \textit{obs} \\ \textsc{rest} \mid \textsc{first} \mid \textsc{sr} & \textit{obs} \end{bmatrix} \lor \begin{bmatrix} \textit{utterance-segs} & \\ \textsc{first} \mid \textsc{sr} \mid \textsc{voice} & \boxed{1} \\ \textsc{rest} \mid \textsc{first} \mid \textsc{sr} \mid \textsc{voice} & \boxed{1} \end{bmatrix}$$

This works like the English example previously used - any segment consisting of two obstruents found in an utterance must have a singular value of *voice* in the SR.

Now we have to formulate a constraint determining the voice of an obstruent in any other case (before a sonorant or a pause). We can, for example, formulate the following:

$$(15)\ \begin{bmatrix} \textit{utterance-segs} & \\ \textsc{first} & \begin{bmatrix} \textit{simple} & \\ \textsc{sr} & \textit{obs} \end{bmatrix} \\ \textsc{rest} & \neg \begin{bmatrix} \textsc{ur} & \textit{obs} \end{bmatrix} \end{bmatrix} \rightarrow \begin{bmatrix} \textit{utterance-segs} & \\ \textsc{first} \mid \textsc{sr} \mid \textsc{voice} & \boxed{1} \\ \textsc{first} \mid \textsc{sr} \mid \textsc{voice} & \boxed{1} \end{bmatrix}$$

The above constraint demands that if any obstruent followed by a non-obstruent is found in an utterance, its surface voicing has to be equivalent to its underlying voicing. This is, however, imprecise for Polish, since the actual situation is like that only in the word-final context. But neither can we restrict the context to word-final in any *word-segs*, because that would demand every single final obstruent to be voiceless, even if the following obstruent is voiced, which would lead to violating the previous formulated constraint (14).

One way of solving the situation would be to expand the structure of a phoneme with a class of morphology-related features, eg. determining if it is a word-final segment or not. This new information is introduced into the *simple* object as a feature NP (for "non-phonetic"):

$$(16)\ \begin{bmatrix} \textit{simple} & \\ \textsc{ur} & \textit{rep} \\ \textsc{sr} & \textit{rep} \\ \textsc{np} \mid \textsc{word-final} & \textit{boolean} \end{bmatrix}$$

With this expansion, we can determine all word-final obstruents in an utterance. We first need to introduce rules to determine the value of feature WORD-FINAL:

$$(17a)\ \begin{bmatrix} \textit{word-segs} & \\ \textsc{first} & \textit{simple} \\ \textsc{rest} & \textit{e-list} \end{bmatrix} \longrightarrow \begin{bmatrix} \textit{word-segs} & \\ \textsc{first} \mid \textsc{np} \mid \textsc{word-final} & + \end{bmatrix}$$

$$(17b)\ \begin{bmatrix} \textit{word-segs} & \\ \textsc{first} & \textit{simple} \\ \textsc{rest} & \textit{segs} \end{bmatrix} \longrightarrow \begin{bmatrix} \textit{word-segs} & \\ \textsc{first} \mid \textsc{np} \mid \textsc{word-final} & - \end{bmatrix}$$

Now, we can restate the previous obstruent devoicing rule to include the new information about word-final segments, but operate at the utterance level:

$$(18a)\ \begin{bmatrix} \textit{utterance-segs} & \\ \textsc{first} & \begin{bmatrix} \textit{simple} & \\ \textsc{sr} & \textit{obs} \\ \textsc{np} \mid \textsc{word-final} & + \end{bmatrix} \\ \textsc{rest} & \neg \begin{bmatrix} \textsc{ur} & \textit{obs} \end{bmatrix} \end{bmatrix} \longrightarrow \begin{bmatrix} \textit{utterance-segs} & \\ \textsc{first} \mid \textsc{sr} \mid \textsc{lg} \mid \textsc{voice} & - \end{bmatrix}$$

$$(18b)\ \begin{bmatrix} \textit{utterance-segs} & \\ \textsc{first} & \begin{bmatrix} \textit{simple} & \\ \textsc{sr} & \textit{obs} \\ \textsc{np} \mid \textsc{word-final} & - \end{bmatrix} \\ \textsc{rest} & \neg \begin{bmatrix} \textsc{ur} & \textit{obs} \end{bmatrix} \end{bmatrix} \longrightarrow \begin{bmatrix} \textit{utterance-segs} & \\ \textsc{first} \mid \textsc{ur} \mid \textsc{lg} \mid \textsc{voice} & \boxed{1} \\ \textsc{first} \mid \textsc{sr} \mid \textsc{lg} \mid \textsc{voice} & \boxed{1} \end{bmatrix}$$

In this case, any word-final obstruent before a non-obstruent in a complete utterance is devoiced, and any non-final obstruent retains its underlying value of VOICE.

This way also, the value of the final obstruent in the cluster is always predicted, allowing the preceding obstruents to assimilate their voice to it in order to retain the constraint (14).

Do note that we can establish other morphology-related features in NP, like a feature STEM-FINAL. In the above examples, the features are encoded into *simple* structure, and would have to operate differently when introduced into a *complex* object, though the above case should suffice for the presented example at least.

4 The Issue of Opacity in Turkish

The purpose of this section is twofold: to return to the processes of epenthesis and deletion, and to demonstrate how accounting for underlying forms can be used to deal with opacity-related issues (rule interaction at more than one level), usually problematic in monostratal frameworks.

In Turkish (based on data for the OT analysis in Sanders 2003, 5.3), two separate processed occur: 1. consonant clusters are broken through epenthesis, and 2. /k/ is deleted intervocalically when a suffix beginning with a vowel follows. However, in the case where /k/ is followed by a consonant, the two take effect at the same time: /k/ is deleted, but triggers epenthesis nonetheless:

1. /bɑʃ/ + /m/ → [bɑʃɯm]
2. /ɑjɑk/ + /ɯ/ → [ajaɯ]

But: 3. /ɑjɑk/ + /m/ → [ɑjɑɯm]

The above description is somehow simplified - there are lexical exceptions to this rule, and the /k/ may not completely disappear (in some contexts it may become another consonant, most importantly /j/ before front vowels, or its deletion may lengthten the preceding vowel) - this also depends on dialectal variation. However, here, the case of complete disappearance is assumed, mainly to account for destructive processes in which ghost segments are undesired.

One way to account for such a process in HPSG is to invoke morphology, which is indeed a viable solution. However, I will attempt to demonstrate that with the notion of the underlying representation, HPSG can handle such cases of opacity purely through phonological constraints.

In a typical monostratal framework, introducing the constraints prohibiting both consonant clusters and [k] before vowels could, possibly, lead to a situation where neither [ɑjɑkɯm] nor [ɑjɑkm] are considered well-formed while the form [ɑjɑɯm] is, but, first of all, we would have no way to arrive at it, and, more importantly, where any cluster of two vowels would have to be acceptable, while in Turkish, that is not exactly the case - the vowel clusters, aside from borrowings, emerge almost uniquely from the deletion of /k/. To account for this fact, the framework would have to postulate the presence, but not articulation, of /k/, as a ghost segment in the cases where it is deleted, but still present for the purpose of epenthesis.

Within the framework presented, it is possible to eliminate the need for such non-surfacing phonemes by translating the two rules (epenthesis and k-deletion) to involve two different levels of representation. This should not be confused with "ordering" the rules, as the two constraints apply simultaneously, but take into account the UR and the SR separately.

To begin with, epenthesis can be formulated by stating that a consonant (here, for simplification, just /m/) can surface either as a single phone, or as a segment:

(19) $$\begin{bmatrix} \textit{ph-str} \\ \textsc{ur}\ \ [\text{m}] \end{bmatrix} \rightarrow \begin{bmatrix} \textit{simple} \\ \textsc{sr}\ \ [\text{m}] \end{bmatrix} \vee \begin{bmatrix} \textit{complex} & \\ \textsc{sr} & \begin{bmatrix} \textit{complex-str} & \\ \textsc{first} & [\text{ɯ}] \\ \textsc{rest} & [\text{m}] \end{bmatrix} \end{bmatrix}$$

Now, because epenthesis takes into account the underlying structure of the word, we can translate this rule into the framework by formulating a constraint demanding than any underlying consonant must be followed by a surface vowel:

(20) $$\begin{bmatrix} \textit{segs} \\ \textsc{first} \mid \textsc{ur}\ \ \textit{cons} \end{bmatrix} \rightarrow$$

$$\begin{bmatrix} \textit{segs} \\ \textsc{rest}\ \ \textit{e-list} \end{bmatrix} \vee \begin{bmatrix} \textit{segs} \\ \textsc{rest} \mid \textsc{first} \mid \textsc{sr}\ \ \textit{vow} \end{bmatrix} \vee \begin{bmatrix} \textit{segs} \\ \textsc{rest} \mid \textsc{first} \mid \textsc{sr} \mid \textsc{first}\ \ \textit{cons} \end{bmatrix}$$

The context for k-Deletion is the occurrence of a following vowel (in simplified terms, again). Therefore, the necessary constraint would demand than any underlying /k/ followed by a vowel must not surface, ie. be an *empty* subtype *segs* object, appending nothing to the SR-LIST:

(21) $$\begin{bmatrix} \textit{segs} & \\ \textsc{first} \mid \textsc{ur} & [\text{k}] \\ \textsc{rest} \mid \textsc{first} \mid \textsc{sr} & \textit{vowel} \end{bmatrix} \vee \begin{bmatrix} \textit{segs} & \\ \textsc{first} \mid \textsc{ur} & [\text{k}] \\ \textsc{rest} \mid \textsc{first} \mid \textsc{sr} \mid \textsc{first} & \textit{vowel} \end{bmatrix} \rightarrow \begin{bmatrix} \textit{segs} \\ \textsc{first}\ \ \textit{empty} \end{bmatrix}$$

The final result of the two constraints (20 & 21) is that the only permissible situation is the one where deletion and epenthesis co-occur. The presence of an underlying /k/ triggers an epenthesis, but the /k/ does not surface itself, because it is followed by a surface vowel. The presence of a non-surfacing /k/ may also be used to formulate a constraint demanding the aforementioned vowel lengthening.

5 The Conclusion

With the presented examples and the description, I hope to have shown that it is possible to have a functional phonological framework utilising underlying forms in HPSG, which would tackle neutralisation and opacity without going into arbitrary complexity. Although other proposals for handling phonology in HPSG exist and, indeed, are constantly being developed, the approach presented here aims to be widely applicable and resolve phonetic alternations on purely phonological grounds, while still leaving a lot of space for expansion and not being detached from the structures of morphology and syntax.

References

Bird, Steven. 1995. *Computational Phonology: A Constraint-Based Approach*. Studies in Natural Language Processing, Cambridge.

Bird, Steven and Klein, Ewan. 1994. Phonological Analysis in Typed Feature Systems. *Computational Linguistics* 20, 455–491.

Bonami, Olivier and Boyé, Gilles. 2006. Deriving Inflectional Irregularity. In Stefan Müller (ed.), *The Proceedings of the 13th International Conference on Head-Driven Phrase Structure Grammar*, pages 361–380, Stanford: CSLI Publications.

Höhle, Tilman N. 1999. An Architecture for Phonology. In Robert D. Borsley and Adam Przepiórkowski (eds.), *Slavic in Head-Driven Phrase Structure Grammar*, pages 61–90, Stanford, CA.

Prince, Alan and Smolensky, Paul. 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell.

Rubach, Jerzy. 1982. *Analysis of Phonological Structures*. Studies in Natural Language Processing, Warsaw: Pañstwowe Wydawnictwo Naukowe.

Sanders, Nathaniel R. 2003. *Opacity and Sound Change in the Polish Lexicon*. Ph. D.thesis, University of California, Santa Cruz.

Shoun, Abby. 2005. Interpreting and generating formal grammars: A study of HPSG phonology. Honors Thesis: University of North Carolina.

Phonological change and grammaticalization in HPSG: The case of French final consonants

Jesse Tseng

Cognition Langues Langage Ergonomie
Équipe de recherche en syntaxe et sémantique
CLLE-ERSS UMR 5263
CNRS & University of Toulouse

Proceedings of the HPSG09 Conference

Georg-August-Universität Göttingen, Germany

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

This paper explores the use of HPSG for modeling historical phonological change and grammaticalization, focusing on the evolution of the pronunciation of word-final consonants in Modern French. The diachronic evidence is presented in detail, and interpreted as two main transitions, first from Old French to Middle French, then from Middle French to the modern language. The data show how the loss of final consonants, originally a phonological development in Middle French, gave rise to the grammaticalized external sandhi phenomenon known as consonant liaison in modern French. The stages of development are analyzed formally as a succession of HPSG lexical schemas in which phonological representations are determined by reference to the immediately following phonological context.

1 Introduction

The prevalence of silent final consonants is a striking feature of French orthography. Even English speakers with no direct knowledge of French may be aware of this, if they know the approximate spelling and pronunciation of familiar loan words such as those in (1a). On the other hand, the equally familiar examples in (1b) show that final consonants are pronounced in some French words.

(1) a. s'il vous plaît [si(l)vuplɛ], merci beaucoup [mɛʁsiboku], rendez-vous [ʁɑ̃devu], faux pas [fopɑ], coup d'état [kudeta]
b. cul-de-sac [sak], bonjour [bɔ̃ʒuʁ], apéritif [apeʁitif], Noël [nɔɛl]

As we will see in more detail below, the final consonants in all of these words correspond to sounds that were pronounced in older stages of French, but were then subject to a process of deletion that targeted different consonants and different series of words to varying degrees. The resulting distribution of pronounced vs. silent consonants was further complicated by normative pressure and orthographic influences ("spelling pronunciation"), as well as analogical tendencies, with significant but haphazard effects.

The preservation of silent final consonants in French orthography is thus motivated by historical considerations, and enforces distinctions in writing that are no longer made in speech. For instance, the singular and plural nouns in (2a) and the verb forms in (2b) all share the same pronunciation:

(2) a. cou/cous "neck(s)", coup/coups "strike(s)", coût/coûts "cost(s)" [ku]
b. couds/coud [ku] (from *coudre* "sew")

More significantly, the consonants in question may be silent in some contexts but pronounced in others, giving rise to synchronically active $\emptyset{\sim}$C alternations. First, the addition of a (vowel-initial) inflectional or derivational suffix can "reactivate"

[†]This research was undertaken as part of the PHONLEX project, directed by Jacques Durand, with support from the French National Research Agency (ANR).

the stem-final consonant. For example, the masculine singular adjective *bourgeois* is pronounced [buʁʒwa], but the feminine *bourgeoise* [buʁʒwaz] and the derived form *bourgeoisie* [buʁʒwazi] both contain a pronounced intervocalic "s" (realized as [z]). Further examples of this morphological alternation are shown in (3):

(3) découpage [dekupaʒ] *vs.* coup [ku]
    débutante [debytɑ̃t] *vs.* début [deby]

Similarly, and most importantly for our purposes, a normally silent final consonant may be pronounced in connected speech when followed immediately by a vowel-initial word. This $\emptyset{\sim}$C alternation is a well-known feature of French pronunciation known as "consonant liaison". The examples in (4) are all plural NPs in which both words carry the plural marker "s", which as illustrated in (2a) above is normally silent. However, because the second word in (4a) begins with a vowel, the liaison consonant [z] appears as a contextually-licensed phonological realization of plural marking. This [z] cannot appear in (4b), where the second word is consonant-initial.

(4) a. Champs-Élysées [ʃɑ̃zelize], États-Unis [etazyni]
    b. champs fleuris [ʃɑ̃fløʁi] ("flowery fields"),
       États Généraux [etaʒeneʁo] ("Estates-General")

Further examples of liaison in [z] can be found in the expressions *Beaux-Arts* [bozaʁ] and *vis-à-vis* [vizavi]. Other frequently occurring liaison consonants are [t] (*prêt-à-porter* [pʁɛtapɔʁte]) and [n] (*bon appétit* [bɔnapeti], vs *bon voyage* [bɔ̃vwajaʒ]).

The analysis of consonant liaison has been the subject of active debate, particularly in generative phonology. An early approach assumed underlying phonemic forms containing a final consonant, which was then deleted in the appropriate contexts — i.e., before another consonant or before a prosodic boundary — by a "truncation" rule (Schane, 1968). The data above can be dealt with using such a rule, which broadly speaking reproduces the historical evolution responsible for the modern forms. But not all cases of liaison can be adequately analyzed in terms of truncation, and more concrete approaches assuming representations closer to the surface forms can be shown to provide a more complete account of the data. The analyses proposed in the HPSG literature naturally tend to follow this surface-oriented, non-transformational approach. See for example Bonami and Boyé (2003) for the morphophonology of prenominal adjectives and Bonami et al. (2004) for the interaction of syntactic and phonological constraints in liaison.

From a diachronic perspective, however, the abstract representations and operations considered in more recent work to be unnecessary and unmotivated for synchronic analysis become the principal objects of study, corresponding to historically attested or reconstructed forms and their evolution over time. Analyses of language change are thus surface-oriented and transformational at the same time.

The hypothesis adopted in this paper is that while the grammar of a language can change radically within the space of a few generations, this global change is the sum of smaller, individual changes that can be modeled in terms of successive, overlapping alternative grammars corresponding to periods of variation (i.e. coexisting, competing analyses) eventually leading to reanalysis.[1]

2 Early developments

The sound changes that led to the development of French from Vulgar Latin have been extensively studied and are relatively well established, although authors often disagree on points of detail and chronology.[2] The loss of final consonants mentioned in the previous section as the source of consonant liaison began at the end of the Old French[3] period (from the 13th century onwards), but a number of changes from earlier periods are also relevant, and will be outlined here.

A major difference between Latin and French is the position of word stress. Polysyllabic words in Latin were stressed on the penultimate or antepenultimate syllable, while in French word stress falls on the final full syllable. This difference is not the result of massive stress shifts in French. In most words, the stress remained on the same syllable, but all following syllables at the end of the word were systematically reduced.

Already in Vulgar Latin, the antepenultimate stress pattern was largely eliminated by deletion of the post-tonic vowel. Later (but still in the pre-literary period of Gallo-Romance) final syllables were reduced: *a* was weakened to become the central vowel e̥ (the precusor of modern "mute *e*"), while most other vowels were deleted altogether.

(5) a. tábula > tábla > táble̥ "table"
    b. cólapu(m) > cólpo > colp "strike"

Final consonants (other than *m*) were preserved and did not stand in the way of the reduction/loss of final vowels:

(6) a. béllas > béle̥s "beautiful-*fpl*", béllos > bels "beautiful-*mpl*"
    b. pórtat > pórte̥t "[he] carries", ténet > tient "[he] holds"

Vowel deletion created many new final consonants and consonant clusters. These underwent devoicing (in the case of the obstruents d, θ, v, $z > t$, θ, f, s), and most clusters of three consonants were simplified by deleting the second element.

(7) a. novu(m) > neuf "new", grande(m) > grand > grant "large"
    b. cólapos > colps > cols "strikes", témpus > temps > tens "time"

[1]Cf. the approach to grammaticalization of Harris and Campbell 1995. For an earlier approach to formalizing reanalysis in HPSG to model syntactic change, see Bender and Flickinger (1999)).

[2]See e.g. Bourciez and Bourciez (1967); Fouché (1961); Zink (1986).

[3]The following abbreviations and approximate chronology are adopted in this paper: Old French "OFr" (10th–13th cent.), Middle French "MidFr" (14th–16th cent.), Modern French "ModFr" (17th cent. to present day).

These changes bring us to the period of the earliest surviving OFr literary texts (11th century). From this point on, we have textual evidence of the effects of sound changes in progress, keeping in mind that written forms are only an indirect representation of contemporary pronunciation.

The erosion of final consonant clusters continued, extending to sequences of two consonants. Various changes affected the first consonant in such clusters: vocalization of *l* to *u*, deletion of obstruents, but preservation of *r* and *n*. The second consonant was usually maintained. One effect of this change was that stem consonants were frequently deleted in favor of inflectional suffixes (*s* or *t*).

(8) a. vívere > vivre̥ "to live", vívo > vif "[I] live"
vs. vívit > vift > vit "[he] lives"

b. colp > coup "strike" *vs.* col(p)s > cous "strikes"
sáccu(m) > sac "sack" *vs.* sáccos > sacs > sas "sacks"

This had particularly significant consequences for nominal morphology, because the stem allomorphy in words like (8b) helped trigger a change in the status of the final stem consonant (see §4.2).

Single final consonants were maintained through the end of the OFr period, with one exception: *θ*. In most cases this consonant developed from an intervocalic *t* or *d* in Latin, which became word-final after the deletion of final vowels. The sound weakened and fell silent by the end of the 11th century:

(9) a. marítu(m) > mariθ > mari "husband"
nepóte(m) > neveuθ > neveu "nephew"

b. fíde(m) > feiθ > foi "faith"
mercéde(m) > merceiθ > merci "mercy"

Another important source of *θ* was the 3sg verb ending *-t*, which already showed signs of weakness in Vulgar Latin. Following the loss of final vowels in Gallo-Romance, this *t* was reinforced if it came into contact with another consonant. In such cases, *t* was maintained (10a), even if the reinforcing consonant was subsequently lost through cluster simplification. Following a vowel, on the other hand, *t* > *θ* was lost at the same time as the cases in (9).

(10) a. dórmit > dormt > dort "[he] sleeps",
présit > prist > prit "[he] took"

b. pórtat > pórte̥θ > pórte̥ "[he] carries"
dormí(vi)t > dormíθ > dormí "[he] slept"

The earliest OFr texts still contain written forms with final "t/d" (e.g. Paris and Pannier, 1872, pp. 97–99), but these letters rapidly disappeared from the orthography, except in exceptional cases like *et* "and". This consonant left no phonological traces in later stages of French. It should be mentioned, however, that orthographic "t" was later reintroduced analogically in some 3rd person verbs like *dormit* "[he] slept" and *fut* "[he] was", which then rejoined forms like (10a) where *t* had always been preserved.[4]

3 Middle French

In the second half of the 12th century, a new wave of deletions began, affecting all remaining final consonants. The process was a very gradual one, however, continuing through MidFr and beyond. For various reasons, these changes in pronunciation did not generally lead to stable orthographic changes. The relatively phonetic spelling of the 12th century began to lag behind the evolution of the language, and silent letters became an increasingly pervasive feature of French orthography.

3.1 Texts and evidence

The texts of this period do offer occasional indications of consonant loss, including the simple omission of the relevant letters (*troi* instead of *trois* "three", *naturé* instead of *naturel* "natural") or the substitution of non-etymological — presumably also silent — consonants (*coureux* for *coureur* "runner", *sant* for *sang* "blood").[5] Poetry is a particularly rich source of evidence, because the loss of final consonants made available many new pairs of rhyming words: *vert*:*vers*, *rechief*:*bouclier*.[6] But given the conservative nature of poetic pronunciation, such rhymes only became accepted long after the loss of the consonants in popular speech. Furthermore, the ends of verses constitute a highly specific prosodic context where words were not necessarily pronounced as they would have been in connected speech.

Contemporary metalinguistic descriptions confirm, in fact, that the pronunciation of final consonants varied according to the immediately following context. The practice of pronouncing the same written word in distinct ways depending on what follows, which as we saw in §1 is so characteristic of ModFr, was already in place, in some form, by the 13th century. One of the rare linguistic texts from this period, the *Orthographia Gallica*, contains the following rule: "Whenever a word beginning with a consonant follows a word ending in a consonant, the consonant of the preceding word must not be pronounced, even though it is written, for example *apres manger* must be pronounced *apre manger*."[7]

The available evidence points to a weakening of word boundaries in late OFr, such that a -C#C- boundary came to be treated like a medial consonant cluster, and thus subject to various simplification processes that had already left their mark within words in OFr. In general, the first consonant in such sequences was deleted, but different combinations were presumably affected at different times, and to varying degrees. One would expect, for instance, for final obstruents to have survived

[4] For the history of the analogical epenthetic *t* in inversion constructions like *Porte-t-il ?* "Does he carry?", see Tseng (2008).

[5] Fouché (1961, p. 663, 783), Brunot (1966, p. 430)

[6] Fouché (1961, p. 664, 783)

[7] Translation of Rule VIII, Stürzinger (1884, p. 17–18).

longer before words starting with "r" or "l", cf. the hypothetical example *petit rost* "small roast" of Morin (1986), in which the [t] should have been temporarily protected from deletion as part of the resyllabified complex onset [tr]. Unfortunately, the texts of this period do not allow us to reconstruct the progression of the sound changes to this level of detail, and by the end of MidFr, all consonant initial words constituted triggers for deletion of a preceding final consonant. Here are some examples from the earliest recognized grammars of French, dating from the mid-16th century (Thurot, 1883):

(11) a. sans cause, soubz couleur, ung combat tel, faictz plaisans, suis sayn
san cause, sou couleur, un comba tel, fai plaisans, sui sayn
"without cause, under color, such a combat, pleasant facts, [I] am healthy" (Palsgrave, 1530)

b. Les femmes sont bonnes "The women are good"
Lé femme son bones (Sylvius, 1531)

Not all final consonants were affected uniformly by this process. According to Palsgrave, *m*, *n*, and *r* were not deleted in preconsonantal contexts; the same three exceptions were already mentioned in the *Orthographica Gallica*. Other grammarians of Palsgrave's time say that final *r* was in fact deleted, at least in some words, such as infinitives in *-er*. On the other hand, they recommend the pronunciation of many consonants that Palsgrave says are silent (in particular *f*, *l*, and *c/q*). Upon closer examination, the increasingly abundant phonetic descriptions of this period (16th–17th cent.) contain many contradictory details, reflecting the different authors' individual opinions about a system that contained areas of instability and variation, for reasons that will be discussed in the following sections.

Before a vowel-initial word, consonants were preserved, as the weakening of the word boundary -C#V- led to the resyllabification of the consonant from coda to onset position. The fricatives [s] (written "s/x/z") and [f], which after the final cluster simplifications described in §2 occurred only after vowels or the sonorants *r/n*, underwent voicing to [z] and [v], respectively, in words like *cors* "body" and *vif* "alive". Final stops were not affected in this way, so for example final "t/d" retained the unvoiced pronunciation [t] before a vowel in words like *tout* "all" and *quand* "when". (With only minor modifications, this is still how liaison consonants are pronounced in ModFr.)

The following examples from Saint-Lien (1580) illustrate the preservation of final consonants in pre-vocalic contexts, which is obviously the origin of consonant liaison in ModFr (Thurot, 1883):[8]

(12) a. tout ainsi que tu fais aux autres "just as you do unto others"
tou tin si ke tu fai zau zautres

b. vous estes un homme de bien "you are a good man"
vou zeste zun nome de bien

[8]See also Livet (1859, p. 508).

The early grammarians mention isolated exceptions to this rule concerning prevocalic contexts. For example, as mentioned at the end of §2, the final "t" of *et* "and" was purely orthographic, and was never pronounced, even before a vowel. Such cases rapidly multiplied in ModFr, as the status of final consonants evolved.

Finally, we need consider the pronunciation of final consonants with no immediately following word. Given the hypothesis that the changes described above resulted from the weakening of word boundaries at the end of the OFr period, final consonants should not have been affected in these "pre-pausal" contexts. One manuscript of the *Orthographia Gallica* seems to confirm this: "But at the ends of sentences or in the middle of a sentence at a pause, [consonants] can be pronounced".[9] The grammarians of the 16th century maintained this general rule, as we can see from the ends of the examples in (11–12), but with many exceptions. For Palsgrave, for example, final *m*, *n*, *r*, and *s/x/z* were distinctly pronounced, but *c*, *f*, *l*, *p*, and *t* were "but remissely sounded" (13a). On the other ahnd, final *t* and *p* following *a/e* retained their full sound (13b).

(13) a. auec, soyf, fil, beaucoup, mot
aue, soy, fi, beaucou, mo
"with, thirst, thread, much, word"
b. chat, debat, ducat, combat, hanap, duvet, regret, entremet "dessert"

Again, there was much disagreement from author to author concerning individual words or series of words. The overall tendency in the transition to ModFr was for more and more final consonants to fall silent in pre-pausal contexts.

3.2 Summary of sound changes

The rather jumbled picture that the early grammars present is the product of new forces that were partially dismantling the phonetic changes of the preceding period. To summarize the earlier changes, recall that in OFr period, the erosion of final unstressed syllables created a rich inventory of final consonants, and consonant-final words had the same pronunciation in all syntactic/prosodic contexts. The MidFr sound changes described in the previous section affected final consonants according to the immediately following phonological context:

(14) a. Final consonants were lost before a following consonant-initial word.
b. Final consonants were preserved before a following vowel-initial word, with voicing of [s] (and [f]).
c. Final consonants were preserved before a pause.

This purely phonological formulation is an idealization that ignores syntactic and lexical conditions that are likely to have existed, but for which we have insufficient evidence. The process eventually extended to all consonants, but some words were affected much later than others, in particular several series of words ending in *r*,

[9] Translation of H79, Stürzinger (1884, p. 18).

l, and *f* (Fouché, 1961, p. 669–70). The process lost steam and gave way to other developments towards the end of the MidFr period (that is, before the 16th century and the publication of the first grammars).

After the application of (refconx), some words still had a single, context-independent pronunciation, e.g. those ending in a full vowel. But consonant-final words developed two distinct pronunciations: a long form, corresponding to the original, historical pronunciation and used in pre-pausal and pre-V contexts, and a short form, derived from the long form by truncation and used only in pre-C contexts. Words originally ending in [s], and some ending in [f], developed a distinct pre-V long form in [z]/[v], so three contextually-determined forms in all.

4 Modern French

In this section we trace the development of the ModFr pronunciation of final consonants as the new changes already visible in the grammatical descriptions of the 16th century took hold. The effects of these changes, which were not purely phonological in nature and were highly unpredictable, led to major changes in the inventory and distribution of contextually alternating forms.

4.1 From Middle French to Modern French

While the roots of ModFr consonant liaison are already clearly visible in the idealized system described in §3.2, the pronunciation of this stage differed from the ModFr system in several respects.

First, many words no longer have distinct contextual forms in ModFr. For example, the noun *coup* "strike" now always has a silent consonant, while the preposition *avec* "with" always has a pronounced final [k]. Second, for words that do still have distinct forms, their distribution is no longer determined exclusively by the following phonological context. In particular, the short form, originally restricted to pre-C contexts, is now often found before vowels, for example in cases of unrealized optional liaison (*toujours ici* "still here" [tuʒuʁisi]).

Finally, the pre-pausal context in ModFr has realigned with the pre-C context. This means, significantly, that the short form has become the form of the word used in isolation, i.e. the citation form. While the citation form does not necessarily reveal the basic or "underlying" phonological form, it does represent the core phonological content of the word, perceived by speakers as sufficient for its identification. As an example, the phonological form [boku] is recognized as the word *beaucoup* in ModFr, while [bokup] is a contextually restricted form that cannot be uttered in isolation. The situation at the end of MidFr, after the application of the changes in (14), was the opposite: the short form [boku] would have been unacceptable out of context, because the citation form of the word was [bokup].

The usage described in 16th century grammars does not exactly reflect the results of the MidFr sound changes in (14); the effects of further developments can already be observed at this time. Recall from (13), for example, that for Palsgrave, many final consonants were silent before a pause. The following transcription by H. Estienne (1582) gives an example of late MidFr pronunciation (Livet, 1859, p. 381–82):

(15) Vous me dites tousiours que vostre pays est plus grand de beaucoup et
Vou me dite touiours que votre pays est[?] plu gran de beaucoup e
plus abondant que le nostre, et que maintenant vous pourriez bien y viure
plus abondan que le notre, e que maintenan vou pourrie bien y viure
à meilleur marché que nous ne viuons depuis trois mois en ceste ville :
à meilleur marché que nou ne viuon depui troi mois en cete ville :
mais tous ceux qui en viennent parlent bien vn autre langage
mai tou ceux qui en viennet parlet bien vn autre langage

"You always tell me that your country is much larger and more abundant than ours, and that now you could live well there, more cheaply than we have been living for three months in this city: but all those who come from there speak another language"

Etienne's transcription is mostly consistent with the effects of the MidFr sound changes, keeping in mind that he does not indicate voicing alternations, and that it is not possible to distinguish pronounced and unpronounced nasals using his naive notation.

The final consonant is pronounced in *tousiours*, *ceux*, and *viennent*, although they are followed by consonants. These are not exceptions to truncation (14a), but instances of pre-pausal pronunciation (14c), reflecting the presence of prosodic boundaries before sentential complements and around relative clauses. Estienne explains that these consonants could be dropped in rapid speech. They must be dropped in ModFr.

The pronounced final consonants in *meilleur* and *parlent* do constitute exceptions to (14a). As mentioned above, words ending in *r* were among the last to be affected by truncation, and thus among the first to respond to normative and analogical influences working to revert the change. The [r] of *meilleur* was restored in pre-C contexts before the end of MidFr, and survived in ModFr. The pre-C pronunciation of *t* in *parlent* was also a normative reaction, to prevent the merger of singular and plural 3rd person verb forms. This pronunciation, unlike the previous one, was not adopted in ModFr. Note finally that if Etienne's transcription of *est* with two pronounced final consonants in pre-C context is accurate, it represents a completely artificial spelling pronunciation, recommended to my knowledge by no other grammarians, and totally abandoned in ModFr.

The transcribed passage contains no exceptions to the rule requiring the pronunciation of consonants in pre-V contexts (14b). We can see that liaison was more systematic at this stage, realized whenever the phonological context allowed it, for example in the sequences *pays est*, *beaucoup et*, and *mois en*. This pronunciation is no longer posible today, because ModFr imposes additional syntactic constraints on liaison (e.g., no liaison between subject and verb, no liaison after a prepositional phrase).

4.2 Contextual alternations in Modern French

This section examines in more detail the various ways in which contextually alternating forms were reorganized and reanalyzed in ModFr.

We have seen that the pronunciation of words like *toujours* and *beaucoup* before a pause changed in early ModFr, from the long forms [tujuʁs] and [bokup] to the short forms [tujuʁ] and [boku]. This could be seen as an extension of the MidFr sound change, with final consonant truncation spreading from pre-C to prepausal contexts. This cannot be the only explanation, however, for a number of reasons. First, contemporary descriptions do not document a process of gradual phonetic loss. It is true that, immediately after providing the transcriptions in (13a), Palsgrave writes, "how be it, the consonant shall have some lyttell sounde" (ch. 27). Similar recommendations continue into the 17th century: "Il ne faut pas la prononser trop distinctemant" (Dobert, 1650). It is unclear, however, just what the phonetic interpretation of such remarks should be. Other authors explicitly recognize the co-existence of two competing pronunciations, one with and one without the final consonant: "cette lettre [*p*] est indifferente… quelques personnes font cette lettre muëtte, mais il vaut mieux la prononcer" (De la Touche, 1696).[10] The change evidently involved two overlapping usages, one of which eventually replaced the other, and not a progressive phonetic erosion (e.g. [p] $>$ [ɸ] $>$ $\emptyset$).

A second argument against treating pre-pausal "truncation" as a sound change is that it did not apply systematically. Although many words lost their final consonant pre-pausally, many others retained, or even regained theirs. As one illustration of this, three of Palsgrave's examples in (13a) now have pronounced final consonants: *avec*, *soif*, *fil*. And finally, it seems unlikely that pre-pausal truncation (when it occurred) could be an extension of pre-C truncation, since this last process was no longer productive at this stage; we observe no new deletions after the 16th century, and in fact, final consonants were reappearing for many words in pre-C contexts. An important factor here is the adoption of large numbers of Latinate borrowings in the learned usage of this period. These words reintroduced many consonantal sequences that had disappeared from the inherited lexicon, and undermined the phonotactic pressures that once motivated (14a).

We can conclude that the changes affecting the pronunciation of final consonants in ModFr were therefore not primarily phonological. They were instead guided by functional pressures (the tendency to neutralize unnecessary distinctions) and, to a surprising extent, normative influence. The contextual alternation of word forms introduced in MidFr was a costly complication in the grammar. It may have eased pronunciation, but it offered absolutely no other functional advantages. On the contrary, stem consonants and inflectional suffixes were deleted in a

[10]Cited by Thurot (1883).

significant proportion of word tokens.[11] The language compensated by developing other strategies, such as making determiners and subject pronouns obligatory, giving rise in the end to a system where for the most part speakers could simply do without the information so unreliably encoded by final consonants.

It would have made sense, given these circumstances, for contextual alternations to be eliminated altogether. This is in fact what happened for many words, in particular for the entire class of singular nouns (outside of fixed expressions). We have already seen, for example, that pre-C truncation came relatively late for many words ending in *r/l/f* (e.g. *trésor*, *calcul*, *relief*), and that normative forces were often successful in reversing its effects (Fouché, 1961, p. 669–70). The final consonant was also restored in many monosyllables, e.g. *duc*, *cap*, *chef*. This tendency is partly explained by functional considerations: a final consonant represents a major portion of the phonological content of a monosyllable, and without it, many words become homophonous and therefore ambiguous.

The question is, then, why final consonants were not restored more systematically, since there would always be some functional advantage to be gained. Moreover, since the final consonant was still pronounced in pre-V and pre-pausal contexts, its "restoration" was a simple matter of generalizing the form used in these contexts to pre-C contexts. (This process was obviously an analogical change, not a sound change.) The fact is, however, that most nouns did not follow this path: instead, contextual alternation was eliminated by deleting the final consonant across the board.

This was a process that had already started in OFr: recall from §2 that nouns regularly lost their final consonant in the plural, through cluster simplification (8). The presence of the plural marker [s] presumably served as a clue to the listener that a consonant might be missing, and moreover this consonant was still systematically pronounced in the singular. But this alternation meant that the final consonant was no longer absolutely necessary for recognizing the word. When the MidFr sound changes applied, the stem-final consonant disappeared even from the singular form, in pre-C contexts, and the status of this consonant as part of the phonological identity of the word was further weakened. It was still pronounced in pre-pausal and pre-V contexts in the singular, but for many nouns, this proved to be insufficient motivation for maintaining the original form of the noun.

Polysyllabic nouns generally had a rich enough phonological content to do without one consonant: *appétit*, *estomac*. And though monosyllabic nouns showed more resistance, as explained above, in many cases they too lost their final consonant. For example, for nouns like *drap* "sheet" or *clef* "key", which appeared more frequently in the plural, or *coup*, which occurred frequently in phrases of the form *coup de*, usage favored forms with a silent final stem consonant (Fouché, 1961, p. 676–77). In such cases, the consonant disappeared completely, eliminating contextual alternation in favor of a single, truncated form. Again, this was a case of form

[11] To take one example, at least one out of three occurrences of the nominal case and number marker *s* became silent, according to Zink (1989, p. 36).

replacement, not a phonological process of truncation: the pattern for these nouns was for the pre-C form to generalize to pre-pausal contexts, and then later to pre-V liaison contexts.

The phonological identity of singular common nouns thus changed in one of two ways, depending on whether the historical final consonant was lost everywhere, including pre-V and pre-pausal contexts, as in the case of *clef* (now also written "*clé*") or restored everywhere, including pre-C contexts, as in *chef*. Recall from (14b) that final *s/f* were originally pronounced [z]/[v] in pre-V contexts. Nouns where final *s/f* were restored as non-alternating final consonants eventually stopped undergoing voicing, and presented a single phonological form in all contexts: e.g. *bœuf* "bull" is now pronounced invariably with an [f], but in early ModFr we find transcriptions like *le beuv et la vache* "the bull and the cow" (Raillet, 1664) and *du beu và la mode* "dish of braised beef" (De la Touche, 1696).

For other classes of words, contextual distinctions were not completely neutralized; these are the words that participate in consonant liaison in ModFr. The majority of these words are inflected forms (e.g. plural nouns and adjectives, conjugated verbs), which means that the final consonant corresponds to a grammatical ending (or part of it) and is not part of the stem. This explains why the pronunciation of these words developed more or less uniformly, without the haphazard lexical variation that we just observed for singular nouns. Furthermore, the only consonants involved here are [z] and [t].

The fact that inflectional suffixes encode morphosyntactic information may explain why these consonants were not lost altogether, and the fact that this information is often redundantly encoded in more than one place in the sentence may explain why they were not restored across the board. Instead, ModFr has simply retained a version of the MidFr system, with a prononced final consonant in certain contexts, and a silent consonant elsewhere. But compared to MidFr, the contexts where the final consonant is pronounced have been reduced severely: it now only occurs in some pre-V contexts, and not at all before a pause.

Before exploring the reasons for this development, let us mention the other classes of words that have maintained contextual forms in ModFr. These include closed-class items (pronouns, determiners, conjunctions), but also many content words (prepositions, adjectives, adverbs). In these cases, the final consonant can be part of the root or a derivational suffix, like *-eux* or *-ment*. It is not surprising to find, within these same word classes, examples of words where the final consonant was fully restored in all contexts, e.g. *il* "he", *leur* "their", *bref* "brief" (often involving the final consonants *r/l/f*, as we see here). There are extremely few cases of across-the-board generalization of the truncated form (possible examples include *hors* "outside", *bientôt* "soon"), because the proclitic nature of most of the members of these classes ensured the survival of liaison in pre-V contexts. In the case of adjectives, the liaison consonant was also preserved by analogy with the feminine forms: e.g. *petit/petite* "small", *premier/première* "first".

Just as for the inflected forms in [t] and [z], the final consonants of these other alternating words are no longer pronounced in ModFr before a pause (or in isolation). As explained above for singular nouns (which have basically followed the same evolution, taken one step further), the prevalence of truncated forms in MidFr diminished the role of final consonants, and the core phonological identity of these words was eventually "updated" to reflect this. Another way to view this shift is to say that in MidFr, the unmarked form of a word was its long form, and the truncated form had to be licensed by a special context (pre-C). Once the language adapted to rely less on the presence of the final consonant, the the truncated form was able to take over as the unmarked form, and gradually spread analogically to pre-pausal contexts. For many words, both forms existed as stylistic variants in this context until the end of the 17th century (recall the quotations at the beginning of this section). The short form eventually won out, and the realization of the final consonant became restricted to an ever smaller set of pre-V liaison contexts. In current French, competition between long and short forms can be observed in many pre-V contexts (the phenomenon of optional liaison).

It should be mentioned, finally, that a few words in ModFr seem to preserve the MidFr distribution of contextual forms, with a truncated form in pre-C contexts but not in pre-pausal contexts:

(16) a. *huit femmes* [ɥifam] '8 women', *huit hommes* [ɥitɔm] '8 men', *il y en a huit* [ɥit] 'there are 8'
b. *dix femmes* [difam] '10 women', *dix hommes* [dizɔm] '10 men', *il y en a dix* [dis] 'there are 10'

These words (which will not be included in the formal analysis of the following section) can be considered to be remnants of the MidFr system. Because of their frequency, and the types of constrcuctions in which they appear, they have managed to avoid the more dominant paths of development described above. These words exhibit a good deal of instability, in part as a result of pressure from the more prevalent pattern, but we cannot conclude that they consitute a completely non-productive class.

5 HPSG formalization

The foregoing discussion described two transitions in the evolution of French final consonants: the sound change introducing contextual forms in MidFr, and different paths of simplification and reanalysis of the contextual alternation in ModFr.

5.1 Phonological context

The changes in question involve the phonological content of word forms, but they are conditioned by the properties of the surrounding context. One way of handling this kind of phono-syntactic interaction is to enrich lexical representations with information about the phonology of adjacent elements. I adopt a variant of the PHON-CONTEXT model of Asudeh and Klein (2002), which defines the following constraint on phrasal constructions:

(17) *construction* →

$$\left[\text{DTRS}\ \left\langle \begin{array}{l}\left[\text{PHON} \mid \text{P-CTXT}\ \ \boxed{1}\right], \boxed{1}\left[\text{PHON} \mid \text{P-CTXT}\ \ \boxed{2}\right], \ldots, \\ \qquad\qquad \boxed{n}\left[\text{PHON} \mid \text{P-CTXT}\ \ \textit{p-ctxt}\right]\end{array} \right\rangle\right]$$

Each daughter in the construction is given full access to the *sign* of the immediately following daughter. It is clear that this formulation is too unconstrained; exactly how much contextual information should be made visible in this way is an open empirical question. In the following analysis, alternating words only need to refer to the first segment of the phonology of the immediately following word (and to one more abstract feature, to be introduced below). Also note that, unlike Asudeh and Klein, I do not a *nil* context for the last daughter in (17). This value needs to be left underspecified, in case the construction is embedded with a larger construction,[12] or instantiated as *nil* by a root utterance constraint.

5.2 Introduction of contextual forms

We being by sketching an analysis of the OFr system, the starting point for the transition summarized in §3.2. At this stage, consonant-final words showed no contextual alternation at syntactic word boundaries. In other words, an OFr word can be assigned a lexical entry with a simple PHON value, encoding the unitary pronunciation of the word in all contexts, and making no use of the P-CTXT apparatus just introduced. Phonological processes active at this time (final devoicing, final cluster simplification) did give rise to alternations between forms of the same lexeme (masculine vs. feminine, singular vs. plural), which later became grammaticalized as instances of paradigmatic stem allopmorphy, for example *vif/vive* "alive", *coup/cous* "strike(s)", cf. the examples in (8). The significance of this development was discussed in §4.2, but its formal analysis is not directly relevant for our purposes, since it involves relations between the lexical entries of distinct inflected word forms.

The sound changes in (14) introduce contextual alternations in the pronunciation of a single word. This development can be modeled by assigning consonant-final words lexical entries with complex PHON specifications, with disjunctive clauses corresponding to the phonological contexts giving rise to form alternation.

(18)
$$\left[\text{PHON}\ \begin{array}{l}\begin{bmatrix}\text{SEGS} & \boxed{1}\\ \text{P-CTXT} & \textit{nil}\end{bmatrix} \vee \begin{bmatrix}\text{SEGS} & \mathsf{s\text{-}z}(\boxed{1})\\ \text{P-CTXT} & \left[\text{SEGS}\ \left\langle \textit{vow}, \ldots \right\rangle\right]\end{bmatrix}\\ \qquad \vee \begin{bmatrix}\text{SEGS} & \mathsf{trunc}(\boxed{1})\\ \text{P-CTXT} & \left[\text{SEGS}\ \left\langle \textit{cons}, \ldots \right\rangle\right]\end{bmatrix}\end{array}\right]$$

[12] In fact the rightmost daughter should structure-share its P-CTXT with the mother construction, so that contextual information can be passed down through levels of syntactic embedding to the relevant lexical element.

In this analysis, the pre-pausal form (encoded by the first disjunct, specifying a null P-CTXT) is taken as the basic form, corresponding to the historically original form, inherited from OFr. The pre-V form is identical except that final [s] and [f] undergo voicing; this is indicated by the phonological function s-z applied to the basic form $\boxed{1}$. The pre-C form is derived by truncation of the final consonant of the basic form.

The adverb *toujours*, for instance, has the PHON value shown in (19a), with three distinct pronunciations, while *beaucoup* has just two (19b), because [p] is not affected by s-z. (I assume modern phonetic values for vowels and consonants elsewhere in the word, for expository purposes.)

(19) a.
$$\begin{bmatrix}\text{SEGS} & \boxed{1}\left\langle \text{t,u,ʒ,u,ʁ,s}\right\rangle \\ \text{P-CTXT} & \textit{nil}\end{bmatrix} \vee \begin{bmatrix}\text{SEGS} & \mathsf{s\text{-}z}(\boxed{1}) = \left\langle \text{t,u,ʒ,u,ʁ,z}\right\rangle \\ \text{P-CTXT} & \begin{bmatrix}\text{SEGS} & \left\langle \textit{vow},\ldots\right\rangle\end{bmatrix}\end{bmatrix}$$
$$\vee \begin{bmatrix}\text{SEGS} & \mathsf{trunc}(\boxed{1}) = \left\langle \text{t,u,ʒ,u,ʁ}\right\rangle \\ \text{P-CTXT} & \begin{bmatrix}\text{SEGS} & \left\langle \textit{cons},\ldots\right\rangle\end{bmatrix}\end{bmatrix}$$

b.
$$\begin{bmatrix}\text{SEGS} & \boxed{1}\left\langle \text{b,o,k,u,p}\right\rangle \\ \text{P-CTXT} & \textit{nil}\end{bmatrix} \vee \begin{bmatrix}\text{SEGS} & \mathsf{s\text{-}z}(\boxed{1}) = \boxed{1} \\ \text{P-CTXT} & \begin{bmatrix}\text{SEGS} & \left\langle \textit{vow},\ldots\right\rangle\end{bmatrix}\end{bmatrix}$$
$$\vee \begin{bmatrix}\text{SEGS} & \mathsf{trunc}(\boxed{1}) = \left\langle \text{b,o,k,u}\right\rangle \\ \text{P-CTXT} & \begin{bmatrix}\text{SEGS} & \left\langle \textit{cons},\ldots\right\rangle\end{bmatrix}\end{bmatrix}$$

As discussed in §4.2, the role of the final consonant was weakened by the frequent occurrence of the truncated form. This triggered various developments in the next stage of the language.

5.3 Transitions to ModFr pronunciations

For the majority of words, the major change in ModFr was the introduction of variation in pre-pausal contexts. The original long form and the truncated form co-existed for a time (20a), before the eventual triumph of the truncated form (20b).

(20) a.
$$\begin{bmatrix}\text{PHON} & \begin{matrix}\begin{bmatrix}\text{SEGS} & \boxed{1} \vee \boxed{2} \\ \text{P-CTXT} & \textit{nil}\end{bmatrix} \vee \begin{bmatrix}\text{SEGS} & \mathsf{s\text{-}z}(\boxed{1}) \\ \text{P-CTXT} & \begin{bmatrix}\text{SEGS} & \left\langle \textit{vow},\ldots\right\rangle\end{bmatrix}\end{bmatrix} \\ \vee \begin{bmatrix}\text{SEGS} & \boxed{2}\,\mathsf{trunc}(\boxed{1}) \\ \text{P-CTXT} & \begin{bmatrix}\text{SEGS} & \left\langle \textit{cons},\ldots\right\rangle\end{bmatrix}\end{bmatrix}\end{matrix}\end{bmatrix}$$

b. $$\rightsquigarrow \begin{bmatrix} \text{PHON} & \begin{array}{l} \begin{bmatrix} \text{SEGS} & \boxed{2} \\ \text{P-CTXT} & \textit{nil} \vee \begin{bmatrix} \text{SEGS} & \left\langle \textit{cons}, \ldots \right\rangle \end{bmatrix} \end{bmatrix} \\ \vee \begin{bmatrix} \text{SEGS} & \mathsf{liaison}(\boxed{2}) \\ \text{P-CTXT} & \begin{bmatrix} \text{SEGS} & \left\langle \textit{vow}, \ldots \right\rangle \end{bmatrix} \end{bmatrix} \end{array} \end{bmatrix}$$

A number of important shifts are involved in the transition to (20b). The form $\boxed{2}$ is now the more frequent form, and the citation form. The historical long form loses its status of basic form. In fact, for words like *toujours*, the original form with final [s] no longer appears in any contexts; we are left with only the two "derived" pronunciations [tuʒuʁ] and [tuʒuʁz]. The relationship between these forms is consequently reinterpreted as shown in (20b): $\boxed{2}$ is now the basic form, and the pre-V form is derived from it by a new process, labeled liaison.

The function liaison cannot represent a simple phonological process. The relation between liaison forms and non-liaison forms is grammaticalized in the form of a two-slot paradigm, which is used in the analysis of all manifestations of liaison in ModFr, including those that have historical origins other than the final consonant deletion described throughout this paper. The slots of the paradigm can be filled in in several different ways. In all of the examples considered up to now, the liaison form is derived from the non-liaison form by the addition of an extra final consonant. This "latent" consonant can correspond to an unpredictable (historical) root consonant (21a), or it can be systematically associated with the grammatical features of the word (b). In such cases the identity of the latent consonant must be encoded somewhere in the lexical representation of the word, but not as part of its core phonological content.[13] The liaison form can be suppletive (21c,d), or it can be defective (e).

(21)

	non-liaison	liaison form	
a.	boku	bokup	*beaucoup* 'a lot'
b.	pəti	pətiz	*petits* 'small.*pl*'
c.	sə	sɛt	*ce* / *cet* 'this'
d.	nuvo	nuvɛl	*nouveau* / *nouvel* 'new' (prenominal)
e.	fʁɑ̃	*	*franc* 'frank' (prenominal)
f.	ku	ku	*coup* 'blow' / *cou* 'neck' / *coût* 'cost'

And finally, words that show no liaison alternation in ModFr, such as singular nouns, simply have identical forms in both slots of their paradigm (f).

The lexical schema in (20b) thus underwent a further step of reinterpretation: the morphologization of the relationship between the two forms.

[13] Bonami et al. (2004) introduce the idea of a phonological "appendix" for encoding latent consonants for liaison and morphological derivation.

(22) $$\begin{bmatrix} \text{PHON} & \begin{bmatrix}\text{SEGS} & \boxed{a}\\ \text{P-CTXT} & nil \vee \begin{bmatrix}\text{SEGS} & \langle cons, \ldots\rangle\end{bmatrix}\end{bmatrix} \vee \begin{bmatrix}\text{SEGS} & \boxed{b}\\ \text{P-CTXT} & \begin{bmatrix}\text{SEGS} & \langle vow, \ldots\rangle\end{bmatrix}\end{bmatrix} \\ \text{MORPH} & \left[\begin{array}{ll} \multicolumn{2}{l}{\textit{liaison-paradigm}}\\ \text{NON-LIAIS-FORM} & \boxed{a}\\ \text{LIAIS-FORM} & \boxed{b}\end{array}\right]\end{bmatrix}$$

This informal representation is meant to show that neither form is derived from the other in the phonology. Instead, the forms are organized in a paradigm in the morphological component of the lexical entry, where the various possible relationships, or the lack of relationship, between the two forms can be modeled.

ModFr has also seen an evolution in the nature of the contextual conditions. While these were closely correlated with the phonological content of the following word in earlier stages, there are situations where this no longer the case in ModFr. We assume that consonant-initial words in MidFr became associated with an abstract feature [−LIAISON-TRIGGER], encoding the fact that they could not license the appearance of a liaison form. The switch to a non-phonological feature is crucial for the class of "aspirated *h*" words, which lost their initial consonant in early ModFr period (e.g. *hache* 'axe': MidFr [haʃœ] ⇝ ModFr [aʃ]). They still fail to trigger liaison today, despite being vowel-initial phonologically.

(23) $$\begin{bmatrix}\text{PHON} & \begin{bmatrix}\text{SEGS} & \langle \text{h}, \ldots\rangle\end{bmatrix}\\ \text{LTRIG} & -\end{bmatrix} \rightsquigarrow \begin{bmatrix}\text{PHON} & \begin{bmatrix}\text{SEGS} & \langle vow, \ldots\rangle\end{bmatrix}\\ \text{LTRIG} & -\end{bmatrix}$$

The constraints on liaison in ModFr refer to the value of the lexically-specified feature [±LTRIG], instead of directly inspecting the SEGMENTS list of the licensing word. We can represent this move by modifying the P-CTXT connstraints inas follows:

(24) $$\begin{bmatrix} \text{PHON} & \begin{bmatrix}\text{SEGS} & \boxed{a}\\ \text{P-CTXT} & \begin{bmatrix}\text{LTRIG} & -\end{bmatrix}\end{bmatrix} \vee \begin{bmatrix}\text{SEGS} & \boxed{b}\\ \text{P-CTXT} & \begin{bmatrix}\text{LTRIG} & +\end{bmatrix}\end{bmatrix} \\ \text{MORPH} & \left[\begin{array}{ll} \multicolumn{2}{l}{\textit{liaison-paradigm}}\\ \text{NON-LIAIS-FORM} & \boxed{a}\\ \text{LIAIS-FORM} & \boxed{b}\end{array}\right]\end{bmatrix}$$

There are other clear signs of the grammaticalization of liaison in ModFr and its shift away from a purely phonological phenomenon. The strict association between liaison forms and liaison contexts expressed in all of the preceding lexical schemas must be relaxed, because in many syntactic environments in ModFr, liaison is optional. The only general constraint is that a liaison form must be immediately followed by a [+LTRIG] word:

(25) a. *beaucoup aimer* [bokueme] / [bokupeme] ‘like a lot’

b. *beaucoup manger* [bokumɑ̃ʒe] / *[bokupmɑ̃ʒe] ‘eat a lot’

(26)
$$\begin{bmatrix} \textsc{phon} & \begin{bmatrix}\textsc{segs} & \boxed{a}\end{bmatrix} \vee \begin{bmatrix}\textsc{segs} & \boxed{b} \\ \textsc{p-ctxt} & \begin{bmatrix}\textsc{ltrig} & +\end{bmatrix}\end{bmatrix} \\ \textsc{morph} & \begin{bmatrix} \textit{liaison-paradigm} & \\ \textsc{non-liais-form} & \boxed{a} \\ \textsc{liais-form} & \boxed{b}\end{bmatrix}\end{bmatrix}$$

It follows that liaison forms cannot appear in isolation or before a pause. Non-liaison forms are subject to no contextual constraints in this generic lexical entry schema, but particular syntactic combinations (head-specifier phrases, head-subject phrases) can impose additional conditions.

The lexical schema in (26) is the last stage of the analysis that will be presented here, but it should be mentioned that the grammaticalization of liaison in ModFr calls into question the reliance on P-CTXT constraints. The P-CTXT approach is appropriate for sandhi phenomena that are primarily phonologically conditioned, because it gives a word direct access to the PHON values of its neighbors. While it is technically possible to refer to non-phonological information via P-CTXT, given the powerful formulation of the constraint in (17), such proposals must be carefully motivated.[14] As we can see in (26), only one contextual constraints is still in force at the lexical level in ModFr, it does not refer directly to phonological information, but to the abstract feature LTRIG.

See Bonami et al. (2004) and Bonami et al. (2005) for a treatment of ModFr liaison in terms of constraints on syntactic combinations, where the grammaticalized remnants of phonological context constraints are modeled using the interaction of two interface features (LTRIG, also introduced here, and LFORM, encoding the liaison status of the alternating word). Those proposals can be considered to be a further reanalysis step, following on from the succession of analyses presented here. The current paper serves to situate synchronic HPSG analyses of French liaison in their historical context.

References

Asudeh, Ash and Klein, Ewan. 2002. Shape Conditions and Phonological Context. In Frank van Eynde, Lars Hellan and Dorothee Beermann (eds.), *Proceedings of the 8th International HPSG Conference*, pages 20–30, Stanford, CA: CSLI Publications.

[14] See Tseng (2005) for the issue of locality in phonological selection.

Bender, Emily and Flickinger, Dan. 1999. Diachronic Evidence for Extended Argument Structure. In Gosse Bouma, Erhard Hinrichs, Geert-Jan M. Kruijff and Richard T. Oehrle (eds.), *A Collection of Papers on Formal Grammar, Head-Driven Phrase Structure Grammar, and Categorial Grammar*, pages 3–19, Stanford, CA: CSLI Publications.

Bonami, Olivier and Boyé, Gilles. 2003. La nature morphologique des allomorphies conditionnées. Les formes de liaison des adjectifs en français. In Bernard Fradin, Georgette Dal, Françoise Kerleroux, Nabil Hathout, Marc Plénat and Michel Roché (eds.), *Les Unités morphologiques / Morphological Units*, Villeneuve d'Ascq: Silex.

Bonami, Olivier, Boyé, Gilles and Tseng, Jesse. 2004. An integrated approach to French liaison. In Gerhard Jäger, Paola Monachesi, Gerald Penn and Shuly Wintner (eds.), *Proceedings of Formal Grammar 2004*, pages 29–45, Nancy: Loria.

Bonami, Olivier, Boyé, Gilles and Tseng, Jesse. 2005. Sur la grammaire des consonnes latentes. *Langages* 158, 89–100.

Bourciez, Édouard and Bourciez, Jean. 1967. *Phonétique française. Étude historique*. Paris: Klincksieck.

Brunot, Ferdinand. 1966. *Histoire de la langue française des origines à nos jours. Tome I. De l'époque latine à la Renaissance*. Paris: Armand Colin.

Fouché, Pierre. 1961. *Phonétique historique du français. Volume III: Les consonnes et index général*. Paris: Klincksieck.

Harris, Alice C. and Campbell, Lyle. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.

Livet, Charles-Louis. 1859. *La Grammaire française et les Grammairiens du XVIe siècle*. Paris: Didier.

Morin, Yves-Charles. 1986. On the morphologization of word-final consonant deletion in French. In Henning Andersen (ed.), *Sandhi Phenomena in the Languages of Europe*, pages 167–210, Berlin: Mouton de Gruyter.

Paris, Gaston and Pannier, Léopold. 1872. *La vie de saint Alexis: poème du XIe siècle et renouvellements des XIIe, XIIIe, et XIVe siècles*. Bibliothèque de l'École des hautes études, No. 7, Paris: A. Franck.

Schane, Sanford. 1968. *French phonology and morphology*. Cambridge, MA: MIT Press.

Stürzinger, Johann Jakob (ed.). 1884. *Orthographia Gallica : Ältester Traktat, über französische Aussprache und Orthographie*. Altfranzösische Bibliothek, No. 8, Heilbronn: Gebr. Henninger.

Thurot, Charles. 1883. *De la prononciation française depuis le commencement du XVIe siècle d'après les témoignages des grammairiens. Tome second*. Paris: Imprimerie nationale.

Tseng, Jesse. 2005. Prepositions and complement selection. In Aline Villavicencio and Valia Kordoni (eds.), *Second ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, pages 11–19, University of Essex.

Tseng, Jesse. 2008. L'inversion pronominale: histoire et analyse. In *Actes du 1er Congrès mondial de linguistique française*, Paris.

Zink, Gaston. 1986. *Phonétique historique du français*. Paris: Presses Universitaires de France, 6. edition.

Zink, Gaston. 1989. *Morphologie du français médiéval*. Paris: Presses Universitaires de France, fourth edition.

On the copula:
from a Fregean to a Montagovian treatment

Frank Van Eynde

University of Leuven

Proceedings of the HPSG09 Conference

Georg-August-Universität Göttingen, Germany

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

The analysis of the copula as a semantically vacuous word in mainstream HPSG is appropriate for some of its uses, such as the progressive and the passive, but not for its use in clauses with a predicative complement. In such clauses the copula denotes a relation of coreference between the indices of the subject and the predicative complement.

1 The Fregean treatment

The copula belongs to a class of verbs which take a subject-oriented predicative complement. Some typical members of this class are *become*, *remain* and *seem*, as used in (1a). Semantically, these verbs are treated as functions which take a single clausal argument, as in (1b). Ignoring tense, this formula also represents (1c).

(1) a. John seems sad.
  b. *seem*(*sad*(*John*))
  c. It seems that John is sad.

In terms of the typed feature structure (TFS) notation of HPSG the combination of the verb with its predicative complement can be expressed as in the following AVM of the German *erscheint klug* 'seems clever', quoted from Müller (2002, 105).

(2)
$$\left[\begin{array}{ll} \text{CAT}\,|\,\text{SUBCAT} & \langle \text{NP}_{\boxed{1}}\,,\ (\text{NP}[\textit{ldat}]_{\boxed{2}})\ \rangle \\ \text{CONTENT} & \left[\begin{array}{ll} \multicolumn{2}{l}{\textit{erscheinen}} \\ \text{EXPERIENCER} & \boxed{2}\ \textit{index} \\ \text{SOA} & \left[\begin{array}{ll} \multicolumn{2}{l}{\textit{klug}} \\ \text{THEME} & \boxed{1}\ \textit{index} \end{array}\right] \end{array}\right] \end{array}\right]$$

In words, the verb *erscheint* 'seems' assigns the S(TATE-)O(F-)A(FFAIRS) role to its predicative complement *klug* 'clever' and the latter assigns the THEME role to the subject of the verb. Besides, *erscheint* assigns the EXPERIENCER role to its optional dative NP complement. Its equivalent in English is the optional PP[*to*], as used in (3).

(3) John seems sad to me.

That the PP[*to*] is an argument of the verb and not of the adjective is clear from the fact that its paraphrase is (4a), rather than (4b).

(4) a. It seems to me that John is sad.

† For their comments on previous versions I thank Ivan Sag, Gert Webelhuth, Stefan Müller, Doug Arnold and the anonymous reviewers of the HPSG-2009 programme committee.

b. It seems that John is sad to me.

Turning now to the copula, if it were analyzed along the same lines as *seem*, (5a) would be represented as in (5b), but what one finds instead is (5c).

(5) a. John is sad.
b. *be*(*sad*(*John*))
c. *sad*(*John*)

The treatment of the copula as semantically vacuous can be traced back to Gottlob Frege, who explicitly claimed that: "it can be replaced by a verbal affix; for example, instead of saying 'this leaf is green' one can say 'this leaf greens'." Frege (1892). Some linguistic evidence for this claim is provided by the observation that the omission of the copula does not affect the meaning of the clause, as illustrated in (6).

(6) a. John seems (to be) sad.
b. With John (being) ill we cannot go on holiday.

In some languages this also holds for the finite forms, more specifically the present tense, as in the Russian (7).

(7) Ona xorosij vrac.
she good doctor
'She is a good doctor.'

Similar observations have been made about the finite forms of the copula in African American Vernacular English, Japanese, Hungarian, Arabic and Mauritian Creole, see a.o. Bender (2001), Dalrymple et al. (2004) and Henri and Abeillé (2007).

The assumption of semantic vacuity is also adopted in HPSG. In Pollard and Sag (1994, 147), for instance, the CONTENT value of the copula is identified with that of its predicative complement.

(8) $$\begin{bmatrix} \text{CAT} \mid \text{SUBCAT} & \left\langle \text{NP}, \text{XP}[+\text{PRD}] : \boxed{1} \right\rangle \\ \text{CONTENT} & \boxed{1} \end{bmatrix}$$

In words, the copula selects an NP and a predicative XP whose CONTENT value is identical to the one of the copula itself.

2 Problems with the Fregean treatment

Characteristic of the Fregean treatment of the copula is the discrepancy between syntactic and semantic structure: What is syntactically the head of the clause is absent from the semantic representation. Technically, this kind of mismatch is easy to model in a TFS-based grammar, and there is evidence that this treatment is indeed appropriate for the passive and progressive uses of the copula, as will be shown in section 4. However, for its use in predicational structures, as in (5), this treatment is less felicitous for a number of reasons. I will discuss four.

2.1 The semantic type of the nominal predicates

As suggested by Frege's paraphrase of 'this leaf is green' as 'this leaf greens', he assumes that the predicative complement, i.c. *green*, takes on a verbal role, reducing the copula's role to that of a verbal affix. The equivalent of this assumption in HPSG is the stipulation that the predicative complement denotes a state of affairs. More specifically, while the predicative complement can belong to any syntactic category (N, A, V, P), its CONTENT value is invariably of type *soa* (*state-of-affairs*). Objects of that type are canonically assigned to verbs and VPs, and consist of a list of quantifiers, ordered in terms of scope, and a nucleus, as exemplified by the representation of *visit* in (9).

(9) $$\begin{bmatrix} \textit{soa} & \\ \textsc{quants} & \textit{list}\Big(\textit{quant-rel}\Big) \\ \textsc{nucleus} & \begin{bmatrix} \textit{visit-rel} & \\ \textsc{visitor} & i \\ \textsc{visited} & j \end{bmatrix} \end{bmatrix}$$

The assignment of the *soa* type to the predicative complements not only reflects the Fregean treatment, it also follows from the analysis of the copula in (8): Since the combination of the copula with its predicative complement is a VP and since the CONTENT value of a VP is of type *state-of-affairs*, it follows, given the structure sharing in (8), that the predicative complement must denote a state of affairs as well.

This, however, is a problem for the nominal predicates, since nominals have a CONTENT value of type *scope-object*. Objects of that type consist of an index and a set of restrictions on its reference, as exemplified by the representation of *table* in (10).

(10) $$\begin{bmatrix} \textit{scope-obj} & \\ \textsc{index} & \boxed{1}\ \textit{index} \\ \textsc{restr} & \Big\{\textit{table}(\boxed{1})\Big\} \end{bmatrix}$$

As a consequence, the nominals must undergo a **type shift** when they are used in predicative position. The notion of type shift was introduced in Partee (1987). In Pollard and Sag (1994, 360) it is modeled in terms of the following lexical rule.[1]

(11) PREDICATIVE NP LEXICAL RULE:

$$\begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{PRD} & - \end{bmatrix} \\ \text{SUBJ} & \langle\ \rangle \end{bmatrix} \\ \text{CONTENT} & \begin{bmatrix} \textit{scope-obj} & \\ \text{INDEX} & \boxed{1} \\ \text{RESTR} & \boxed{2}\ \textit{set}(\textit{psoa}) \end{bmatrix} \end{bmatrix} \Rightarrow \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{PRD} & + \end{bmatrix} \\ \text{SUBJ} & \langle \text{XP}_{\boxed{1}} \rangle \end{bmatrix} \\ \text{CONTENT} & \boxed{2} \end{bmatrix}$$

In words, for every nonpredicative noun which denotes a scope-object, there is a homonymous predicative noun which denotes the set of restrictions which are part of the scope-object ($\boxed{2}$). In the type hierarchy of Pollard and Sag (1994), which treats the RESTRICTION value as a set of parametrized states of affairs, this rule yields a semantic object which can be identified with the CONTENT value of the copula.[2] A consequence of this treatment is that the nouns are systematically ambiguous.

A lexical rule is not the only possible way to model the type shift in HPSG terms. Another possibility is proposed in Müller (2009). Quoting Kasper (1995), Müller points out that the lexical rule of Pollard and Sag (1994) is inappropriate for the analysis of nominal predicates which contain an adjunct. Given the canonical HPSG treatment of adjuncts, the prenominal adjective in (12), for instance, selects an N-bar head and identifies its own index with that of the noun.

(12) John is a good candidate.

However, if the noun is in predicative position, it has no index! To repair this Müller (2009) applies the type shift at the level of the full NP, rather than at the lexical level. To model this he employs a unary syntactic rule which transforms nonpredicative NPs into predicative ones.

[1] Pollard and Sag (1994) uses the term *nominal-object* for what is called a *scope-object* in Ginzburg and Sag (2000). I use the latter term.

[2] In the type hierarchy of Ginzburg and Sag (2000), which treats the RESTRICTION value as a set of facts, the type shift has to be modeled in another way, but since the equivalent of (11) in Ginzburg and Sag (2000) does not mention the CONTENT values, it is not made clear how this is done.

(13) PREDICATIVE NP PROJECTION SCHEMA:

$$
\textit{np-pred-phrase} \Rightarrow
\begin{bmatrix}
\text{SYNSEM}|\text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{PRD} & + \\ \text{SUBJ} & \langle \text{NP}_{\boxed{1}} \rangle \end{bmatrix} \\ \text{SPR} & \langle\,\rangle \\ \text{COMPS} & \langle\,\rangle \end{bmatrix} \\ \text{CONTENT}|\text{INDEX} & \boxed{0} \end{bmatrix} \\
\text{C-CONT} & \begin{bmatrix} \text{RELS} & \left\langle \begin{bmatrix} \textit{equal-rel} & \\ \text{ARG0} & \boxed{0}\ \textit{event} \\ \text{ARG1} & \boxed{1} \\ \text{ARG2} & \boxed{2} \end{bmatrix} \right\rangle \\ \text{H-CONS} & \langle\,\rangle \end{bmatrix} \\
\text{NON-HEAD-DTRS} & \left\langle \begin{bmatrix} \text{SYNSEM}|\text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{noun} \\ \text{SPR} & \langle\,\rangle \\ \text{COMPS} & \langle\,\rangle \end{bmatrix} \\ \text{CONTENT} & \begin{bmatrix} \textit{npro} & \\ \text{INDEX} & \boxed{2}\ \textit{index} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle
\end{bmatrix}
$$

In words, the rule turns a fully saturated nonpronominal NP which denotes a scope-object (= the non-head-daughter) into a predicative NP which selects a subject and which denotes an object of type *event* (= the mother). The C-CONT attribute captures the constructional aspects of the semantic composition. In this case, it represents a requirement of equality between the indices of the subject ($\boxed{1}$) and the NP daughter ($\boxed{2}$). Since the INDEX value of the NP mother ($\boxed{0}$) is inherited by the copula, the latter has an index of type *event*.

This treatment avoids the problem with (11), since the type shift is now applied after the addition of the adjuncts. At the same time, since (13) explicitly requires a fully saturated NP daughter, it does not subsume the determinerless predicates in (14) and the German (15).

(14) Sylvia is mayor of Seattle.

(15) Er ist Lehrer.
he is teacher
'He is a teacher.'

To cover these, Müller (2009) keeps a version of lexical rule (11). It will presumably be more constrained than (11), since only some of the (singular count) nouns can be used in this way (typically nouns denoting roles, functions and professions), but since the paper does not spell out the rule, this is left unclear.

Together, rule (13) and the implicit lexical rule solve the mismatch, but the price to pay is a systematic ambiguity for the NPs and for those nouns which can be used without determiner in predicative position.

2.2 Quantified predicate nominals

As admitted in Pollard and Sag (1994, 360), the canonical HPSG treatment does not account for the semantic contribution of the determiner in predicate nominals.[3] This is hardly surprising, since the Fregean analysis on which it is based has the same problem. To show this, let us compare the treatment of the indefinite article in (16) with that in (17).

(16) a. John knows a teacher.
b. ∃*x* [*teacher*(*x*) & *know*(*John*, *x*)]

(17) a. John is a teacher.
b. *teacher*(*John*)

In the analysis of (16a) the contribution of the indefinite article is captured in terms of the existential quantifier, but in the analysis of (17a) the article is assumed to be semantically vacuous, just like the copula.[4] This not only introduces another discrepancy between syntactic and semantic structure, it also raises the question of how predicative NPs with another determiner, such as *the* or *my*, have to be differentiated from those with the indefinite article.

2.3 Stipulation of an ambiguity for the copula

Another problem for the treatment of the predicate nominals concerns the pronouns and the proper nouns. They can be used in postcopular position, as exemplified in (18), but semantically it makes no sense to treat them as states-of-affairs or events. In fact, Stefan Müller's unary rule (13) explicitly requires the NP daughter to be nonpronominal.

(18) a. Cicero is Tully.
b. The winner is Jimmy Logan.
c. That must be her.
d. That book is mine.

[3]This criticism does not apply to the analysis in Müller (2009).
[4]For more discussion of this point, see Allegranza (2006, 78).

To handle these it is commonly assumed that the copula is not used in its predicational sense here, but in an equational or identificational sense. Also this assumption is due to Frege (1892) and has been very influential, both in logic and linguistics, see a.o. Pollard and Sag (1987, 66), Declerck (1988), Mikkelsen (2005) and Müller (2009).

In spite of its wide-spread acceptance, though, few have bothered to spell out what it is that distinguishes the predicational use from the identifying use. Matters would be easy, of course, if the latter would simply coincide with the combinations with proper nouns and pronouns, but this is not the case. On the one hand, there are other kinds of NPs that are canonically treated as complements of the identifying copula, such as the definite ones in (19).

(19) a. Clara is her youngest sister.
b. Tim is the man with the black tie in the left corner.

On the other hand, there are combinations with proper nouns or pronouns in which the copula has its usual predicating sense, as in (20).

(20) a. This was characteristic of Helen. A fine person in many ways, but this ability to forget completely the true state of our finances and start dreaming up major new spending opportunities, this was very Helen. [quoted from Kazuo Ishiguro, *Nocturnes. Five stories of music and nightfall.* Faber & Faber, 2009. page 130]
b. This movie is SO Woody Allen.
c. Susan is somebody we can trust.
d. Cicero is not just anybody; he is the greatest orator of all time.

This makes it very hard to formulate any criteria for drawing the distinction between the identifying and the predicating *be*. Moreover, the distinction sometimes gets in the way. Speaking of the treatment of pied piping in NPs, Ginzburg and Sag (2000, 195) remarks that "this analysis provides an account of examples like *I wondered [whose cousin] she was pretending to be* __, if we assume that complements of the identity copula are also predicative NPs." In other words, the treatment of pied piping is more uniform and straightforward if we do NOT distinguish between the predicational and the identifying senses.

2.4 Assignment of the EXPERIENCER role

As already pointed out in section 1, some of the predicate selecting verbs, such as *seem*, take an optional experiencer. Such verbs can obviously not be treated as semantically vacuous, since otherwise there is no way to assign the EXPERIENCER role to the relevant NP or PP. As a consequence, if the copula can take an optional experiencer, it follows that it cannot be semantically vacuous. The following evidence from Dutch suggests that this is indeed the case.

(21) a. Dat lijkt/is me echt te duur.
that seems/is me really too expensive.
'That seems/is really too expensive to me.'

b. Het lijkt/is ons nu wel duidelijk dat ze niet zullen komen.
it seems/is us now – clear that they not will come
'It seems/is clear to us now that they won't come.'

c. Het juiste aantal bleek/was hen nog niet bekend.
the exact number appeared/was them still not known
'The exact number appeared/was not yet known to them.'

Given that the pronominal objects *me* 'me', *ons* 'us' and *hen* 'them' are canonically treated as complements of resp. *lijken* 'seem' and *blijken* 'appear' and that they receive the EXPERIENCER role from these verbs, it would only be logical to treat them in the same way in the combination with the copula. Conversely, if one decides instead to treat the pronominal objects as adjuncts or as raised arguments in the case of the copula, then it would only be logical to treat them in the same way when they are combined with *seem* or *appear*, contrary to the canonical practice.

Notice, furthermore, that the combination is not only possible with adjectival predicates, but also with nominal ones, as in (22), and with prepositional ones, as in (23).

(22) a. Wat dit betekent is me nog steeds een raadsel.
what this means is to-me still always a puzzle
'What this means is still a puzzle to me.'

b. Het is ons een waar genoegen.
it is us a real pleasure
'It is a real pleasure to us.'

(23) a. Dat kereltje is ons tot last.
that guy-DIM is us to burden
'That little guy is a real burden for us.'

b. Dat is me om het even.
that is me about the same
'It is all the same to me.'

2.5 Summing up

The Fregean treatment of the copula complicates the treatment of the predicate nominals, requiring a type shift which makes the nominals systematically ambiguous, it does not account for the semantic contribution of the determiner in predicate nominals, it presupposes a distinction between predicating and identifying uses, which is very hard to substantiate, and it does not account for the assignment of the EXPERIENCER role.

As a final remark, notice that the main linguistic argument in favor of the treatment is not very strong. The omissibility of the copula in certain contexts and in certain languages, as in (6-7), is not by itself an argument for semantic vacuity. Otherwise, the existence of languages without articles, such as Latin and Russian, would entail that the articles do not contribute any content either, also in languages which have them. This, it goes without saying, is a conclusion which few semanticists would be happy to welcome.

3 A Montagovian treatment

The mismatch between syntactic and semantic structure which is characteristic of the Fregean treatment of the copula did not particularly appeal to Richard Montague. His insistence on compositionality made him more sympathetic to a treatment in which the copula is treated along the same lines as the other verbs. His analysis is briefly presented in 3.1, translated in HPSG terms in 3.2 and demonstrated to be superior to the Fregean treatment in 3.3. It will also be shown to be extensible to other verbs that select a predicative complement in 3.4.

3.1 The Quine-Montague proposal

The Montagovian treatment can be traced back to a proposal in Quine (1960, 114–118): "the sign '=' of identity is a relative term; thus a transitive verb, we might say ... Like any such term it joins singular terms to make a sentence. The sentence thus formed is true if and only if those component terms refer to the same object." (p. 115)

In terms of the PTQ model (Montague, 1974, 247–270) with its distinct representations for disambiguated English (DL) and intensional logic (IL), Richard Montague treated the copula as a transitive verb in disambiguated English and as the relation of identity in intensional logic. Defining the link between them is the following translation rule.[5]

(24) **be** translates into $\lambda\mathcal{P}\ \lambda x\ \mathcal{P}\ \{\hat{y}\ [\check{}x = \check{}y]\}$.

In other words, it is not only the identifying or equational *be* that is assumed to denote the identity relation, but also the predicational *be*: "our uniform symbolization of *be* will adequately cover both the *is* of identity and the *is* of predication." (Montague, 1974, 267).[6] As an illustration, let us take the analysis of (25).

[5] In the PTQ notation *j, m, ...* are constants of type entity (*e*), *u, v, ...* are variables of type entity (*e*), *x, y, ...* are variables for individual concepts (<*s*, *e*>), *P, Q, ...* are variables for properties of individual concepts (<*s*,<<*s*, *e*>, *t*>>), and the rounded $\mathcal{P}$, $\mathcal{Q}$, ... are variables for properties of properties of individual concepts (<*s*,<<*s*,<<*s*, *e*>, *t*>>, *t*>>) (Montague, 1974, 260).

[6] The same claim is made in Montague (1970): "the 'is' of such formulas as 'v_0 is a horse' may be identified with the 'is' of identity, and the indefinite singular term 'a horse' treated, as usual, existentially."

(25) Mary is a woman.

(26) a. $\lambda x\, \exists y\, [\mathit{woman}(y)\ \&\ {}^{\vee}x = {}^{\vee}y]$

b. $\exists y\, [\mathit{woman}(y)\ \&\ m = {}^{\vee}y]$

c. $\mathit{woman}({}^{\wedge}m)$

(26a) is the IL formula for the VP *is a woman*. It results from applying the IL representation of the copula, given in (24), to the intensionalized representation of *a woman*, followed by three λ reductions. Truth-conditionally, (26a) stands for the set of individual concepts that can truthfully be said to be a woman. (26b) is the IL formula for the sentence as a whole. It results from applying the IL formula of the subject *Mary* to the intension of the IL formula for the VP, i.e. (26a), followed by two λ reductions. In a final step, the variable in (26b) is replaced by a constant, yielding (26c); this replacement is possible since the variable and the constant are co-extensional.

In contrast to the Fregean treatment, this analysis does not reduce the role of the copula to that of a verbal affix. Instead, it assumes that the copula denotes a relation, just like the other verbs. It also captures the contribution of the determiner in predicate nominals, and it provides a uniform treatment of the predicating and identifying *be*.

3.2 Back to HPSG

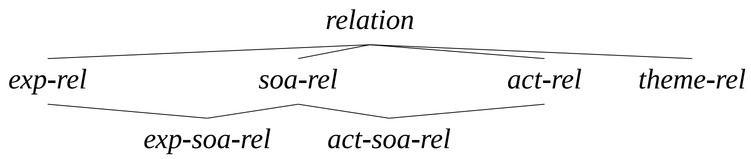
Following the lead of Quine and Montague, I do not treat the copula as semantically vacuous, but rather as denoting a relation of coreference between the indices of the subject and the predicative complement. This implies that the latter belongs to an index bearing type, in other words that its CONTENT value is of type *scope-object*, rather than of type *state-of-affairs*.

Besides, I add an optional argument whose index provides the value of the EXPERIENCER attribute. The resulting AVM looks as follows.[7]

(27)
$$\begin{bmatrix} \text{ARG-ST} & \left\langle \text{X}_{\boxed{1}}\ , (\text{Y}_{\boxed{2}}\ ,)\ \text{Z}_{\boxed{3}} \right\rangle \\ \text{SS}\,|\,\text{LOC}\,|\,\text{CONTENT} & \begin{bmatrix} \mathit{soa} & \\ \text{NUCL} & \begin{bmatrix} \mathit{exp\text{-}soa\text{-}rel} & \\ \text{EXPERIENCER} & \boxed{2}\ \mathit{index} \\ \text{SOA} & \begin{bmatrix} \mathit{soa} & \\ \text{NUCL} & \begin{bmatrix} \mathit{coref\text{-}rel} & \\ \text{THEME} & \boxed{1}\ \mathit{index} \\ \text{ATTRIBUTE} & \boxed{3}\ \mathit{index} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

[7] I follow the more recent practice in HPSG of modeling the selection of syntactic arguments in terms of the ARG-ST feature, rather than in terms of the SUBCAT feature.

(27) subsumes all verbs which take a subject-oriented predicative complement, including the copula. Those verbs take three syntactic arguments which each have a CONTENT value of type *scope-object*, and denote a state of affairs. Its nucleus is a relation of a type that is subsumed by both *soa-rel* and *exp-rel*, which implies that it has both an EXPERIENCER and a SOA attribute. Technically, this can be modeled in terms of a hierarchy of relational types, as in Davis (2001).



Each type is associated with a corresponding semantic role.

(28) $\begin{bmatrix} \textit{act-rel} & \\ \text{ACTOR} & \textit{index} \end{bmatrix}$ $\begin{bmatrix} \textit{soa-rel} & \\ \text{SOA} & \textit{soa} \end{bmatrix}$ $\begin{bmatrix} \textit{exp-rel} & \\ \text{EXP} & \textit{index} \end{bmatrix}$ $\begin{bmatrix} \textit{theme-rel} & \\ \text{THEME} & \textit{index} \end{bmatrix}$

The value of the SOA attribute is a state of affairs and has as its nucleus the relation of co-reference, which holds between the indices of the subject and the predicative complement.[8] Notice that these indices are co-referent but not token-identical. Token-identity would be too strong a requirement, since the presence of PERSON, NUMBER and GENDER features in the HPSG indices would then impose agreement for these features between the subject and the predicative complement, thus erroneously excluding (30).

(29) $\begin{bmatrix} \textit{index} & \\ \text{PERSON} & \textit{person} \\ \text{NUMBER} & \textit{number} \\ \text{GENDER} & \textit{gender} \end{bmatrix}$

(30) a. If I were you, ...
b. We are a good team.

The resulting analysis bears a resemblance to Stefan Müller's analysis of the German *erscheinen* in (2). The only important difference concerns the assumption that the predicative complement denotes a scope-object, rather than a state of affairs.

Having shown how the Montagovian treatment can be expressed in the HPSG notation, I will now demonstrate how the resulting analysis solves the problems with the Fregean treatment.

[8]If the subject has a non-referential index, as in *it is Friday*, the THEME role is left unassigned. The same holds for the EXPERIENCER role, if there is no constituent which expresses it.

3.3 Solving the problems with the Fregean treatment

The four problems with the Fregean treatment which were discussed in section 2 disappear.

First, there are no complications anymore with the semantic type of the nominal predicates, since their usual type, i.e. *scope-object*, is exactly what the copula and the other predicate selectors require. In other words, there is no need for type shifting. Moreover, this treatment does not cause any problems for the non-nominal predicates. Adjectival predicates, for instance, can be treated as scope-objects as well. In fact, the standard predicate logic treatment of adjectives is essentially the same as that of common nouns. In the same way that the common noun *dog* stands for the set of dogs, represented by '$x \mid dog(x)$', the adjective *tall* stands for the set of all things tall, represented by '$x \mid tall(x)$'. Prepositional predicates fit the mould as well. The PP in *she is in Paris*, for instance, denotes the set of all things in Paris, represented by '$x \mid in(x, paris)$'. For a more lengthy demonstration that all all types of predicative complements denote a scope-object, see Van Eynde (2008).

Second, the semantic contribution of the determiner in predicative nominals can be integrated in the usual way. For the indefinite article, this has already been spelt out in (26a-26b): It contributes an existential quantifier which is then omitted in the substitution of a constant for the variable. This treatment also works for sentences with a quantified subject, as in (31), and for sentences with a predicative nominal that is introduced by another determiner, such as *no* in (32).[9]

(31) a. Every candidate is a woman.
b. $\forall u\ [candidate(u) \rightarrow \exists v\ [woman(v)\ \&\ u = v]]$
c. $\forall u\ [candidate(u) \rightarrow woman(u)]$

(32) a. Kim is no fool.
b. $\neg\ \exists u\ [fool(u)\ \&\ u = k]$
c. $\neg\ fool(k)$

Third, there is no need to differentiate between predicational and identifying uses of the copula.

Fourth, the optional second argument can be assigned the EXPERIENCER role in the same way as the second argument of a verb like *seem*.

3.4 An extension

Besides the fact that the Montagovian treatment solves the problems with the Fregean one, it also has the advantage of being easily extensible to clauses with an object-oriented predicate, as in (33).

[9]These are formulae of first order logic, in which the variables and the proper nouns denote entities and in which the common nouns denote sets of entities. They are, hence, simpler than Montague's IL representations.

(33) I consider him a winner.

The only difference between the selectors of subject-oriented predicates and verbs like *consider* concerns the linking between the syntactic arguments and their semantic roles. Whereas the first argument supplies the theme and the (optional) second one the experiencer in the case of *seem* and the copula (27), it is the other way round in the case of *consider* and its cognates.

(34)
$$\begin{bmatrix} \text{ARG-ST} & \left\langle X_{\boxed{1}}, Y_{\boxed{2}}, Z_{\boxed{3}} \right\rangle \\ \text{SS}\,|\,\text{LOC}\,|\,\text{CONTENT} & \begin{bmatrix} \mathit{soa} & \\ \text{NUCL} & \begin{bmatrix} \mathit{exp\text{-}soa\text{-}rel} & \\ \text{EXPERIENCER} & \boxed{1}\ \mathit{index} \\ \text{SOA} & \begin{bmatrix} \mathit{soa} & \\ \text{NUCL} & \begin{bmatrix} \mathit{coref\text{-}rel} & \\ \text{THEME} & \boxed{2}\ \mathit{index} \\ \text{ATTRIBUTE} & \boxed{3}\ \mathit{index} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Notice that the CONTENT value contains the same coreference relation as in (27). It is, hence, unnecessary to assume a phantom occurrence of *be* to get this effect.

Much the same can be said about the use of *made* in (35).

(35) She made me happy.

The only difference with *consider* is that the subject has a more active role. This can be modeled by assigning it a NUCLEUS value of type *act-soa-rel*, so that the first argument is linked with the ACTOR role.

The treatment is also extensible to the use of *with* in the so-called absolute construction. A relevant example is the one in (6), repeated here:

(36) With John ill we cannot go on holiday.

The only difference between this use of *with* and the predicate selecting verbs is that it never takes an experiencer, so that its CONTENT value is less complex.

(37)
$$\begin{bmatrix} \text{ARG-ST} & \left\langle X_{\boxed{1}}, Y_{\boxed{2}} \right\rangle \\ \text{SYNSEM}\,|\,\text{LOC}\,|\,\text{CONTENT} & \begin{bmatrix} \mathit{soa} & \\ \text{NUCL} & \begin{bmatrix} \mathit{coref\text{-}rel} & \\ \text{THEME} & \boxed{1}\ \mathit{index} \\ \text{ATTRIBUTE} & \boxed{2}\ \mathit{index} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Notice also here that there is no need to assume a phantom occurrence of *be*.

3.5 Summing up

Building on the proposals of Quine and Montague I have developed an HPSG treatment of the copula which solves the problems with the Fregean treatment and which is straightforwardly extensible to other predicate selectors.

To avoid misunderstandings it is worth stressing that (27) subsumes those uses of the copula in which it combines with a predicative complement. Its other uses require another treatment, as will be demonstrated in section 4.

4 Other uses of *be*

Since (27) explicitly requires the predicative complement to denote a scope-object, it does not subsume the combination of *be* with a VP complement that denotes a state-of-affairs, as in (38).

(38) a. They are going home.
b. She was bitten by a big black dog.
c. You are to leave this room at once.

The progressive and the passive *be*, as used in (38a) and (38b), do not introduce a new state of affairs, but inherit the one of their participial complement, as spelled out in (39).

(39)
$$\begin{bmatrix} \text{ARG-ST} & \left\langle \text{NP, VP}[\textit{ptc}] : \boxed{2} \right\rangle \\ \text{SYNSEM} \mid \text{LOC} \mid \text{CONTENT} & \boxed{2}\ \textit{soa} \end{bmatrix}$$

The modal *be*, as used in (38c), introduces a state of affairs which is distinct from the one of its infinitival complement; it takes the latter as the value of its SOA argument, just like the other modals.

(40)
$$\begin{bmatrix} \text{ARG-ST} & \left\langle \text{NP, VP}[\textit{inf}] : \boxed{2} \right\rangle \\ \text{SYNSEM} \mid \text{LOC} \mid \text{CONTENT} & \begin{bmatrix} \textit{soa} & \\ \text{NUCL} & \begin{bmatrix} \textit{soa-rel} & \\ \text{SOA} & \boxed{2}\ \textit{soa} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

This, admittedly, results in a modicum of lexical ambiguity, but as compared to the distinction between the predicating and identifying uses of the copula, the distinctions between predicating *be*, progressive *be*, passive *be* and modal *be* are easy to capture and resolve. Moreover, they are independently motivated by the fact that the predicating *be* corresponds to the most commonly used copular verbs of other languages, such as *zijn* in Dutch, *sein* in German, and *être* in French, whereas the progressive, passive and modal *be* either have no translational equivalent or one that differs from the copula. The Dutch equivalent of passive *be*, for instance, is *worden*, rather than *zijn*, the one of modal *be* is *moeten* 'must' or *hebben te* 'have to', and the progressive *be* has no equivalent in Dutch.[10]

5 Conclusion

The anaylsis of the copula as a semantically vacuous word is appropriate for some of its uses, such as the progressive and the passive, but not for its use in clauses with a predicative complement. In such clauses, it denotes a relation of co-reference between the indices of the subject and the predicative complement. Moreover, it takes an optional experiencer.

References

Allegranza, Valerio. 2006. *The Signs of Determination. Constraint-based Modelling across Languages*. Peter Lang.

Bender, Emily. 2001. *Syntactic variation and linguistic competence: The case of AAVE copula absence*. Ph. D.thesis, Stanford University.

Dalrymple, Mary, Dyvik, Helge and Holloway King, Tracy. 2004. Copular complements: closed or open? In M. Butt and T. Holloway King (eds.), *Proceedings of the LFG04 Conference*, pages 188–198, Stanford: CSLI Publications.

Davis, Anthony. 2001. *Linking by types in the hierarchical lexicon*. Dordrecht: CSLI Publications.

Declerck, Renaat. 1988. *Studies on copular sentences, clefts and pseudoclefts*. Dordrecht/Leuven: Foris and Leuven University Press.

Frege, Gottlob. 1892. Ueber Begriff und Gegenstand. *Vierteljahrsschrift fuer wissenschaftliche Philosophie* 16, 192–205.

Ginzburg, Jonathan and Sag, Ivan. 2000. *Interrogative Investigations*. Stanford: CSLI.

Henri, Fabiola and Abeillé, Anne. 2007. The syntax of copular construction in Mauritian. In S. Müller (ed.), *On-Line Proceedings of HPSG 2007*, pages 130–148, Stanford University: CSLI Publications.

Kasper, Robert T. 1995. Semantics of recursive modification, hPSG Workshop in Tübingen.

Mikkelsen, L. 2005. *Copular clauses. Specification, predication and equation*. Amsterdam/Philadelphia: John Benjamins.

[10]The combination of the copula with the Dutch *aan het* + INF cannot be considered a translational equivalent of the English progressive.

Montague, Richard. 1970. English as a formal language. In B. Visentini (ed.), *Linguaggi nella socitеà e nella tecnica'*, Milano: Edizioni di Comunità.

Montague, Richard. 1974. *Formal philosophy*. New Haven: Yale University Press.

Müller, Stefan. 2002. *Complex predicates*. Stanford: CSLI.

Müller, Stefan. 2009. On predication. In Stefan Müller (ed.), *On-Line Proceedings of HPSG 2009*, page this volume, Stanford University: CSLI Publications.

Partee, Barbara. 1987. Noun Phrase Interpretation and Type-shifting Principles. In Jeroen Groenendijk, Dick de Jongh and Martin Stokhof (eds.), *Studies in Discourse Representation Theory and the Theory of generalized quantifiers*, pages 115–143, Dordrecht: Foris.

Pollard, Carl and Sag, Ivan. 1987. *Information-based Syntax and Semantics*. Stanford: CSLI Publications.

Pollard, Carl and Sag, Ivan. 1994. *Head-driven Phrase Structure Grammar*. Stanford/Chicago: CSLI Publications and University of Chicago Press.

Quine, W.V.O. 1960. *Word and object*. Cambridge, Massachusetts: MIT Press.

Van Eynde, Frank. 2008. Predicate complements. In Stefan Müller (ed.), *On-Line Proceedings of HPSG 2008*, pages 253–273, Stanford University: CSLI Publications.